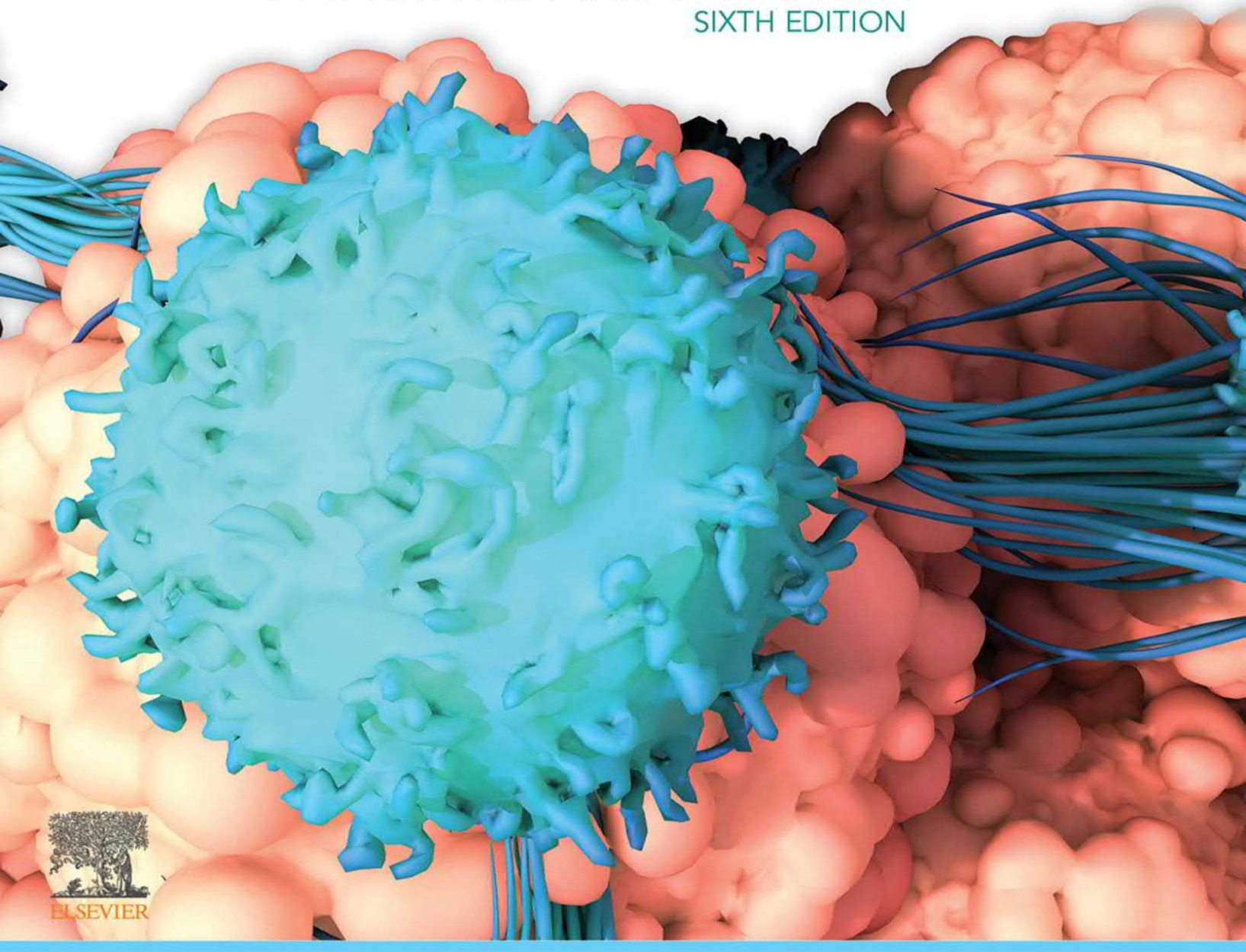
Edited by **Robert R. Rich**
Thomas A. Fleisher • Harry W. Schroeder, Jr.
Cornelia M. Weyand • David B. Corry
Jennifer M. Puck

Enhanced DIGITAL VERSION Included

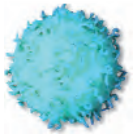
CLINICAL Immunology

PRINCIPLES AND PRACTICE

SIXTH EDITION

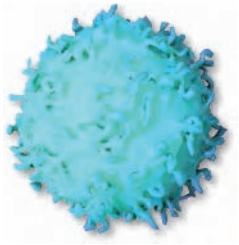


ELSEVIER



CLINICAL Immunology

PRINCIPLES AND PRACTICE



CLINICAL Immunology

PRINCIPLES AND PRACTICE

SIXTH EDITION

Robert R. Rich, MD
Dean and Professor Emeritus
Medicine
University of Alabama at Birmingham;
Former Senior Vice President for Medicine
School of Medicine
University of Alabama at Birmingham
Birmingham, AL, United States

Thomas A. Fleisher, MD
Executive Vice President
American Academy of Allergy, Asthma, and Immunology
Milwaukee, WI, United States;
Scientist Emeritus
National Institutes of Health Clinical Center
Bethesda, MD, United States

Harry W. Schroeder Jr., MD, PhD
Professor
Medicine, Microbiology, and Genetics
University of Alabama at Birmingham
Birmingham, AL, United States

Cornelia M. Weyand, MD, PhD
Professor Emerita
Medicine
Stanford University
Stanford, CA, United States;
Professor
Medicine and Immunology
Mayo Clinic Alix School of Medicine
Rochester, MN, United States

David B.Corry, MD
Professor
Medicine and Pathology and Immunology
Baylor College of Medicine
Houston, TX, United States

Jennifer M. Puck, MD
Professor of Pediatrics
University of California San Francisco
San Francisco, CA, United States

First edition 1996
Second edition 2001
Third edition 2008
Fourth edition 2013
Fifth edition 2019
Sixth edition 2023

Notices

Practitioners and researchers must always rely on their own experience and knowledge in evaluating and using any information, methods, compounds, or experiments described herein. Because of rapid advances in the medical sciences, in particular, independent verification of diagnoses and drug dosages should be made. To the fullest extent of the law, no responsibility is assumed by Elsevier, authors, editors, or contributors for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions, or ideas contained in the material herein.

ISBN: 9780702081651

Content Strategists: Robin Carter, Mary Hegeler
Content Development Specialists: Louise Cook, Jennifer Ehlers, Malvika Shah
Project Manager: Andrew Riley
Design: Amy Buxton
Illustration Manager: Deanna Sorenson
Marketing Manager: Kate Bresnahan

Printed in India

Last digit is the print number: 9 8 7 6 5 4 3 2 1

CONTENTS

PART X Transplantation

PART XI Diagnostic Immunology

PREFACE TO THE FIRST EDITION

Clinical immunology is a discipline with a distinguished history, rooted in the prevention and treatment of infectious diseases in the late nineteenth and early twentieth centuries. The conquest of historical scourges such as smallpox and (substantially) polio and relegation of several other diseases to the category of medical curiosities is often regarded as the most important achievement of medical science of the past fifty years. Nevertheless, the challenges facing immunologists in the efforts to control infectious diseases remain formidable; HIV infection, malaria, and tuberculosis are but three examples of diseases of global import that elude control despite major commitments of monetary and intellectual resources.

Although firmly grounded in the study and application of defenses to microbial infection, since the 1960s, clinical immunology has emerged as a far broader discipline. Dysfunction of the immune system has been increasingly recognized as a pathogenic mechanism that can lead to an array of specific diseases and failure of virtually every organ system. Paradoxically, although the importance of the immune system in disease pathogenesis is generally appreciated, the place of clinical immunology as a practice discipline has been less clear. As most of the noninfectious diseases in the human immune system eventually lead to the failure of other organs, it has been organ-specific subspecialists who have usually dealt with their consequences. Recently, however, the outlook has begun to change as new diagnostic tools increasingly allow the theoretical possibility of intervention much earlier in disease processes, often before irreversible target organ destruction occurs. More importantly, this theoretical possibility is increasingly realized as clinical immunologists find themselves in the vanguard of translating molecular medicine from laboratory bench to patient bedside.

In many settings, clinical immunologists today function as primary care physicians in the management of patients with immune-deficiency, allergic, and autoimmune diseases. Indeed many influential voices in the clinical disciplines of allergy and rheumatology support the increasing coalescence of these traditional subspecialties around their intellectual core of immunology. In addition to his or her role as a primary care physician, the clinical immunologist is increasingly being looked to as a consultant, as scientific and clinical advances enhance his or her expertise. The immunologist with a "generalist" perspective can be particularly helpful in the application of unifying principles of diagnosis and treatment across the broad spectrum of immunologic diseases.

Clinical Immunology: Principles and Practice has emerged from this concept of the clinical immunologist as both primary care physician and expert consultant in the management of patients with immunologic diseases. It opens in full appreciation of the critical role of fundamental immunology in this rapidly evolving clinical discipline. Authors of basic science chapters were asked, however, to cast their subjects in a context of clinical relevance. We believe the result is a well-balanced exposition of basic immunology for the clinician.

The initial two sections on basic principles of immunology are followed by two sections that focus in detail on the role of the immune system in defenses against infectious organisms. The approach is two-pronged. It begins first with a systematic survey of immune responses to pathogenic agents followed by a detailed treatment of immunologic deficiency syndromes. Pathogenic mechanisms of both congenital and acquired immune deficiency diseases are discussed, as are the infectious complications that characterize these diseases. Befitting its importance, the subject of HIV infection and AIDS receives particular attention, with separate chapters on the problem of infection in the immunocompromised host, HIV infection in children, anti-retroviral therapy, and current progress in the development of HIV vaccines.

The classic allergic diseases are the most common immunologic diseases in the population, ranging from atopic disease to drug allergy to organ-specific allergic disease (e.g., of the lungs, eye, and skin). They constitute a foundation for the practice of clinical immunology, particularly for those physicians with a practice orientation defined by formal subspecialty training in allergy and immunology. A major section is consequently devoted to these diseases, with an emphasis on pathophysiology as the basis for rational management.

The next two sections deal separately with systemic and organ-specific immunologic diseases. The diseases considered in the first of these sections are generally regarded as the core practice of the clinical immunologist with a subdisciplinary emphasis in rheumatology. The second section considers diseases of specific organ failure as consequences of immunologically mediated processes that may involve virtually any organ system. These diseases include, as typical examples, demyelinating diseases, insulin-dependent diabetes mellitus, glomerulonephritides, and inflammatory bowel diseases. It is in the management of such diseases that the discipline of clinical immunology will have an increasing role as efforts focus on intervention early in the pathogenic process and involve diagnostic and therapeutic tools of ever-increasing sophistication.

One of the major clinical areas in which the expertise of a clinical immunologist is most frequently sought is that of allogeneic organ transplantation. A full section is devoted to the issue of transplantation of solid organs, with an introductory chapter on general principles of transplantation and management of transplantation rejection followed by separate chapters dealing with the special problems of transplantation of specific organs or organ systems.

Appreciation of both the molecular and clinical features of lymphoid malignancies is important to the clinical immunologist regardless of subspecialty background, notwithstanding the fact that primary responsibility for the management of such patients will generally fall to the hematologist/oncologist. A separate section is consequently devoted to the lymphocytic leukemias and lymphomas that constitute the majority of malignancies seen in the context of a clinical immunology practice. The separate issues of immune responses to tumors and immunological strategies to treatment of malignant diseases are subjects of additional chapters.

Another important feature is the attention to therapy of immunologic diseases. This theme is constant throughout the chapters on allergic and immunologic diseases, and because of the importance the editors attach to clinical immunology as a therapeutic discipline, an extensive section is also devoted specifically to this subject. Subsections are devoted to issues of immunologic reconstitution, with three chapters on the treatment of

immunodeficiencies, malignancies, and metabolic diseases by bone marrow transplantation. Also included is a series of chapters on pharmaceutical agents currently available to clinical immunologists, both as anti-allergic and anti-inflammatory drugs, as well as newer agents with greater specificity for T cell-mediated immune responses. The section concludes with a series of chapters that address established and potential applications of therapeutic agents and approaches that are largely based on the new techniques of molecular medicine. In addition to pharmaceutical agents, the section deals in detail with such subjects as apheresis, cytokines, monoclonal antibodies and immunotoxins, gene therapy, and new experimental approaches to the treatment of autoimmunity. The book concludes with a section devoted to approaches and specific techniques involved in the diagnosis of immunologic diseases. The use of the diagnostic laboratory in the evaluation of complex problems of immunopathogenesis has been a hallmark of the clinical immunologist since the inception of the discipline, and many clinical immunologists serve as directors of diagnostic immunology laboratories. Critical assessment of the utilization of techniques ranging from lymphocyte cloning to flow cytometric phenotyping to molecular diagnostics is certain to continue as an important function of the clinical immunologist, particularly in his or her role as an expert consultant.

In summary, we have intended to provide the reader with a comprehensive and authoritative treatise on the broad subject of clinical immunology, with particular emphasis on the diagnosis and treatment of immunological diseases. It is anticipated that the book will be used most frequently by the physician specialist practicing clinical immunology, both in his or her role as a primary physician and as a subsequent consultant. It is hoped, however, that the book will also be of considerable utility to the non-immunologist. Many of the diseases discussed authoritatively in the book are diseases commonly encountered by the generalist physician. Indeed, as noted, because clinical immunology involves diseases of virtually all organ systems, competence in the diagnosis and management of immunological diseases is important to virtually all clinicians. The editors would be particularly pleased to see the book among the references readily available to the practicing internist, pediatrician, and family physician.

Robert R. Rich
Thomas A. Fleisher
Benjamin D. Schwartz
William T. Shearer
Warren Strober
1996

PREFACE TO THE SIXTH EDITION

At the time this preface was prepared, our world was being arguably faced with its worst infectious pandemic since the global influenza A infection of the early 20th century. That pandemic infected ~500 million people, or ~1/3 of the total world population, and caused at least 50 million deaths. The current 21st century COVID-19 pandemic was estimated in September 2021 to have infected >230 million individuals out of 7.7 billion (~3%), causing ~4.71 million deaths; and these numbers are also staggering and are necessarily incomplete. But as we learn more about the prevention, diagnosis, and care of persons with COVID-19 infections (both symptomatic and asymptomatic), we have reason to believe that despite the enormous case load, we will not approach the death rate of the 1918–20 pandemic.

This improved survival reflects in part the knowledge that has been gained over the past century from research in immunology, which has been contributed by investigators and clinicians throughout the world. This knowledge has advanced our ability to prevent and manage pandemics in general, giving us hope to face future emerging infectious disease challenges as well.

Although clinical immunology and this book are strongly grounded in the study of and application to microbial infections, both the book and the discipline are far broader. Dysfunction of the immune system is recognized as a pathogenic mechanism that can lead to diseases and failure of virtually every organ system. Fortunately, advances in the prevention and treatment of immunologic diseases also offer enhanced intervention, often before irreversible target organ damage or destruction has occurred. Thus, clinical immunologists are frequently in the vanguard of translating molecular medicine from the laboratory bench to the patient bedside.

In the United States, clinical immunologists often function as primary care providers for patients with a wide variety of disorders of immune function, including immune deficiency, allergic disease, and autoimmunity. As a result, although clinical immunology has not been constituted in this country as a formal subspecialty, many influential voices in the "official" disciplines of allergy and rheumatology regard themselves foremost as clinical immunologists. We trust that this textbook will prove useful to both clinical generalists and clinical immunology subspecialists.

The book opens with three sections dedicated to the fundamental sciences that underlie clinical immunology. However, authors of the basic science chapters were asked to cast their chapters in a context of clinical relevance. We believe this has been accomplished. The basic science chapters are followed by sections on Immune Deficiency and Immune Regulatory Disorders; Allergic Diseases; Systemic Immune Diseases; Organ-Specific Inflammatory Disorders; Immunology and Immunotherapy of Neoplasia; Medical Management of Immunologic Diseases; Transplantation of Tissues and Organs; and the Technologies of Diagnostic Immunology.

We have preserved features in the book that were well received in previous editions. Chapters are generously illustrated, and all chapters contain a Key Concepts summary Box (commonly in bulleted form) as well as an *On the Horizon* box, in which authors look to research opportunities for important advances over the next five to ten years. Furthermore, due to the extraordinarily cross-disciplinary nature of clinical immunology, it is our hope that investigators working in one area might find new ideas and opportunities in the *On the Horizon* boxes outside their primary area of focus. Other boxes similarly summarize content with *Clinical Relevance, Clinical Pearls*, and *Therapeutic Principles*.

Now in its 6th edition since it was first published in 1996, this book represents the achievements and provides access to the expertise of ~200 individual contributors. As editors, we are deeply grateful for the thousands of hours of work put into this effort by our colleagues in Clinical Immunology around the globe. We also note the exceptional support that the preparation of this edition has received from publishing experts at Elsevier, in particular Robin Carter, Louise Cook, Jennifer Ehlers, and Andrew Riley. Thank you so much, Robin, Louise, Jennifer, and Andrew. The task would have been impossible without your expertise.

Robert R. Rich
Thomas A. Fleisher
Harry W. Schroeder Jr.
Cornelia M. Weyand
David B. Corry
Jennifer M. Puck

LIST OF CONTRIBUTORS

The editors would like to acknowledge and offer grateful thanks for the input of all previous editions' contributors, without whom this new edition would not have been possible.

Roshini Sarah Abraham, PhD
Professor of Clinical Pathology; Director, Diagnostic Immunology Laboratory
Department of Pathology and Laboratory Medicine
Nationwide Children's Hospital and The Ohio State University College of Medicine
Columbus, OH, United States

Behdad Afzali, MBBS, PhD, MRCP
Earl Stadtman Investigator
Kidney Diseases Branch
National Institute of Diabetes and Digestive and Kidney Diseases
National Insitute of Health
Bethesda, MD, United States

Ana Águeda, MD
Department of Rheumatology
Centro Hospitalar do Baixo Vouga
Aveiro, Portugal

Cem Akin, MD, PhD
Professor of Medicine
Division of Internal Medicine
University of Michigan
Ann Arbor, MI, United States

Cristina Albanesi, PhD
Laboratory of Experimental Immunology
Istituto Dermopatico dell'Immacolata - Istituto di Ricovero e Cura a Carattere Scientifico (IDI-IRCCS)
Rome, Italy

Brendan Antiochos, MD
Department of Medicine
Johns Hopkins University
Baltimore, MD, United States

Cynthia Aranow, MD
Feinstein Institutes for Medical Research
Northwell Health
Manhasset, NY, United States

John P. Atkinson, MD
Department of Medicine
Washington University School of Medicine
St. Louis, MO, United States

Thomas M. Aune, PhD
Professor
Department of Medicine
Vanderbilt University Medical Center
Nashville, TN, United States

Subash Babu, MBBS, PhD
Scientific Director
ICER
National Institutes of Health-NIRT-ICER
Chennai, India;
Staff Scientist, Laboratory of Parasitic Diseases
National Institute of Allergy and Infectious Diseases (NIAID)
Bethesda, MA, United States

Justin M. Balko, PharmD, PhD
Associate Professor of Medicine
Associate Professor of Pathology, Microbiology and Immunology
Vanderbilt University Medical Center
Nashville, TN, United States

Mark Ballow, MD
Professor of Pediatrics
Department of Pediatrics
Morsani College of Medicine
St Petersburg, FL, United States;
Former Chief, Division of Allergy and Immunology
Department of Pediatrics
SUNY Buffalo
Buffalo, NY, United States

Rachel Bean, MD
Special Volunteer
Division of Intramural Research
National Institute of Allergy and Infectious Diseases
National Institutes of Health
Bethesda, MD, United States

Alexia Belavgeni, MSc
Department of Nephrology
University Hospital Carl Gustav Carus
Technische Universität Dresden
Dresden, Germany

Claudia Berek, PhD
Department of B Cell Immunology
Deutsches Rheuma Forschungszentrum
Berlin, Germany

Timothy Beukelman, MD, MSCE
Associate Professor
Department of Pediatrics, Division of Rheumatology
University of Alabama at Birmingham Heersink School of Medicine
Birmingham, AL, United States

Vivien Beziat, PhD
Assistant Professor
Laboratory of Human Genetics of Infectious Diseases
Institut Imagine/University of Paris
Paris, France

Lynn Bimler, PhD
Postdoctoral Fellow
Department of Medicine
Baylor College of Medicine
Houston, TX, United States

J. Andrew Bird, MD
Associate Professor
Division of Allergy and Immunology
Department of Pediatrics
University of Texas Southwestern Medical Center;
Director
Food Allergy Center
Children's Medical Center
Dallas, TX, United States

Sarah E. Blutt, PhD
Associate Professor
Molecular Virology and Microbiology
Associate Professor
Molecular and Cellular Biology
Baylor College of Medicine
Houston, TX, United States

Mark Boguniewicz, MD
Professor, Division of Allergy-Immunology
Department of Pediatrics
National Jewish Health and University of Colorado School of Medicine
Denver, CO, United States

Bertrand Boisson, PhD
Assistant Professor
Laboratory of Human Genetics of Infectious Diseases
Institut Imagine/University of Paris
Paris, France;
St. Giles Laboratory of Infectious Diseases
The Rockefeller University
New York, NY, United States

Stéphanie Boisson-Dupuis, PhD
St Giles Laboratory of Human Genetics of Infectious Diseases
the Rockefeller University
New York, NY, United States

Elena Borzova, PhD
Professor
Department of Dermatology and Venereology
I.M. First Moscow State Medical University
Moscow, Russian Federation

Maria Bottazzi, PhD
Associate Dean
National School of Tropical Medicine
Baylor College of Medicine
Houston, TX, United States

Prosper N. Boyaka, PhD
Professor
Veterinary Biosciences
The Ohio State University
Columbus, OH, United States

John M. Bridges, MD
Associate and Fellow, Clinical Immunology and Rheumatology
Children's Hospital of Alabama
University of Alabama at Birmingham
Birmingham, AL, United States

Sarah K. Browne, MD
Medical Officer
Division of Vaccines and Related Product Applications
Food and Drug Administration
Silver Spring, MD, United States

A. Wesley Burks, MD
Dean of the UNC School of Medicine and CEO of UNC Health
Vice Chancellor for Medical Affairs
University of North Carolina
Chapel Hill, NC, United States

Jacinta Bustamante, MD, PhD
Associate Professor
Human Genetics of Infectious Diseases – GHMI
Imagine Institute/University of Paris
Paris, France

Jean-Laurent Casanova, MD, PhD
Levy Family Professor
The Rockefeller University;
Investigator
Howard Hughes Medical Institute;
Head
St. Giles Laboratory of Human Genetics of Infectious Diseases
Laboratory of Human Genetics of Infectious Diseases
Institut Imagine/University of Paris
Paris, France

Alice Y. Chan, MD, PhD
Department of Pediatrics
University of California San Francisco
San Francisco, CA, United States

Edwin S. L. Chan, MB, ChB, FRCPC
Adjunct Associate Professor of Medicine
Department of Medicine
New York University School of Medicine
New York, NY, United States

Walter Winn Chatham, MD
Professor of Medicine, Clinical Immunology and Rheumatology
University of Alabama at Birmingham
Birmingham, AL, United States

Javier Chinen, MD, PhD
Allergist and Immunologist
Baylor College of Medicine
The Woodlands, Texas Children's Hospital
Humble, TX, United States

Lisa Christopher-Stine, MD
Assistant Professor of Pathology and Laboratory Medicine
University of Rochester Medical Center
Rochester, NY, United States

Emily Coates, PhD
Director of Translational Science
Clinical Trials Program
Vaccine Research Center
National Institute of Allergy and Infectious Diseases
Bethesda, MD, United States

Andrew P. Cope, BSc, MBBS, PhD, FRCP
Centre for Rheumatic Diseases
King's College London
London, UK

David B. Corry, MD
Professor
Medicine and Pathology and Immunology
Baylor College of Medicine
Houston, TX, United States

Joana Cosme, MD
Allergy and Clinical Immunology Consultant
Department of Immunoallergology
Centro Hospitalar Universitário de Lisboa Norte
Lisbon, Portugal

Randy Q. Cron, MD, PhD
Professor of Pediatrics and Medicine
University of Alabama at Birmingham;
Director of Pediatric Rheumatology
Children's of Alabama
Birmingham, AL, United States

Marinos C. Dalakas, MD, FAAN
Professor of Neurology
Chief Neuromuscular Division
Department of Neurology
Thomas Jefferson University
Philadelphia, PA, United States

Sara M. Dann, PhD
Assistant Professor
Internal Medicine
University of Texas Medical Branch
Galveston, TX, United States

Satya Das, MD, MSCI
Assistant Professor
Medicine
Vanderbilt University Medical Center
Nashville, TN, United States

Molly M. Daughety, MD
Physician
Hematology Oncology
Duke University Medical Center
Durham, NC, United States

Betty Diamond, MD
Feinstein Institutes for Medical Research
Northwell Health
Manhasset, NY, United States

Angela Dispenzieri, MD
Professor of Medicine and Laboratory Medicine
Mayo Clinic
Rochester, MN, United States

Stephen R. Durham, MA, MD, FRCP
Professor of Allergy and Respiratory Medicine
Allergy and Clinical Immunology
NHLI, Imperial College London
London, UK

Todd N. Eagar, PhD
Assistant Professor
Pathology and Genomic Medicine
Houston Methodist Hospital Research Institute
Director of HLA and Transplant Immunology Laboratory
Pathology and Genomic Medicine
Houston Methodist Hospital
Houston, TX, United States;
Assistant Professor
Laboratory Medicine
Weill Cornell Medical College
New York, NY, United States

Michelle Al-Hosni, MD
Allergy and Clinical Immunology Fellow
University of Michigan
Ann Arbor, MI, United States

Sarah Elitzur, MD
Senior Physician
Pediatric Hematology-Oncology
Schneider Children's Medical Center
Petah Tikva, Israel

Craig A. Elmets, MD
Professor and Chair
Dermatology
University of Alabama at Birmingham
Birmingham, AL, United States

Doruk Erkan, MD
Associate Professor
Rheumatology
Hospital for Special Surgery
New York, NY, United States

Thomas A. Fleisher, MD
Executive Vice President
American Academy of Allergy, Asthma and Immunology
Milwuakee, WI, United States;
Scientist Emeritus, National Institutes of Health Clinical Center
Bethesda, MD, United States

Luz Fonacier, MD
Professor of Medicine
Department of Medicine
NYU Long Island School of Medicine;
Head of Allergy
Department of Medicine
NYU Langone Hospital, Long Island;
Training Program Director
Allergy and Immunology
Department of Medicine
NYU Langone Hospital, Long Island
Mineola, NY, United States

Andrew P. Fontenot, MD
Professor of Medicine
Department of Medicine
University of Colorado Anschutz Medical Campus
Aurora, CO, United States

George Fragoulis, MD, PhD
Rheumatologist
Institute of Infection, Immunity and Inflammation
University of Glasgow
Glasgow, UK;
First Propedeutic Internal Medicine
National and Kapodistrian University of Athens
Athens, Greece

Ivo M. B. Francischetti, MD, PhD
Hematopathology Fellow
Laboratory of Pathology
National Institutes of Health
Bethesda, MD, United States

Tilo Freiwald, MD
Research Fellow
Immunoregulation Section
Kidney Diseases Branch
National Institute of Diabetes and Digestive and Kidney Diseases
National Institutes of Health
Bethesda, MD, United States

Anthony J. Frew, MD, FRCP†
Professor of Allergy and Respiratory
Medicine
Department of Respiratory Medicine
Royal Sussex County Hospital
Brighton, UK

Kohtaro Fujihashi, DDS, PhD
Project Professor
Division of Clinical Vaccinology International Research and
Development Center for Mucosal Vaccines
The Institute of Medical Science The University of Tokyo
Tokyo, Japan;
Professor Emeritus
Pediatric Dentistry
The University of Alabama at Birmingham
Birmingham, AL, United States

Massimo Gadina, PhD
Chief
Translational Immunology Section
National Institute of Arthritis and Musculoskeletal and Skin Diseases
National Institutes of Health
Bethesda, MD, United States

Laurent Gapin, PhD
Professor
Department of Immunology and Microbiology
University of Colorado Anschutz Medical Campus
Aurora, CO, United States;
Department of Immunology and Genomic Medicine
National Jewish Health
Denver, CO, United States

Moshe E. Gatt, MD
Senior Hematologist
Hematology
Hadassah Hebrew University Medical Center
Jerusalem, Israel

M. Eric Gershwin, MD, PhD
Chief Division of Rheumatology Allergy and Clinical Immunology
Department of Internal Medicine
University of California
Davis, CA, United States

Susan L. Gillespie, MD, PhD
Associate Professor
Pediatrics
Baylor College of Medicine
Houston, TX, United States

Lynn K. Gordon, MD, PhD
Professor Emeritus
Department of Ophthalmology
David Geffen School of Medicine at UCLA
Los Angeles, CA, United States

Jörg J. Goronzy, MD, PhD
Department of Medicine
Mayo Clinic College of Medicine and Science
Rochester, MN, United States

Clive E. H. Grattan, MA, MD, FRCP
Consultant Dermatologist
St. John's Institute of Dermatology
London, UK

Neil S. Greenspan, MD, PhD
Professor of Pathology
University Hospitals Cleveland Medical Center
Case Western Reserve University
Cleveland, OH, United States

Anna Gschwend, MD, PhD
Allergology/Immunology Specialist
University Clinic for Pulmonary Medicine and Allergology
Inselspital/University Hospital
Bern, Switzerland

Claire E. Gustafson, PhD
Scientist III
Allen Institute for Immunology
Seattle, WA, United States

Tillie-Louise Hackett, PhD
Professor
Anesthesiology, Pharmacology and Therapeutics
Centre for Heart Lung Innovation
University of British Columbia
Vancouver, BC, Canada

Robert G. Hamilton, PhD, DABMLI
Professor
Medicine and Pathology
Johns Hopkins University School of Medicine
Baltimore, MD, United States

Myra Happe, PhD
Scientific Advisor
Clinical Trials Program, Vaccine Research Center
National Institute of Allergy and Infectious Diseases
National Institutes of Health
Bethesda, MD, United States

Leonard C. Harrison, MB, BS, MD, DSc, DMedSci, FRACP, FRCPA, FAHMS
Professor
Population Health and Immunity Division
Walter and Eliza Hall Institute of Medical Research
Parkville, Melbourne, VIC, Australia;
Professor
Medical Biology
University of Melbourne
Melbourne, VIC, Australia

Arthur Helbling, MD
Professor; Head of the Allergology and Clinical Immunology
Policlinic University Clinic for Pulmonary Medicine and Allergology
Inselspital/University Hospital
Bern, Switzerland

Emmaline Heckmann
Research Scientist
Department of Molecular Virology and Microbiology
Baylor College of Medicine
Houston, TX, United States

Kristin Hogquist, PhD
Professor
Center for Immunology
University of Minnesota
Minneapolis, MN, United States

Tobias M. Hohl, MD, PhD
Member
Department of Medicine
Memorial Sloan Kettering Cancer Center
New York, NY, United States

Steven M. Holland, MD
Director
Division of Intramural Research
National Institute of Allergy and Infectious Diseases
National Institutes of Health (NIH)
Chief
Immunopathogenesis Section
National Institute of Allergy and Infectious Diseases
National Institutes of Health (NIH)
Bethesda, MD, United States

Peter J. Hotez, MD, PhD
Dean, National School of Tropical Medicine
Professor, Pediatrics and Molecular and Virology and Microbiology
Head, Section of Pediatric Tropical Medicine
Baylor College of Medicine;
Director, Sabin Vaccine Institute Texas Children's Hospital Center for Vaccine Development
Houston, TX, United States

Katherine Houser, PhD
Senior Scientific Advisor
NationaI Institutes of Health
Bethesda, MD, United States

Nicholas D. Huntingdon, BSc, PhD
Laboratory Head
Department of Biochemistry and Molecular Biology
Monash University
Clayton, VIC, Australia

Tracy Hwangpo, MD, PhD
Instructor of Medicine
Department of Medicine
University of Alabama at Birmingham
Birmingham, AL, United States

Shai Izraeli, MD
Chair
Pediatric Hematology and Oncology
Schneider Children's Medical Center of Israel
Petach Tikva, Israel

Elaine S. Jaffe, MD
Head, Hematopathology Section
Laboratory of Pathology
National Cancer Institute
National Institutes of Health
Bethesda, MD, United States

Sirpa Jalkanen, MD, PhD
Professor
MediCity
University of Turku
Turku, Finland

Anuja Java, MD
Assistant Professor of Medicine
Department of Internal Medicine
Washington University School of Medicine in St. Louis;
Assistant Professor of Medicine
Department of Internal Medicine
Veterans Affairs Medical Center
St Louis, MO, United States

Douglas B. Johnson, MD, MSCI
Associate Professor of Medicine
Department of Medicine
Vanderbilt University Medical Center
Nashville, TN, United States

Tory P. Johnson, PhD
Department of Neurology
Johns Hopkins University
Baltimore, MD United States

Michael B. Jordan, MD
Professor
Pediatrics, Divisions of Immunobiology and Bone Marrow Transplantation and Immune Deficiency
University of Cincinnati
Cincinnati Children's Hospital
Cincinnati, OH, United States

Shyam R. Joshi, MD
Assistant Professor of Medicine
Section of Allergy and Immunology
Oregon Health and Science University
Portland, OR, United States

Emmanuelle Jouanguy, PhD
Associate Professor
Laboratory of Human Genetics of Infectious Diseases
Imagine Institute/University of Paris
Paris, France

Henry J. Kaminski, MD
Chairman, Neurology
George Washington University
Washington, DC, United States

Stefan H. E. Kaufmann, PhD
Professor and Director Emeritus
Max Planck Institute for Infection Biology
Berlin, Germany

David A. Khan, MD
Professor of Medicine and Pediatrics
Division of Allergy and Immunology
University of Texas Southwestern Medical Center
Dallas, TX, United States

Farrah Kheradmand, MD
Professor
Pathology, Immunology, Medicine
Baylor College of Medicine
Houston, TX, United States

Dilawar Singh Khokar, MD
Allergy and Clinical Immunology Fellow
University of Michigan
Ann Arbor, MI, United States

Paneez Khoury, MD MHSc
Human Eosinophil Section
Laboratory of Parasitic Diseases
National Institute of Allergy and Infectious Diseases
National Institutes of Health
Bethesda, MD United States

Bruce S. Klein, MD
Gerard B. Odell Professor and Shirley S. Matchette Professor
Department of Pediatrics;
University of Wisconsin, Madison;
Professor, Internal Medicine;
Professor, Internal Medicine, Medical Microbiology and Immunology
University of Wisconsin, Madison
Madison, WI, United States

Amy D. Klion, MD
Human Eosinophil Section
Laboratory of Parasitic Diseases, National Institute of Allergy and Infectious Diseases
National Institutes of Health
Bethesda, MD, United States

Donald B. Kohn, MD
Professor
Microbiology, Imunology and Molecular Genetics
University of California, Los Angeles
Los Angeles, CA, United States

Michihito Kono, MD, PhD
Assistant Professor
Department of Rheumatology, Endocrinology and Nephrology
Faculty of Medicine
Hokkaido University
Sapporo, Japan

Robert Korngold, PhD
Emeritus Member
Center for Discovery and Innovation
Hackensack Meridian Health
Nutley, NJ, United States

Vasiliki Koulouri, MD, MSc
PhD Student
Department of Physiology
National and Kapodistrian University of Athens
School of Medicine
Athens, Greece

Douglas B. Kuhns, PhD
Principal Scientist
Neutrophil Monitoring Lab
Applied/Developmental Research Directorate
Frederick National Laboratory for Cancer Research
Frederick, MD, United States

Hrishikesh S. Kulkarni, MD, MSCI
Instructor
Medicine, Division of Pulmonary and Critical Care Medicine
Washington University School of Medicine
St. Louis, MO, United States

Caroline Y. Kuo, MD
Division of Allergy and Immunology
Department of Pediatrics
David Geffen School of Medicine at UCLA
Los Angeles, CA, United States

Linda L. Kusner, PhD
Associate Research Professor
Pharmacology and Physiology
George Washington University
Washington, DC, United States

Arash H. Lahouti, MD
Assistant Professor of Pathology and Laboratory Medicine
University of Rochester Medical Center
Rochester, NY, United States

Laura C. Lane, MD
Gynecologist
Sky Ridge Medical Center
Lone Tree, CO, United States

Arian Laurence, PhD, MRCP, FRCPath
Visiting Research Fellow
Nuffield Department of Medicine
Oxford University
Oxford, UK;
Clinical Research Fellow
Institute of Cellular Medicine
University of Newcastle
Newcastle upon Tyne, UK

Joyce S. Lee, MD
Associate Professor
Department of Medicine
University of Colorado Anschutz Medical Campus
Aurora, CO, United States

S. Thera Lee, PhD
Professor
Center for Immunology University of Minnesota
Minneapolis, MN, United States

Donald Y. M. Leung, MD, PhD
Edelstein Family Chair of Pediatric Allergy-Immunology
Department of Pediatrics
National Jewish Health
Professor of Pediatrics
Department of Pediatrics
University of Colorado Medical School
Denver, CO, United States

Ofer Levy, MD, PhD
Director, Precision Vaccines Program
Division of Infectious Diseases
Boston Children's Hospital;
Professor of Pediatrics
Harvard Medical School
Boston, MA, United States;
Associate Member
Broad Institute of MIT and Harvard
Cambridge, MA, United States

Dorothy E. Lewis, PhD, MS
Research Scientist
Faculty Development
Houston Methodist Research Institute;
Adjunct Professor
Pathology and Immunology
Baylor College of Medicine;
Adjunct Professor
Internal Medicine
University of Texas Health Science Center
Houston, TX, United States

Evan Li, MD
Assistant Professor
Department of Medicine
Baylor College of Medicine
Houston, TX, United States

Peter Libby, MD
Cardiovascular Specialist
Cardiovascular Medicine
Brigham and Women's Hospital/Harvard Medical School
Boston, MA, United States

Andrew H. Lichtman, MD, PhD
Professor of Pathology
Brigham and Women's Hospital;
Professor of Pathology, Harvard Medical School
Boston, MA, United States

Andreas Linkermann, MD
Professor
Department of Nephrology
University Clinic Carl Gustav Carus
Dresden, Saxony, Germany

Michail S. Lionakis, MD, DSc
Chief, Fungal Pathogenesis Section
Deputy Chief, Laboratory of Clinical Immunology and Microbiology
National Institute of Allergy and Infectious Diseases (NIAID)
National Institutes of Health (NIH)
Bethesda, MD, United States

M. Kathryn Liszewski, PhD
Assistant Professor of Medicine
Medicine/Rheumatology
Washington University School of Medicine
St. Louis, MO, United States

Michael D. Lockshin, MD
Director, Barbara Volcker Center
Hospital for Special Surgery
Weill Cornell Medical College
New York, NY, United States

Debra Long Priel, MS
Clinical Services Program
Leidos Biomedical Research, Inc.
Frederick, MD, United States

Adi Zoref Lorenz, MD
Instructor
Hematology Institute
Meir Medical Center
Sackler Faculty of Medicine
Tel Aviv, Israel;
Division of Immunobiology
Cincinnati Children's Hospital
Cincinnati, OH, USA

Ralf J. Ludwig, MD
Professor
Lübeck Institute of Experimental Dermatology
University of Lübeck
Lübeck, Germany

Amber Luong, MD, PhD
Professor
Otorhinolaryngology - Head and Neck Surgery
McGovern Medical School of The University of
Texas Health Science Center
Houston, TX, USA

Raashid Ahmed Luqmani, MD
Oxford NIHR Biomedical Research Centre University of Oxford
Oxford, UK

Meggan Mackay, MD, MS
Feinstein Institutes for Medical Research
Northwell Health
Manhasset, NY, United States

Alfred Mahr, MD, MPH, PhD
Professor
Clinic for Rheumatology
Kantonsspital St. Gallen
St. Gallen, Switzerland

Tamir Malley, MD
Royal Brompton Hospital
London, UK

Elinor C. Mannon, MD
Department of Physiology
Medical College of Georgia at Georgia Regents University
Augusta, GA, United States

Peter J. Mannon, MD
Professor of Medicine
Chief, Division of Gastroenterology and Hepatology
University of Nebraska Medical Center
Omaha, NE, United States

Roslyn B. Mannon, MD
Professor of Medicine, Pathology and Microbiology
University of Nebraska Medical Center
Omaha, NE, United States

Michael P. Manns, MD
Hannover Medical School
Hannover, Germany

Anthony Maresso, PhD
Associate Professor
Virology and Microbiology
Baylor College of Medicine
Houston, TX, United States

Scott M. Matson, MD
Assistant Professor
Pulmonary, Critical Care and Sleep Medicine
University of Kansas
Kansas City, KS, United States

Clio P. Mavragani, MD, PhD
Assistant Professor
Department of Physiology
National and Kapodistrian University of Athens
Athens, Greece

Craig L. Maynard, PhD
Assistant Professor
Pathology and Medicine
University of Alabama at Birmingham
Birmingham, AL, United States

Douglas R. McDonald, MD, PhD
Assistant Professor in Pediatrics
Immunology
Boston Children's Hospital
Boston, MA, United States

Françoise Meylan, PhD
National Institute of Arthritis and Musculoskeletal and Skin Diseases
National Institutes of Health
Translational Immunology Section
Bethesda, MD, USA

Stephen D. Miller, PhD
Professor
Department of Microbiology-Immunology
Northwestern University Medical School
Chicago, IL, United States

Anna L. Mitchell, MBBS, MRes, PhD
Consultant Endocrinologist
Department of Endocrinology
Newcastle upon Tyne NHS Hospitals Foundation Trust
Newcastle upon Tyne, UK

Dimitri S. Monos, PhD
Professor
Pathology and Laboratory Medicine
Perelman School of Medicine
University of Pennsylvania
Children's Hospital of Philadelphia
Philadelphia, PA, United States

Scott N. Mueller, PhD
Professor
Microbiology and Immunology
The University of Melbourne
Melbourne, VIC, Australia

Catharina M. Mulders-Manders, MD, PhD
Infectious Diseases Specialist
Department of Internal Medicine
Infectious Disease Section
Radboudumc Expertise Centre for Immunodeficiency and Autoinflammation
Radboud University Medical Centre
Nijmegen, The Netherlands

Pashna N. Munshi, MD
Associate Clinical Director
Stem Cell Transplant and Immunotherapy Program
Lombardi Comprehensive Cancer Center
Washington, DC, United States

Philip M. Murphy, MD
Chief, Laboratory of Molecular Immunology
National Institute of Allergy and Infectious Diseases
National Institutes of Health
Bethesda, MD, United States

Pierre Noel, MD
Professor of Medicine
Division of Hematology-Oncology
Mayo College of Medicine
Scottsdale, AZ, USA

Luigi D. Notarangelo, MD
Chief, Laboratory of Clinical Immunology and Microbiology
National Institute of Allergy and Infectious Diseases (NIAID)
National Institutes of Health (NIH)
Bethesda, MD, United States

Cristiane J. Nunes-Santos, MD
Postdoctoral Visiting Fellow
Department of Laboratory Medicine
National Institutes of Health Clinical Center
Bethesda, MD, United States

Robert L. Nussbaum, MD
Chief Medical Officer
Medical Affairs
Invitae Corporation;
Volunteer Clinical Faculty
Department of Medicine
University of California San Francisco
San Francisco, CA, United States

Thomas B. Nutman, MD
Chief, Laboratory of Parasitic Diseases
National Institute of Allergy and Infectious Diseases;
Head, Clinical Parasitology Section
Laboratory of Parasitic Diseases
National Institute of Allergy and Infectious Diseases;
Head, Helminth Immunology Section
Laboratory of Parasitic Diseases
National Institute of Allergy and Infectious Diseses
Bethesda, MD, United States

Stephen L. Nutt, BSc, PhD
Laboratory Head, Immunology Division
The Walter and Eliza Hall Institute of Medical Research
Parkville, VIC, Australia

Lorraine O'Neill, MD
The EULAR Centre for Arthritis and Rheumatic Disease
St. Vincent's University Hospital
Elm Park and University College Dublin
Dublin, Ireland

John J. O'Shea, MD
Director, NIAMS Intramural Research Program
National Institutes of Health (NIH)
Bethesda, MD, United States

Thomas L. Ortel, MD, PhD
Professor
Departments of Medicine and Pathology
Duke University Medical Center
Durham, NC, United States

Sung-Yun Pai, MD
Senior Investigator
Immune Deficiency Cellular Therapy Program
Center for Cancer Research, National Cancer Institute
National Institutes of Health
Bethesda, MD, USA

Mary E. Paul, MD
Professor of Pediatrics
Baylor College of Medicine
Houston, TX, United States

Simon H. S. Pearce, MD
Professor of Endocrinology
Newcastle University Medical School
Newcastle upon Tyne, England

Erik J. Peterson, MD
Associate Professor of Medicine
Division of Rheumatic and Autoimmune Diseases
University of Minnesota
Minneapolis, MN, United States

Stefania Pittaluga, MD, PhD
Staff Clinician
Laboratory of Pathology
National Cancer Institute Bethesda
National Institutes of Health
Bethesda, MD, United States

Francesca Polverino, MD, PhD
Associate Professor of Medicine
Pulmonary and Critical Care Medicine
Baylor College of Medicine
Neurosensory Center of Houston
Houston, TX, United States

Jennifer M. Puck, MD
Professor
Department of Pediatrics
University of California San Francisco
San Francisco, CA, United States

Anne Puel, PhD
Professor
Laboratory of Human Genetics of Infectious Diseases
Institut Imagine/University of Paris
Paris, France

Andreas Radbruch, MD, PhD
Scientific Director
Deutsches Rheuma-Forschungszentrum Berlin
A Leibniz Institute;
Professor for Experimental Rheumatology
Charité – Universitätsmedizin Berlin
Berlin, Germany

Raja Rajalingam, PhD
Professor of Clinical Surgery and Director of Immunogenetics and Transplantation Laboratory
Department of Surgery
University of California San Francisco
San Francisco, CA, United States

Stephen T. Reece, PhD
Director, Antimicrobial Antibodies and Vaccines
Kymab Ltd.
Cambridge, UK

John D. Reveille, MD
Professor and Director, Rheumatology
University of Texas Health Science Center at Houston
Houston, TX, United States

Robert R. Rich, MD
Dean and Professor Emeritus
Medicine
University of Alabama at Birmingham
Former Senior Vice President for Medicine
School of Medicine
University of Alabama at Birmingham
Birmingham, AL, United States

Lauren K. Ridley, MD
Department of Internal Medicine
Division of Rheumatology
McGovern Medical School at The University of Texas Health Science Center
Houston, TX, USA

Andrew R. Romeo, MD
Clinical Assistant Professor
Department of Neurology
University of Michigan
Ann Arbor, MI, United States

Cliona M. Rooney, PhD
Professor
Departments of Pediatrics, Pathology, and Molecular Virology and Microbiology
Center for Cell and Gene Therapy
Baylor College of Medicine
Houston, TX, United States

Antony Rosen, MBChB, BSc (Hons)
Mary Betty Stevens Professor of Medicine, Professor of Pathology
Director, Division of Rheumatology;
Department of Medicine
Vice Dean for Research
School of Medicine
Johns Hopkins University
Baltimore, MD, United States

Sergio D. Rosenzweig, MD, PhD
Chief, Immunology Service
Department of Laboratory Medicine Clinical Center
National Institutes of Health
Bethesda, MD, United States

Barry T. Rouse, DVM, PhD, DSc
Distinguished Professor
Biomedical and Diagnostic Sciences
University of Tennessee
Knoxville, TN, United States

Scott D. Rowley, MD
Associate Clinical Director
Stem Cell Transplant and Cellular Immunotherapy Program
Lombardi Comprehensive Cancer Center;
Assistant Professor of Medicine
Georgetown University School of Medicine
Washington, DC, United States

Umit Murat Sahiner, MD
Professor
Pediatric Allergy
Hacettepe University
Ankara, Turkey;
Academic Visitor
Allergy and Clinical Immunology
Imperial College
London, UK

Shimon Sakaguchi, MD, PhD
Laboratory of Experimental Immunology
WPI Immunology Frontier Research Center
Osaka University, Suita
Osaka, Japan

Whitney Salinas, MD
Allergy and Immunology
Park Lane Allergy and Asthma Center
Baylor University Medical Center
Dallas, TX, United States

Marko Salmi, MD, PhD
Professor
MediCity Research Laboratory
University of Turku
Turku, Finland

Sarah W. Satola, PhD
Associate Professor
Medicine/Infectious Diseases
Emory University School of Medicine
Atlanta, GA, United States

Marcos C. Schechter, MD
Assistant Professor
Department of Medicine
Emory University School of Medicine
Atlanta, GA, United States

Enno Schmidt, MD, PhD
Professor and Director
Lübeck Institute of Experimental Dermatology (LIED);
Senior Consultant
Department of Dermatology
University of Lübeck
Lübeck, Germany

Harry W. Schroeder Jr., MD, PhD
Professor
Division of Clinical Immunology and Rheumatology
Departments of Medicine, Microbiology, and Genetics
University of Alabama at Birmingham
Birmingham, AL, United States

Pamela L. Schwartzberg, MD PhD
Senior Investigator
Laboratory of Immune System Biology
National Institutes of Allergy and Infectious Diseases
National Human Genome Research Institute
National Institutes of Health
Bethesda, MD, United States

Giuseppe Sciumè, PhD
Laboratory Affiliated to Istituto Pasteur Italia – Fondazione Cenci-Bolognetti
Department of Molecular Medicine
Sapienza University of Rome
Rome, Italy

Benjamin M. Segal, MD
Chair
Department of Neurology;
Director
Neurological Research Institute;
Professor
Department of Neurology
The Ohio State University
Columbus, OH, United States

Carlo Selmi, MD
Professor
Rheumatology and Clinical Immunology
Humanitas Research Hospital/Humanitas University
Rozzano, Italy

Amir Sharabi, MD, PhD
Senior Rheumatologist
Rheumatology
Rabin Medical Center
Petach-Tikva, Israel;
Assistant Professor
Microbiology and Immunology
Sackler Faculty of Medicine
Tel-Aviv University
Tel-Aviv, Israel

Kristin Ammon Shimano, MD
HS Associate Clinical Professor
Pediatrics
UCSF Benioff Children's Hospital
San Francisco, CA, United States

Patricia M. Sikorski, PhD
Postdoctoral Scientist
Pharmacology and Physiology
George Washington University
Washington, DC, United States

Anna Simon, MD, PhD
Associate Professor
Internal Medicine, Section Infectious Diseases
Radboudumc Expertise Centre for Immunodeficiency and Autoinflammation
Radboud Universitary Medical Centre
Nijmegen, The Netherlands

Gideon P. Smith, MD, PhD, MPH
Vice Chair
Dermatology
Massachusetts General Hospital/Harvard Medical School
Boston, MA, United States

Joo Y. Song, MD
Associate Clinical Professor
Department of Pathology
City of Hope National Medical Center
Duarte, CA, United States

David S. Stephens, MD
Vice President for Research
Robert W. Woodruff Health Sciences Center
Emory University;
Stephen W. Schwarzmann Distinguished Professor of Medicine;
Chair, Department of Medicine
Emory University School of Medicine
Atlanta, GA, United States

Robin Stephens, PhD
Associate Professor
Department of Internal Medicine
Division of Infectious Diseases
University of Texas Medical Branch
Galveston, TX, United States

Michel M. Sun, MD, PhD
Resident
Department of Ophthalmology
University of California Los Angeles
Los Angeles, CA, United States

Benedetta Terziroli Beretta-Piccoli, MD
Associate Professor
Epatocentro Ticino;
Facoltà di Scienze Biomediche
Università della Svizzera Italiana
Lugano, Switzerland;
Institute of Liver Studies
King's College London Faculty of Life Sciences and Medicine at King's College Hospital
London, UK

Wulf Tonnus, MD
Department of Nephrology
University Hospital Carl Gustav Carus
Technische Universität Dresden
Dresden, Germany

Troy R. Torgerson, MD, PhD
Director of Experimental Immunology
Allen Institute for Immunology
Seattle, WA, United States

Raul Martin Torres, MS, PhD
Professor
Immunology and Microbiology
University of Colorado School of Medicine;
Professor
Department of Biomedical Research
National Jewish Health
Aurora, CO, United States

Jennifer D. Treat, PA-C, MSHS, MPH
Physician Assistant
National Institutes of Health
National Institute of Allergy and Infectious Diseases (NIAID)
Bethesda, MD, United States

George C. Tsokos, MD
Chief
Division of Rheumatology and Clinical Immunology
Beth Israel Deaconess Medical Center;
Professor of Medicine
Harvard Medical School
Boston, MA, United States

Gülbü Uzel, MD
Staff Clinician
National Institute of Allergy and Infectious Diseases (NIAID)
National Institutes for Health
Bethesda, MD, United States

Jude E. Uzonna, DVM, PhD
Professor of Immunology
University of Manitoba
Winnipeg, MB, Canada

Jeroen C. H. van der Hilst, MD, PhD
Infectious Disease Specialist
Infectious Diseases and Immunity
Jessa Hospital;
Associate Professor Immune Pathology
Limburg Clinical Research Centre
University of Hasselt
Hasselt, Belgium

Jos W. M. van der Meer, MD, PhD
Professor of Medicine
Department of Internal Medicine
Radboud University Medical Centre
Nijmegen, The Netherlands

John Varga, MD
John and Nancy Hughes Professor
Department of Medicine
Northwestern University Feinberg School of Medicine
Chicago, IL, United States

Meryl Waldman, MD
Senior Research Physician
Immunoregulation Section
Kidney Diseases Branch
National Institute of Diabetes and Digestive and Kidney Diseases
National Institutes of Health
Bethesda, MD, United States

Jill Weatherhead, MD, PhD
Assistant Professor
Medicine-Pediatrics
Infectious Diseases and Tropical Medicine
Baylor College of Medicine's National School of Tropical Medicine
Houston, TX, United States

Peter Weiser, MD
Associate Professor
Department of Pediatrics, Division of Rheumatology
University of Alabama at Birmingham Heersink School of Medicine
Birmingham, AL, United States

Cornelia M. Weyand, MD, PhD
Professor Emerita
Medicine
Stanford University
Stanford, CA, United States;
Professor
Medicine and Immunology
Mayo Clinic Alix School of Medicine
Rochester, MN, United States

Fredrick M. Wigley, MD
Martha McCrory Professor of Medicine
Division of Rheumatology
Johns Hopkins University School of Medicine
Baltimore, MD, United States

James B. Wing, PhD
Laboratory of Experimental Immunology
Immunology Frontier Research Center
Osaka University
Suita, Osaka, Japan

Kathryn J. Wood, DPhil
Professor Emerita of Immunology
Nuffield Department of Social Sciences
University of Oxford
Oxford, UK

Shyra Wilde, PhD
Department of Microbiology and Immunology
Emory University School of Medicine
Atlanta, GA USA

Hui Xu, PhD
Professor
Dermatology
University of Alabama at Birmingham
Birmingham, AL, United States

Nabiha Yusuf, PhD
Associate Professor
Department of Dermatology
University of Alabama at Birmingham
Birmingham, AL, United States

Christa S. Zerbe, MD, MS
Medical Director
Clinical Patient Services Unit
Laboratory of Clinical Immunology and Microbiology
The National Institutes of Health
Bethesda, MD, United States

Qian Zhang, MD
Assistant Professor
Laboratory of Human Genetics of Infectious Diseases
Institut Imagine/University of Paris
Paris, France;
St. Giles Laboratory of Infectious Diseases
The Rockefeller University
New York, NY, USA

Dina Ben-Yehuda, MD
Department of Hematology
Hadassah-Hebrew University Medical Center
Jerusalem, Israel

Shen-Ying Zhang, MD, PhD
Associate Professor
Laboratory of Human Genetics of Infectious Diseases
Institut Imagine/University of Paris
Paris, France;
St. Giles Laboratory of Infectious Diseases
The Rockefeller University
New York, NY, USA

Arthur W. Zieske, MD
Associate Professor of Clinical Pathology and
Genomic Medicine
Houston Methodist Academic Institute
Houston Methodist Hospital
Weill Cornell Medical College
Houston, TX, United States

DEDICATION

To my wife, Susan Rich

Robert R. Rich

To my wife Mary Fleisher, my partner and supporter for >50 years

Thomas A. Fleisher

To Dixie Lee Schroeder; Harry W. Schroeder III, MD, PhD; Maria Isabel, Anabel and William Schroeder; Jeff, Elena, Liam, Noah, Haddie and Ellie Beck; and Jeannette Schroeder

Harry W. Schroeder Jr.

To Jörg Goronzy and Dominic and Isabel Weyand Goronzy

Cornelia M. Weyand

To my immunology mentors: Mary, Rich, and Farrah

David B. Corry

To Bob Nussbaum, my husband and best friend

Jennifer M. Puck

1

The Human Immune Response

Robert R. Rich and Randy Q. Cron

Clinical immunology is a medical subspecialty largely focused on a specific physiologic process, inflammation, which is essential to good health, particularly in defense against pathogenic organisms, recovery from injury, and containment of neoplasms. However, inflammation, which is mediated by the cells and soluble products of the immune system, is also a powerful contributor to the pathogenesis of diseases that affect virtually every organ system. A consequent challenge for clinical immunologists, both clinicians and basic scientists, is to reduce a dizzying array of disease descriptions to a systematic understanding of pathogenic mechanisms in order to facilitate translation of fundamental concepts and new discoveries into more effective disease prevention or treatment.

This introductory chapter is directed to nonimmunologist clinicians and researchers. It is structured as an introduction to the interacting elements of the human immune system and their disordered functions in diseases. The subtleties, including immunologic or molecular genetic jargon unavoidably used, are described in detail in the chapters that follow.

THE HOST-MICROBE INTERACTION

The vertebrate immune system is a product of eons of evolutionary relationships between rapidly evolving microbial organisms and their much less rapidly reproducing, and hence less adaptable, hosts.[1] In general, the relationship is mutually beneficial, each providing nutrients and other materials essential to the well-being of their partner—the host and its microbiome (Chapter 22). Occasionally, however, a normally beneficial relationship becomes pathologic. Pathogenic microbes can overwhelm the microbiome, invade host tissues, and result in host morbidity or even death. Because the vertebrate host cannot win a battle with microbial invaders by rapid mutation and selection, the immune system uses a strategy of complexity and redundancy, which involves both the individual organism and its collective population.

Reflecting plasticity of the response, specific defenses differ, depending on the nature of the infectious agent and its point of entry and distribution within the body. Regardless of the defense mechanism, an intended outcome is destruction or neutralization of the invading organism. However, a secondary consequence can be collateral damage to host cells. These unfortunate cells can be targeted for damage because they are sites of microbial residence and replication, or they can be damaged as "innocent bystanders." Depending on the site and severity of the host's defensive response, it may be accompanied by local and/or systemic symptoms and signs of inflammation, which may lead to long-lasting tissue dysfunction as a result of tissue remodeling and partial repair.

Adaptive and Innate Immunity

Immune responses are traditionally classified as adaptive (also termed *acquired* or *specific*) and innate (or *nonspecific*) (Table 1.1). The adaptive immune system, present uniquely in species of the phylum Chordata, is specialized for development of an inflammatory response based on recognition of specific "foreign" macromolecules that are predominantly, but not exclusively, proteins, peptides, and carbohydrates. The vast majority of chordate species are jawed vertebrates, and this book addresses adaptive immunity of that subphylum. Its primary effectors are antibodies, B lymphocytes, T lymphocytes, innate lymphoid cells (ILCs), and antigen-presenting cells (APCs). T and B lymphocytes express surface antigen receptors that are clonally specific as a consequence of receptor-gene rearrangements. Expansion of clones of lymphocytes specific for any particular antigen is induced by antigen encounter and consequent activation and proliferation, thereby constituting the basis of immunologic memory.

Innate immune responses are phylogenetically far more ancient, being widely represented in multicellular phyla.[2] Rather than being based on exquisitely specific recognition of a diverse array of macromolecules (i.e., antigens), they are focused on recognition of common molecular signatures of microbial organisms that are not present in vertebrates (Chapter 3).[3] Among these structures, which are termed *pathogen-associated molecular patterns* (PAMPs) or *danger-associated molecular patterns* (DAMPs), are bacterial cell wall constituents, such as mannose-rich oligosaccharides, lipopolysaccharides, peptidoglycans, and several nucleic acid variants, including double-stranded RNA and unmethylated CpG DNA. For both innate and adaptive immune responses, defense effector mechanisms can require either direct cell-to-cell contact or the activity of cytokines (Chapter 14) and chemokines (Chapter 15), which are hormone-like soluble molecules that act in the cellular microenvironment (cell-mediated immunity). Most immune responses include participation of both modes of response.[2–4]

The elements of innate immunity are diverse. They include physical barriers to pathogen invasion (e.g., skin, mucous membranes, cilia, and mucus), as well as an array of cellular and soluble factors that can be activated by secreted or cell surface products of the pathogen, including PAMPs. Recognition of PAMPs by cells in innate immunity, which also commonly function as APCs to the lymphocytes of adaptive immunity, is via cell membrane or cytoplasmic receptors known as pattern recognition receptors (PRRs). PRRs can be either membrane bound or cytoplasmic. Membrane-bound PRRs include Toll-like receptors (TLRs) and C-type lectin receptors (CLRs). Humans express 10 distinct TLRs, which recognize (among others) specific bacterial glycolipids, lipopolysaccharide; viral single-stranded

TABLE 1.1 Features of Innate and Adaptive Immune Systems

Distinguishing Features	
Innate Immunity	**Adaptive Immunity**
Germline-encoded receptors targeting pathogen molecular patterns	Clonally variable receptors generated somatically by rearrangement of gene elements
Does not require immunization	Consequence of B- and/or T-cell activation
Limited memory	Immunologic memory well developed
Includes physical barriers to pathogen	Antibody and cytotoxic T cells

Common Features

Cytokines and chemokines
Complement cascade
Phagocytic cells
Natural killer (NK) cells
"Natural" antibodies

RNA; and bacterial and viral unmethylated CpG DNA. CLRs are particularly important in antifungal innate immunity but also have important roles in defenses against bacteria, viruses, and parasites. They comprise a large family that commonly recognizes microbe-specific carbohydrate ligands or structurally similar lectin-like domains. Cytoplasmic PRRs include RIG-1–like receptors (RLRs) and nucleotide oligomerization domain (NOD)-like receptors (NLRs). RLRs are involved in recognition of viruses through interaction with intracytoplasmic viral double-stranded RNA (dsRNA), and NLRs recognize bacterial peptidoglycan motifs.[4]

Cells of the innate immune system are commonly triggered through activation of the nuclear factor-κB (NF-κB) transcription factor via the MyD88 signaling pathway, thereby inducing an inflammatory response using mechanisms that are broadly shared with those of the adaptive immune system. These include activation of various types of ILCs (e.g., natural killer [NK] cells [Chapter 12]), which are characterized by absence of clonally expressed receptors for specific antigen (see later), activation of granulocytes and other phagocytes (Chapter 39), the secretion of inflammatory cytokines and chemokines, and interactions of the many participants in the complement cascade (Chapter 40). In addition, activation of cells of innate immunity that also act as APCs for the adaptive immune system results in upregulation of membrane molecules such as CD80 and CD86 (among others) that provide the second signal, which along with the T-cell receptor (TCR) for antigen (Chapter 4), can activate antigen-specific T cells (Chapter 10).[5]

Finally, because recognition of pathogens by the innate immune system relies on germline encoded, non-rearranged receptors held in common by the specific cell type, innate immunity is more rapidly responsive. It can initiate in minutes to hours and generally precedes development of a primary adaptive immune response by at least several days.

CELLS OF THE IMMUNE SYSTEM

The major cellular constituents of both innate and adaptive immunity originate in bone marrow, where they differentiate from multipotent hematopoietic stem cells (HSCs) along several pathways to become granulocytes, lymphocytes, and APCs (Chapter 2).

Granulocytes

Polymorphonuclear leukocytes (granulocytes) are classified by light microscopy into four types. By far the most abundant in the peripheral circulation are neutrophils, which are principal effector cells linking the innate and adaptive responses by virtue of their expression of surface receptors for antibody and complement (Chapter 40). They are phagocytic cells that ingest, kill, and degrade microbes and other targets of an immune attack within specialized cytoplasmic vacuoles that contain potent antimicrobial enzymes and oxidative pathways. The phagocytic activity of neutrophils is promoted by their surface display of receptors for antibody molecules (specifically the Fc portion of immunoglobulin G [IgG] molecules) (Chapter 8) and activated complement proteins (particularly the C3b component) (Chapter 40). Neutrophils are the predominant cell type in acute inflammatory infiltrates and are the primary effector cells in immune responses to pyogenic bacteria (Chapter 27).

Eosinophils (Chapter 45) and basophils (Chapter 44) are the other circulating forms of granulocytes. A close relative of the basophil, but derived from distinct bone marrow precursors, is the tissue mast cell, which does not circulate in blood. Eosinophils, basophils, and mast cells are important in defenses against multicellular pathogens, particularly helminths (Chapter 30). Their defensive functions are not based on phagocytic capabilities but on their ability to discharge potent biologic mediators from their storage granules into the cellular microenvironment. This process, termed *degranulation*, can be triggered by antigen-specific IgE molecules that bind to basophils and mast cells via high-affinity receptors for the Fc portion of IgE (FcεR) on their surfaces. In addition to providing a mechanism for anthelmintic host defenses and certain antibacterial responses, this is also the principal mechanism involved in acute (IgE-mediated) allergic reactions (Chapters 43–50).

Lymphocytes

Three broad categories of lymphocytes are identified on the basis of display of particular surface molecules: B cells, T cells, and ILCs. Each of these categories can be further subdivided according to specific function and display of distinguishing cell surface molecules (Chapter 2). All lymphocytes differentiate from common lymphoid stem cells in bone marrow. B cells create their Ig receptors in bone marrow and differentiate into antibody-producing cells in the periphery (Chapter 7). T-cell precursors move from bone marrow to the thymus (or, in some cases, to extrathymic tissue compartments), where they complete their differentiation and selection (Chapter 9).

T cells and B cells are the heart of immune recognition, a property reflecting their clonally specific cell surface receptors for antigen (Chapter 4). The TCR is a heterodimeric integral membrane molecule expressed exclusively by T lymphocytes. B-cell receptors (BCRs) for antigen are membrane immunoglobulin (mIg) molecules of the same antigenic specificity that the B cell and its terminally differentiated progeny, plasma cells, will secrete as soluble antibodies. Memory B cells and nondividing, long-lived plasma cells may account substantially for persistence of antibody responses (including production of autoantibodies) over many years.[6]

Receptors for "antigen" on the third class of lymphocytes, ILCs, are not clonally expressed. ILCs are subdivided into three major groups according to the cytokines that they produce. For example, group 1 ILCs, including NK cells, produce interferon-γ (IFN-γ) and tumor necrosis factor (TNF).[7] ILCs express receptors

for PAMPs and, as such, serve as major effectors of innate immunity. They also recognize target cells that might otherwise elude the immune system (Chapters 2 and 12). Thus recognition of NK cell targets is based substantially on what their targets lack rather than on what they express.

NK cells express receptors of several types for major histocompatibility complex (MHC) class I molecules via killer immunoglobulin-like receptors (KIRs).[8] KIRs are expressed on the plasma membrane of NK cells (and some T cells), which interact with class I molecules to alter NK-cell cytotoxic function. Most KIRs express in their intracellular domain a tyrosine-based inhibitory motif (ITIM) that suppresses NK activity, thereby preventing NK cell activity directed against normal self-cells. In contrast, some KIRs express a tyrosine-based activation motif (ITAM), which amplifies their activity. NK cells will kill target cells unless they receive an inhibitory signal transmitted by an ITIM receptor. Virus-infected cells and tumor cells that attempt to escape T-cell recognition by downregulating their expression of class I molecules become susceptible to NK cell–mediated killing because the NK cells receive an activation signal and/or fail to receive an inhibitory signal through the ITAM- and ITIM-containing MHC class I receptors. The balance between ITIM and ITAM is regulated by the microenvironmental milieu, increasing expression of ITAM in the presence of virus-infected or cancer cells and of ITIM as necessary to maintain self-tolerance and prevent autoimmunity. A high frequency of ITAM-expressing cells has been reported in some patients with autoimmune diseases.[9]

Although NK cell–mediated innate immunity has been long considered to lack immunologic memory, studies suggest that NK cells can exhibit memory of previous encounters with microbes or other antigens, the molecular basis of which remains to be fully elucidated.[10] NK cells can also participate in antigen-specific immune responses by virtue of their surface display of the activating ITAM receptor CD16, which binds the constant (Fc) region of IgG molecules. This enables them to function as effectors of a cytolytic process termed *antibody-dependent cellular cytotoxicity* (ADCC), a mechanism exploited clinically with monoclonal antibody (mAb) therapeutic agents.[11]

In general, pathways leading to differentiation of T cells, B cells, and ILCs are mutually exclusive, representing a permanent lineage commitment. No lymphocytes express both mIg and TCRs. However, a subset of T cells, termed *NKT cells*, exhibit both NK-like cytotoxicity and αβ TCR with limited receptor diversity.

Antigen-Presenting Cells

KEY CONCEPTS

Features of Antigen-Presenting Cells

- Capacity for uptake and partial degradation of protein antigens
- Expression of major histocompatibility complex (MHC) molecules for binding antigenic peptides
- Chemokine receptors to allow colocalization with T cells
- Expression of accessory molecules for interaction with T cells
- Receptors for pathogen- or danger-associated molecular patterns
- Secretion of cytokines that program T helper (Th) cell responses

A morphologically and functionally diverse group of cells, all of which are derived from bone marrow precursors, is specialized for presentation of antigen to lymphocytes, particularly T cells (Chapter 6). Included among such cells are dendritic cells (DCs), monocytes (present in the peripheral circulation), macrophages (solid tissue derivatives of monocytes), cutaneous Langerhans cells (Chapter 23), and constituents of the reticular endothelial system within solid organs. B lymphocytes that specifically capture antigen via their clonally expressed mIg can also function efficiently in antigen presentation to T cells.

Cardinal features of APCs include their expression of both class I and class II MHC (Chapter 5) molecules as well as requisite accessory molecules for T-cell activation (e.g., B7-1, B7-2/CD80, CD86). Upon activation, APC elaborate cytokines that induce specific responses in cells to which they are presenting antigen. In addition to processing and presenting antigen, APCs can regulate activation of the immune system via innate cell surface receptors, which contribute to determination of whether the antigen is pathogen associated.

APCs differ substantially among themselves with respect to mechanisms of antigen uptake and effector functions. Immature DCs show high phagocytic and pathogen-killing activity but low ability to present antigen and activate T cells. DCs that have ingested a pathogen or foreign antigen can be induced to mature by inflammatory stimuli,[12,13] especially via cells of the innate immune system and by direct activation through receptors for PAMPs or DAMPs. Monocytes and macrophages are actively phagocytic, particularly for antibody and/or complement-coated (opsonized) antigens that bind to their surface receptors for IgG and C3b. These cells are also important effectors of immune responses, especially in sites of chronic inflammation. Upon further activation by T-cell cytokines, they can kill ingested microorganisms by oxidative pathways similar to those used by polymorphonuclear leukocytes.

The interaction between B cells acting as APCs and T lymphocytes is notable because the cells are involved in a mutually amplifying circuitry of antigen presentation and response. The process is initiated by antigen capture through B-cell mIg and ingestion by receptor-mediated endocytosis. This is followed by proteolytic antigen degradation and then display to T cells as oligopeptides bound to MHC molecules. Like other APCs, B cells display CD80 and thus provide a requisite second signal to the antigen-responsive T cell via CD28, its accessory molecule for activation (Fig. 1.1; Chapters 4 and 10). As a result of T-cell activation, T-cell cytokines that regulate B-cell differentiation and antibody production are produced. T cells are stimulated to display the surface ligand CD40L (CD154), which can serve as the second signal for B-cell activation through its inducible surface receptor.

BASIS OF ADAPTIVE IMMUNITY

The essence of adaptive immunity is molecular distinction between self constituents and potential pathogens. For simplicity, this is typically summarized as self/nonself discrimination. However, more precisely the issue is one of discrimination between molecular species that are perceived as signaling potential "danger" versus those that are not. This discrimination is a major responsibility of both T cells, B cells, and cells of the innate immune system. This process is more complicated for T cells because it reflects the selection of thymocytes that have generated specific antigen receptors that can bind to self-antigens below a certain threshold of activation and then, upon later encounter, can bind nonself antigenic peptides bound to self-MHC molecules and be activated to engage in effector function.

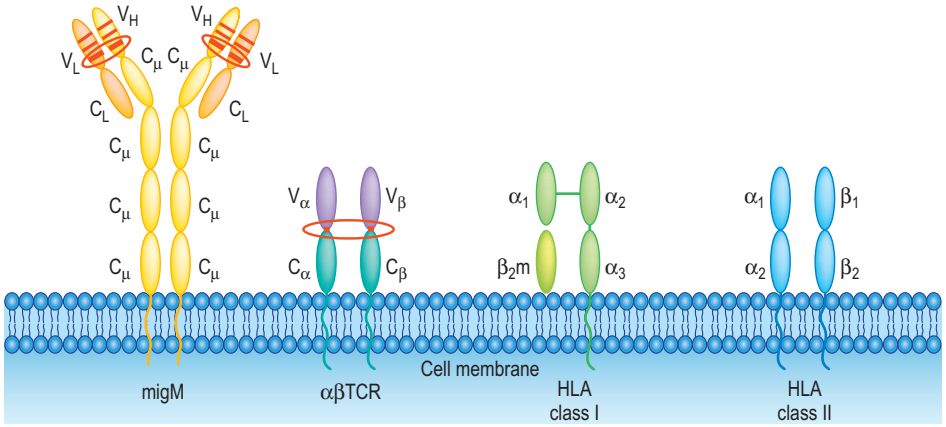


FIG. 1.1 Antigen-Binding Molecules. Antigen-binding pockets of immunoglobulin (Ig) and T-cell receptor (TCR) comprise variable *(V) segments of two chains translated from transcripts that represent rearranged V(D)J or VJ gene elements. Thin red bars* designate two of the complementarity determining regions (CDRs) that form portions of the Ig antigen-binding site. The *red ovals with thick red bars designate the regions of very high sequence variability in both Igs and TCRs that are generated by r*ecombination of the 3′-end of the V gene element with the D and J gene elements or with the J gene element. In the Ig molecule this is designated CDR3. Antigen-binding pockets of Ig molecules are formed by the three-dimensional folding of the heavy and light chains that juxtapose the CDRs of one heavy chain and one light chain. Antigen-binding grooves of MHC molecules are formed with contributions from α_1 and β_1 domains of class II molecules and from α_1 and α_2 domains of class I molecules. All of these molecules are members of the immunoglobulin superfamily. *β_2m*, β_2-Microglobulin; *C*, constant-region domain; *HLA*, human leukocyte antigen; *MHC*, major histocompatibility complex; *mIgM*, membrane immunoglobulin M.

More subtly, the consequence of this selection process is that foreign proteins are recognized as antigens in terms of their ability to initiate an active immune response, whereas self-proteins are tolerated (i.e., are not perceived as antigens). B cells that express self-reactive antibodies are subjected to negative selection in the bone marrow and the periphery. Through PAMPs/DAMPs and other, still undefined, mechanisms, the cells of innate immunity contribute to the essential distinction between commensal (not dangerous) and potentially pathogenic (dangerous) microbes.

T lymphocytes generally recognize antigens as a complex of short linear peptides bound to self-MHC molecules on the surfaces of APCs (Chapter 6). The source of these peptides can be either extracellular or intracellular proteins, and they can be derived from either self or foreign (e.g., microbial) molecules. With the exception of superantigens (SAgs; see later), T cells neither bind antigen in native conformation nor recognize free antigen in solution. The vast majority of antigens for T cells are oligopeptides. However, the antigen receptors of NKT cells can recognize lipid and glycolipid antigens that are presented to them by MHC-like CD1 molecules.[14]

Antigen recognition by T cells differs fundamentally from that by antibodies, which are produced by B lymphocytes and their derivatives. Antibodies are oriented toward recognition of extracellular threats and, unlike T cells, can bind complex macromolecules in their native conformation at cell surfaces or in solution. Moreover, antibodies show less preference for recognition of proteins; antibodies against carbohydrates, nucleic acids, lipids, and simple chemical moieties can be readily produced. Although B cells can also be rendered unresponsive by exposure to self-antigens, particularly during differentiation in bone marrow, this process does not define foreignness within the context of self-MHC recognition.

Clonal Basis of Immunologic Memory

An essential element of self/nonself discrimination is the clonal nature of antigen recognition. Although the immune system can recognize a vast array of distinct antigens, all of the receptors of a single T cell or B cell (and their clonal progeny) have identical antigen-binding sites and hence a particular specificity (Chapter 4). A direct consequence is the capacity for antigen-driven immunologic memory. This phenomenon derives from the fact that, after an initial encounter with antigen, clones of lymphocytes that can recognize the antigen proliferate and differentiate into effector cells. After interaction with their target, most of these effector cells are consumed or undergo programed cell death. However, a smaller population of long-lived memory cells persists. These memory cells constitute a pool of cells larger than the initial naïve responders. They can elicit a greater and more rapid response upon subsequent antigen encounter. These two hallmarks of adaptive immunity, clonal specificity and immunologic memory, provide a conceptual foundation for the use of vaccines in prevention of infectious diseases (Chapter 87).

Immunologic memory involves not only the T cells charged with antigen recognition but also the T cells and B cells that mediate the efferent limb of an inflammatory response. In its attack on foreign targets, the immune system can exhibit exquisite specificity for the inducing antigen, as is seen in the epitope-specific lysis of virus-infected target cells by cytolytic T cells.

ANTIGEN-BINDING MOLECULES

KEY CONCEPTS

Features of the Immunoglobulin (Ig) Superfamily

- Large family of ancestrally related genes (more than 100 members)
- Most products involved in immune system function or other cell-cell interactions
- Ig superfamily members have one or more domains of ~100 amino acids, each usually translated from a single exon
- Each Ig domain consists of a pair of β-pleated sheets usually held together by an intrachain disulfide bond

Three sets of molecules are responsible for the specificity of adaptive immune responses by virtue of their capacity to bind foreign antigen. These are Igs, TCRs, and MHC molecules (see Fig. 1.1; Chapters 4 and 5). All are products of a very large family of ancestrally related genes, the Ig superfamily, which includes many other molecules essential to induction and regulation of immune responses.[15,16] Members of the Ig superfamily exhibit characteristic structural features. The most notable of these is organization into homologous domains of approximately 110 amino acids that are usually encoded by a single exon with an intradomain disulfide bond. These domains are characteristically configured as antiparallel strands, forming two opposing β-pleated sheets.

Immunoglobulins and T-Cell Receptors

The remarkable specificity of Ig and TCR molecules for antigen is achieved by a mechanism of genetic recombination that is unique to Ig and TCR genes (Chapter 4). The antigen-binding site of both types of molecules lies at the tip of the two juxtaposed, constituent polypeptides and contains contributions from each of the two. In the case of Igs, these are a heavy (H) chain and one of two alternative types of light (L) chains, κ or λ. In the case of TCRs, either of two alternative heterodimers can constitute the antigen-binding molecule, one composed of α and β chains, and the other of λ and δ chains. In the case of the TCR, this antigen-binding site is relatively flat, permitting association between the TCR and the MHC:peptide complex. In the case of Ig, the antigen-binding site can be concave, flat, or form a projection. This permits the Ig to bind to a variety of surface structures, including projections and nooks and crannies.

The polypeptides contributing to both Igs and TCRs can be divided into an antigen-binding amino-terminal variable (V) domain and one or more carboxy-terminal constant (i.e., nonvariable) domains. Ig constant region domains generally include specific sites responsible for the biologic effector functions of the antibody molecule (Chapter 8).

KEY CONCEPTS

Comparison of T- and B-Cell Receptors for Antigen

Similarities

- Members of the immunoglobulin (Ig) superfamily
- Each chain divided into variable and constant regions
- Variable regions constructed by V(D)J rearrangements
- Nongenomic N-nucleotide additions at V(D)J junctions
- Both polypeptide chains contribute to the antigen-binding site
- Exhibit allelic exclusion
- Negative selection against receptors with self-antigen specificity
- Transmembrane signaling involving coreceptor molecules

Differences

- Ig can be secreted; T-cell receptor (TCR) is not
- Ig recognizes conformational antigen (Ag) determinants; TCR recognizes linear determinants
- Ig can bind antigen in solution; TCR binds antigen when presented by a major histocompatibility complex (MHC) molecule on an antigen-presenting cell (APC)
- TCRs are positively selected for self-MHC recognition
- Somatic hypermutation of Ig genes can enhance antigen-binding affinity
- Ig genes can undergo isotype switching
- Ig constant domains express inflammatory effector functions

The most noteworthy feature of the jawed vertebrate immune system is the process of recombination-activating gene (RAG) mediated genetic recombination that generates a virtually limitless array of specific antigen receptors from a rather limited genomic investment. This phenomenon is accomplished by the recombination of genomic segments that encode the variable domains of Ig and TCR polypeptides (Chapter 4).[17] The products of these rearranged gene elements provide a specific B or T cell with its unique antigen receptor. The variable domain of the mature receptor is created by the rearrangement of two or three separate gene segments. These are designated V (variable) and J (joining), for IgL chains and TCR α and γ chains, and V, D (diversity) and J, for IgH and TCR β and δ chains. In addition to rearrangement, N-nucleotide addition also contributes substantially to receptor diversity. N-nucleotide addition results in the insertion, at the time of rearrangement, of one or more nongenomic nucleotides at the junctions between V, D, and J segments through the action of terminal deoxynucleotidyl transferase (TdT).[17] This permits receptor diversity to extend beyond germline constraints. Analysis of the linear sequences of many Ig V region domains has shown that they contain three sites of high sequence variability that have been designated *complementarity determining regions 1–3 (CDR1–3)* to indicate that they are the sites that contact antigen (see Fig. 1.1).

DNA rearrangement involved in generating TCRs and BCRs is controlled by recombinases that are active in early thymocytes and in B precursor cells in bone marrow. The process is sequential and carefully regulated, generally leading to translation of one receptor of unique specificity for any given T or B lymphocyte. This result is achieved through a process termed *allelic exclusion*, wherein only one member of a pair of allelic genes potentially contributing to an Ig or TCR molecule is rearranged at a time.[18]

The process of allelic exclusion is not absolute, and a small number of lymphocytes will express dual functional Ig or TCR transcripts and, in some cases, two distinct surface receptors. However, B cells exclusively rearrange Ig genes, not TCR genes, and vice-versa for T cells. Moreover, after producing a functional heavy chain, B cells sequentially rearrange L chain genes, typically κ before λ. Thus, in normal individuals, the vast majority of B cells express either κ or λ chains, with 1% or less expressing both. Similarly, thymocytes express α and β genes, or γ and δ genes.

There is one feature of V region construction that is essentially reserved to B cells. This is somatic hypermutation (SHM), a process that can continue at discrete times throughout the life of a mature B cell at both the $V_HD_HJ_H$ and V_LJ_L gene exons.[19,20] Because these rearranged gene exons encode the antigen-binding site that contains the specific points of contact with antigen, on occasion the random process of SHM will result in cells expressing mIg with increased affinity for the antigen they recognize. Typically, cells with increased affinity for antigen are activated preferentially, particularly at limiting doses of antigen. Thus the average affinity of antibodies produced during the course of an immune response tends to increase, a process termed *affinity maturation*.

TCRs do not show evidence of SHM. This absence may be related to the focus on selection in the thymus involving corecognition of a self-MHC molecule and self-peptides,[21] (Chapter 9) rather than the continuous process of antigen-driven selection in the periphery by B cells after SHM. Thymic selection results in deletion by apoptosis of the vast majority of differentiating

thymocytes by mechanisms that place stringent boundaries around the viability of a thymocyte with a newly expressed TCR specificity. Once a T cell is fully mature and ready for emigration from the thymus, its TCR is essentially fixed, reducing the likelihood of emergent autoimmune T-cell clones in the periphery.

Receptor Selection

The receptor expressed by a developing thymocyte must be capable of binding with low-level affinity to some particular MHC self-molecule, either class I or II, expressed by a resident thymic epithelial cell or APC. Because their receptors are generated by a process of semirandom joining of rearranging exon segments coupled with N-nucleotide additions, most thymocytes fail this test. They are consequently deleted as not being useful to an immune system that requires T cells to recognize antigen that is bound to self-MHC molecules. Thymocytes surviving this hurdle are said to have been "positively selected" (Fig. 1.2A).[21] Conversely, a small number of thymocytes bind with an unallowably high affinity for a combination of MHC molecule plus antigenic peptide expressed by a thymic APC. Because the peptides available for MHC binding at this site are derived almost entirely from self-proteins, differentiating thymocytes with such receptors are intrinsically dangerous as potentially autoimmune. This deletion of thymocytes with high-affinity receptors for self-MHC plus (presumptively) self-peptide is termed "negative selection" (see Fig. 1.2B),[21] a process that may also involve activity of regulatory T cells (Tregs; Chapter 13).[22,23]

Another feature that distinguishes B cells from T cells is that the cell surface antigen receptors of the former can be secreted in large quantities as antibody molecules, the effector functions of which are carried out in solution or at the surfaces of other cells. Secretion is accomplished by alternative splicing of Ig messenger RNAs (mRNAs) to include or exclude a transmembrane segment that is encoded by the Ig heavy-chain genes.

Immunoglobulin Class Switching

In addition to synthesizing both membrane and secreted forms of Igs, B cells also undergo class switching. Antibody molecules are composed of five major classes (isotypes). In order of abundance in serum, these are IgG, IgM, IgA, IgD, and IgE (Chapters 4 and 8). In humans the IgG class is further subdivided into four subclasses and the IgA class into two subclasses. The class of Ig is determined by the sequence of the constant region of its heavy chain (C_H). The H-chain constant region gene locus is organized with exons that encode each of the Ig isotypes and subclasses located downstream (3′) of the variable (V_H) genes. Thus an antibody-producing cell with a successfully rearranged $V_HD_HJ_H$ exon can change the class of antibody molecule that it

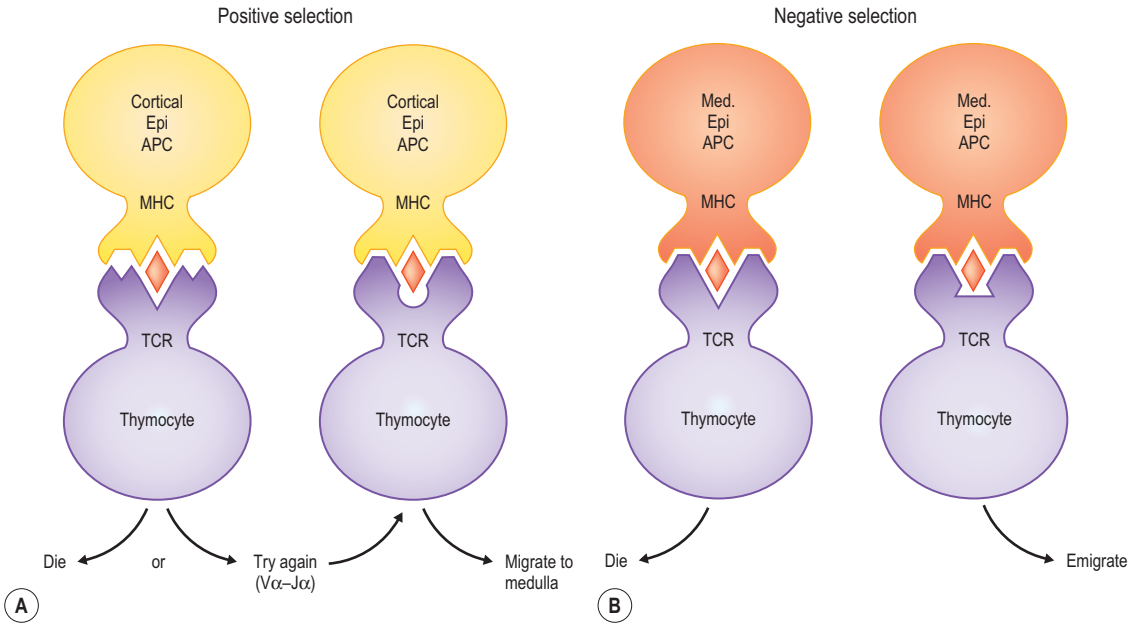


FIG. 1.2 Two-Stage Selection of Thymocytes Based on Binding Characteristics of Randomly Generated T-Cell Receptors. (A) *Positive selection.* "Double-positive" ($CD4^+$, $CD8^+$) thymocytes with T-cell receptors *(TCRs)* capable of *low* avidity binding to some specific self-MHC molecule (either class I or II) expressed by thymic cortical epithelial cells *(Epi)* are positively selected. This process may involve sequential attempts at α gene rearrangement to express an αβ TCR of appropriate self-MHC specificity. If binding is to a class I molecule, the positively selected thymocyte becomes CD8 single-positive, and if to a class II molecule, a CD4 single-positive. Thymocytes that are unsuccessful in achieving a receptor with avidity for either a class I or II self-MHC molecule die by apoptosis. The *solid diamond* represents a self-peptide derived from hydrolysis of an autologous protein present in the thymic microenvironment or synthesized within the thymic epithelium itself. (B) *Negative selection.* "Single-positive" ($CD4^+$ *or* $CD8^+$) thymocytes, positively selected in the thymic cortex, that display TCRs with *high* avidity for the combination of self-MHC plus some self (autologous) peptide present in the thymus are negatively selected (i.e., die) as potentially "autoimmune." Those few thymocytes that have survived both positive and negative selection emigrate to the periphery as mature T cells. *APC*, Antigen-presenting cell; *MHC*, major histocompatibility complex.

synthesizes by utilization of different C_H genes without changing its unique antibody specificity. This process, termed *class switch recombination*, is regulated by cytokines and is accomplished through the action of activation-induced cytidine deaminase.[24]

There is no process comparable to class switch recombination in T cells. The two types of TCRs are products of four independent sets of V-region and C-region genes. A large majority of peripheral blood T cells express αβ TCRs, with a small fraction expressing γδ TCRs (usually ≤5% in peripheral blood). There is a higher representation of γδ T cells in certain tissues, particularly those lining mucous membranes, where they may be specialized for recognition of heavily glycosylated peptides or nonpeptide antigens that are commonly encountered in these tissue compartments. Thymocytes are committed to the expression of either αβ or γδ TCR, and their differentiated progeny (T cells) never change their TCR type in the periphery.

Major Histocompatibility Complex

MHC molecules constitute a third class of antigen-binding molecules. When an MHC class I molecule was initially crystallized, an unknown peptide was found in a binding groove formed by the first two (α_1 and α_2) domains of the molecule. This binding groove has since been established as a general feature of MHC molecules.[25] It is now known that the function of MHC molecules is to present antigen to T cells in the form of oligopeptides that reside within this antigen-binding groove (Chapter 6). The most important difference between the nature of the binding groove of MHC molecules and those of Ig and TCR is that the former does not represent a consequence of gene rearrangement. Rather, all the available MHC molecules in an individual are encoded in a linked array, which in humans is located on chromosome 6 and designated the human leukocyte antigen (HLA) complex.

MHC molecules are of two basic types, class I and class II. Class I molecules are found on the surface of almost all somatic cells, whereas cell surface expression of class II genes is restricted primarily to cells specialized for APC function. Class I molecules have a single heavy chain that is an integral membrane protein composed of three external domains (see Fig. 1.1). The heavy chain is noncovalently associated with β_2-microglobulin, a nonpolymorphic, non–membrane-bound, single-domain Ig superfamily molecule that is encoded in humans on chromosome 15, not linked to the MHC. In contrast, class II MHC molecules comprise two polypeptide chains, α and β (or A and B), of approximately equal size, each of which consists of two external domains connected to a transmembrane region and cytoplasmic tail. Both chains of class II molecules are anchored on the cell by a transmembrane domain, and both are encoded within the MHC. Class I and class II molecules have a high degree of structural homology, and both fold to form a peptide-binding groove on their exterior face, with contribution from the α_1 and α_2 domains for class I molecules and from α_1 and β_1 domains for class II.

There are three class I loci (HLA-A, -B, and -C) and three class II subregions (HLA-DR, -DQ, and -DP) that are principally involved in antigen presentation to T cells (Chapter 5). The functions of other class I and class II genes within this complex are less clear. Some, at least, appear to be specialized for binding (presentation) of peptide antigens of restricted type, source, or function (e.g., HLA-E),[26] and others (e.g., HLA-DM and HLA-DO) appear to be involved in selective antigen processing and loading of antigenic peptides into the binding cleft of the HLA-DR, -DQ, and -DP molecules (Chapter 6).[27] In addition, members of a family of "nonclassic" class Ib molecules, $CD1_{a–d}$, which are encoded on chromosome 1 outside the MHC, are specialized for binding and presentation of lipid and lipid-conjugate antigens to T cells.[14,28]

The HLA complex represents an exceedingly polymorphic set of genes (Chapter 5). Consequently, most individuals are heterozygous at each major locus. In contrast to TCRs and Igs, the genes of the MHC are co-dominantly expressed. Thus, at a minimum, an APC can express six class I molecules and six class II molecules (the products of the two alternative alleles of three class I and three class II loci). This number is, in fact, usually an underestimate, as a consequence of additional complexity in the organization of the class II region.

ANTIGEN PRESENTATION

Because MHC genes do not undergo recombination, the number of distinct antigen-binding grooves that they can form is many orders of magnitude less than that for either TCRs or Igs. Oligopeptides that bind to MHC molecules are the products of self or foreign proteins. They are derived by hydrolytic cleavage within APCs and are loaded into MHC molecules before expression at the cell surface (Chapter 6). Indeed, stability of MHC molecules at the cell surface requires the presence of a peptide in the antigen-binding groove; cells mutant for the loading of peptide fragments into MHC molecules fail to express MHC molecules on their cell surfaces.[29] Because in the absence of infection most hydrolyzed proteins are of self-origin, the binding groove of most MHC molecules contains a self-peptide.

Class I and class II molecules differ from one another in the length of peptides that they bind, usually 8 to 9 amino acids for class I and 14 to 22 amino acids for class II. Although important exceptions are clearly demonstrable, they also generally differ with respect to the source of peptide. Those peptides binding to class I molecules usually derive from proteins synthesized intracellularly (e.g., autologous proteins, tumor antigens, virus proteins, and proteins from other intracellular microbes), whereas class II molecules commonly bind peptides derived from proteins synthesized extracellularly (e.g., extracellular bacteria, nonreplicating vaccines, toxins/allergens). Endogenous peptides are generated by the immunoproteasome and then are loaded into newly synthesized class I molecules in the endoplasmic reticulum following active transport from the cytosol. In contrast, proteolytic breakdown and loading of exogenous peptides into class II molecules occurs in acidic endosomal vacuoles. As a consequence of proteolytic processing and binding into an MHC molecule, T cells see linear peptide epitopes. In contrast, because B-cell antigen recognition requires neither proteolytic processing nor binding into an MHC molecule, B cells recognize native, three-dimensional epitopes.

In addition to the recognition of lipids and lipid-conjugates presented by CD1 molecules, there are other exceptions to the generalization that MHC molecules only present (and T cells only recognize) oligopeptides. It has been known for many years that T cells can recognize haptens, presumably covalently or noncovalently complexed with peptides residing in the antigen-binding groove. This phenomenon is familiar to physicians as contact dermatitis to nonpeptide antigens, such as urushiol (from poison ivy) and nickel ion (Chapter 48). In addition, a newly recognized subset of T cells designated mucosal-associated (semi-) invariant T (MAIT) cells recognize vitamin B_2

(riboflavin) and vitamin B_9 (folate) derivatives bound to MR1, a nonpolymorphic MHC class I–like molecule; these vitamin derivatives are expressed by many strains of bacteria and yeast. Because MAIT cells constitute approximately 5% of human T cells and up to 25% of CD8 cells, their binding specificity suggests a role for these cells in host defenses.[30] In addition, certain human γδ T cells can recognize a variety of nonpeptide phosphoantigens, such as phosphorylated nucleotides, other phosphorylated small molecules, and alkylamines. The role of APCs and MHC-like molecules in presentation of phosphoantigens to γδ T cells remains a subject of investigation.

Another exception to the generalization of T-cell recognition of oligopeptides is represented by a group of proteins termed *superantigens*.[31] SAgs, of which the staphylococcal enterotoxin A represents a prototype, are produced by a broad spectrum of microbes, ranging from retroviruses to bacteria. They differ from conventional peptide antigens in their mode of contact both with MHC class II molecules and TCRs (Chapter 6). They do not undergo processing to oligopeptide fragments. Instead, they bind as intact proteins to class II molecules and TCRs outside the antigen-binding grooves. Their interaction with TCRs is predominantly determined by variable residues of the TCR V_β region. Because SAgs bind more or less independently of the TCRs α chain and the other variable segments of the β chain, they are capable of activating much larger numbers of T cells compared with conventional peptide antigens—hence their name. SAgs cause a wave of T-cell activation, proliferation, and production of proinflammatory molecules that can have profound clinical consequences, leading to development of such diseases as toxic shock syndrome.[31]

LYMPHOCYTE ADHESION AND TRAFFICKING

The capacity to continuously survey the antigenic environment is an essential element of immune function. APCs and lymphocytes must be able to find antigen wherever it occurs. Surveillance is accomplished through an elaborate interdigitated circulatory system of blood and lymphatic vessels that establish connections between the solid organs of the peripheral immune system (e.g., spleen, lymph nodes, and lymphoid structures in mucosal tissues) in which the interactions between immune cells predominantly occur (Chapter 2).

The trafficking and distribution of circulating cells of the immune system is largely regulated by interactions between molecules on leukocyte surfaces and ligands on vascular endothelial cells[32] (Chapter 16). Leukocyte-specific cellular adhesion molecules can be expressed constitutively or can be induced by cytokines (e.g., as a consequence of an inflammatory process).

Several families of molecules are involved in the regulation of lymphocyte trafficking. Particularly important are selectins and integrins, which ensure that mobile cells home to appropriate locations within lymphoid organs and other tissues. Selectins are proteins characterized by a distal carbohydrate-binding (lectin) domain. They bind to a family of mucin-like molecules, the endothelial vascular addressins. Integrins are heterodimers essential for the emigration of leukocytes from blood vessels into tissues. Members of the selectin and integrin families are involved in lymphocyte circulation and homing and are also important in interactions between APCs, T cells, and B cells in the induction and expression of immune responses. Certain endothelial adhesion molecules, mostly members of the Ig superfamily, are similarly involved in promoting interactions between T cells and APCs, as well as in leukocyte transmigration from the vasculature. Receptors for chemokines are important determinants of lymphocyte migration, particularly in guiding tissue-selective cell trafficking.[33]

LYMPHOCYTE ACTIVATION

For both B cells and T cells, initial activation is a two-signal event (Chapters 4 and 10).[34] This generalization is particularly true for immunologically naïve cells that have not been previously exposed to antigen. The first signal is provided by antigen. Most commonly, antigens for B cells are proteins with distinct sites, termed *epitopes*, which are bound by membrane Ig. Such epitopes can be linear, defined by a contiguous amino acid sequence or (more frequently) can be conformationally defined by the three-dimensional structure of the antigen. Epitopes can also be simple chemical moieties (haptens) that have been attached, usually covalently, to amino acid side chains (Chapter 6). In addition to proteins, some B cells have receptors with specificity for carbohydrates and, less commonly, lipids or nucleic acids. Antigens that stimulate B cells can be either in solution or fixed to a solid matrix (e.g., a cell membrane). As previously noted, the nature of antigens that stimulate T cells is more limited. TCRs do not bind antigen in solution but are usually stimulated only by small molecules, primarily oligopeptides, which reside within the antigen-binding cleft of a self-MHC molecule.

The second signal requisite for lymphocyte activation is provided by an accessory molecule expressed on the surface of the APC (e.g., B7/CD80) for stimulation of T cells or on the surface of a helper T cell (e.g., CD40L/CD154) for activation of B lymphocytes. The cell surface receptors for this particular second signal on T cells is CD28 and on B cells is CD40 (Fig. 1.3). Other cell surface ligand-receptor pairs may similarly provide the second signal (Chapters 7 and 10). The growth and differentiation of both T cells and B cells additionally require stimulation with one or more cytokines, which are peptide hormones secreted in small quantities by activated leukocytes and APCs for function in the cellular microenvironment.[35] In the absence of a second signal, cells stimulated only by antigen become unresponsive to subsequent antigen stimulation (i.e., *anergic*) (Chapter 10).[36]

Signal transduction through the antigen receptor is a complex process involving interactions between the specific receptor and molecules coexpressed in the cell membrane.[37] For B cells, this is a heterodimer, Igα/Igβ; and for T cells it is a macromolecular complex, CD3, usually comprising γ, δ, ε, and ζ chains.

Within the cell membrane, antigen receptor stimulation induces phosphorylation of Igα/Igβ or CD3 and hydrolysis of phosphatidylinositol 4,5-bisphosphate by phospholipase C, leading to generation of diacylglycerol (DAG) and inositol 1,4,5-triphosphate (IP_3). As a consequence of signal transduction and secondarily of DAG and IP_3 generation, tyrosine and serine/threonine protein kinases are activated. In turn, these kinases catalyze phosphorylation of a number of signal-transducing proteins. This leads to activation of cytoplasmic transcription factors NF-AT in T cells and NF-κB in both T cells and B cells. These transcription factors then translocate to the nucleus, where they bind to 5′ regulatory regions of genes that are critical to generalized lymphocyte activation (Chapter 10).[38]

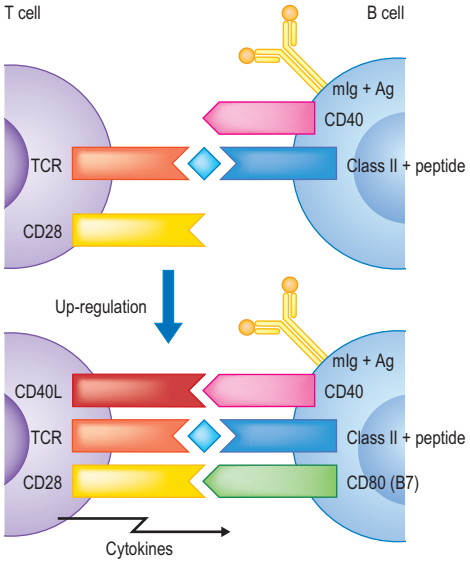


FIG. 1.3 Reciprocal Activation Events Involved in Mutual Simulation of T and B Cells. T cells constitutively express T-cell receptors *(TCRs)* and CD28. B cells constitutively express mIg and major histocompatibility complex (MHC) class II. Antigen bound to mIg is endocytosed and processed to peptide fragments that bind to MHC class II molecules for presentation to TCRs. Activation of B cells by antigen *(Ag)* upregulates their expression of CD80 that interacts with CD28 to activate T cells. This upregulates CD40L (CD154) on the T cell and induces cytokine synthesis. Costimulation of B cells by antigen CD40L and cytokines leads to Ig production. *mIg,* Membrane immunoglobulin.

CELL-MEDIATED IMMUNE RESPONSES

T-Cell Subsets

T lymphocytes expressing an αβ TCR can be divided into two major subpopulations based on the class of MHC molecule that their TCR recognizes and the consequent expression of one of a pair of TCR accessory molecules, CD4 or CD8 (Chapters 4 and 9). Binding of CD4 to class II MHC molecules or CD8 to class I MHC molecules on APCs contributes to the overall strength of intercellular molecular interactions. The ratio of CD4 to CD8 cells in peripheral blood is usually approximately 2:1.

CD4 T Cells, Cytokines, and Chemokines

The activities of CD4 T lymphocytes, commonly referred to as *T helper (Th) cells*, are mediated predominantly through the secretion of cytokines (Chapter 14). Cytokine activity can include autostimulation (autocrine function) if the cell producing the cytokine also expresses a surface receptor for it, or stimulation of other cells in the microenvironment of the cytokine-secreting cell (paracrine function) including B cells, APCs, and other T cells. Although it is now recognized that their biologic effects are broader than implied by their name, many of the principal cytokines active in the immune system are known as *interleukins* (ILs), implying that they are produced by a leukocyte to act on other leukocytes.

The specific profile of cytokines produced by CD4 T cells allows further functional subdivision (Chapters 11 and 14).[35,39] CD4 T cells elaborating the "inflammatory" cytokines involved in effector functions of cell-mediated immunity, such as IL-2 and IFN-γ, are designated Th1 cells. Th2 CD4 T cells synthesize cytokines (Chapter 14), such as IL-4 and IL-13, which control and regulate antibody responses and activate cells that are involved in host defense against parasites. Differentiation of Th1 versus Th2 subsets is a process substantially controlled by positive feedback loops, being promoted particularly by IL-12 in the case of Th1 cells and IL-4 in the case of Th2 cells.

It is important to note that generalizations regarding cytokine activity are usually oversimplifications, reflecting a substantial overlap and multiplicity of functions. For example, although IL-2 was initially identified as a T-cell growth factor, it also significantly affects B-cell differentiation. The prototypic inflammatory cytokine, IFN-γ, which promotes differentiation of effector function of cytotoxic T lymphocytes (CTLs) and macrophages, is also involved in regulation of Ig isotype switching. Similarly, IL-4 is known primarily as a B-cell growth and differentiation factor, but it can also stimulate proliferation of T cells.

A distinct subset of cytokines is a large group of highly conserved cytokine-like molecules, smaller than typical cytokines (~7 to 12 kilodaltons [kDa]), termed *chemo*tactic cyto*kines*, or *chemokines* (Chapter 15).[33] Chemokines are classified on the basis of the number and spacing of specific cysteine residues. They regulate and coordinate trafficking and activation of leukocytes, functioning importantly in host defenses, and also broadly in a variety of nonimmunologic processes, including organ development and angiogenesis. They are characterized by binding to seven-transmembrane-domain G protein–coupled receptors. Of particular interest to clinical immunologists are the chemokine receptors CCR5 and CXCR4. Together with CD4, these molecules are used as coreceptors by human immunodeficiency virus (HIV) to gain entry into target cells (Chapter 41).[40]

Cytokines produced by activated T cells can downregulate as well as initiate or amplify immune responses.[41] Downregulating cytokines include IL-10 (produced by both T cells and B cells) and transforming growth factor-β (TGF-β). The functions of IL-10 in vivo are thought to include both suppression of proinflammatory cytokine production and enhancement of IgM and IgA synthesis. TGF-β, which is produced by virtually all cells, expresses a broad array of biologic activities. These include promotion of wound healing and suppression of both humoral and cell-mediated immune responses.

In addition to their central role in initiation and regulation of immune responses, CD4 T cells are important effectors of cell-mediated immunity (Chapter 11). Through the elaboration of inflammatory cytokines, particularly IFN-γ, they are essential contributors to the generation of chronic inflammation, which is characterized histologically by mononuclear cell infiltrates. Here their principal role is thought to be the activation of macrophages. In some circumstances, CD4 T cells can function as cytotoxic effectors, either directly as CTLs, in which case the killing is "restricted" for recognition of antigen-bound self-MHC class II, or through the elaboration of cytotoxic cytokines, such as lymphotoxin and tumor necrosis factor (TNF).

A third subset of Th cells, designated Th17, has been recognized more recently. With differentiation driven particularly by TGF-β and IL-23 and characterized by the production of the proinflammatory cytokine IL-17, Th17 cells are important in the induction and exacerbation of autoimmunity in a variety of disease models, as well as in host defenses against a broad

spectrum of extracellular bacteria, fungi, and other pathogens.[42] Research continues to identify additional examples of CD4 T cells that may become recognized as distinct subsets. Their function is governed by other predominantly expressed cytokines to achieve specialized effector responses.

One final category of CD4 cells, Tregs, plays a crucial role suppressing the functions of other lymphocytes (Chapter 13). Tregs can differentiate either in the thymus (tTregs) or in the periphery (pTregs). A third category of Tregs are induced in vitro (iTregs).[43] Tregs are commonly characterized by surface expression of CD4 and CD25 and by nuclear expression of the transcription factor Foxp3. Peripheral activation of $CD25^+$ Tregs is via the TCRs. The cells are IL-2 dependent and apparently require cell-to-cell contact for suppressive function. They can suppress functions of both CD4 and CD8 T cells, as well as B cells, NK cells, and NKT cells. In contrast to activation, suppressor effects are independent of the antigen specificity of the target cells. Other Tregs are noted for production of inhibitory cytokines. These include IL-10– and TGF-β–secreting Th3 cells and IL-10–producing Tr1 cells.[44,45]

CD8 T Cells

The best understood function of CD8 T cells is that of CTL effectors.[46] CTLs are particularly important in host defenses against virus-infected cells, where they can kill target cells expressing viral peptides bound to self-MHC class I molecules (Chapter 12). This process is highly specific and requires direct apposition of CTLs and target cell membranes. Bystander cells expressing MHC molecules that have bound peptides that the CD8 T cell does not recognize are not affected. The killing is unidirectional; the CTL itself is not harmed and after transmission of a "lethal hit" it can detach from one target to seek another. Killing occurs via two mechanisms: a death receptor–induced apoptotic mechanism and a mechanism involving insertion of perforins into the target cell membrane to create a pore through which granzymes and other cytotoxic enzymes can be transferred from the CTL into the target cell. CTL activity is enhanced by IFN-γ. As CTL function is dependent on target cell surface display of MHC class I molecules, a principal mechanism of immune evasion by viruses and tumors is elaboration of factors that downregulate class I expression (Chapter 25). However, this increases susceptibility of such cells to cytolysis by NK cells that are activated to attack cells expressing low levels of class I MHC molecules.

ANTIBODY-MEDIATED IMMUNE RESPONSES

The structure of antibodies permits a virtually limitless binding specificity of its antigen-binding site. Antigen binding can then be translated into biologic effector functions based on properties of the larger nonvariable (constant) region of its heavy chains (Fc fragment) (Chapter 8). Moreover, in response to cytokines in the cellular microenvironment, the mechanism of isotype switching enables the antibody-producing cell to switch the exons that are used to encode its heavy-chain constant region and thereby alter the biologic effects of its secreted product without affecting its antigen-binding specificity. With functional heterogeneity determined by isotype, antibody molecules provide a broad-based and efficient defense system against extracellular microbes or foreign macromolecules (e.g., toxins and venoms) (Chapters 8, 27, and 87).

KEY CONCEPTS

Biologic Properties of Immunoglobulin (Ig) Classes

- IgM
  - Monomeric on the cell surface, primarily pentameric in soluble form
  - Principal Ig of the primary immune response
  - Generally restricted to the vascular compartment
  - Antigen receptor (monomer) for most naïve B cells
  - Fixes complement potently
- IgG
  - Monomeric
  - Principal Ig of internal secondary immune responses
  - Binds to Fcγ receptors on neutrophils, monocytes/macrophages, and NK cells
  - Four subclasses, each with a different effector function
  - Fixes complement (except IgG4 subclass)
- IgA
  - Monomeric or dimeric
  - Transported into the gut
  - Principal Ig of mucosal immunity
  - Two subclasses
- IgD
  - Antigen receptor for mature B cells
  - Typically expressed on cells that also express membrane IgM
- IgE
  - Binds to Fcε receptors on mast cells and basophils
  - Antibody of immediate hypersensitivity
  - Important in defenses against helminths

Each antibody class contributes differently to an integrated defense system.[47] IgM is the predominant class formed on initial contact with antigen (primary immune response). As a monomeric structure comprises two light (κ or λ) and two heavy (μ) chains, it is initially expressed as a membrane-bound antigen receptor on the surface of B lymphocytes. The avidity of serum IgM for antigen binding is increased by its organization into a pentamer of five of the monomeric subunits held together by a joining (J) chain. IgM is essentially confined to the intravascular compartment. As a multivalent antigen binder that can efficiently activate ("fix") complement, it is an important contributor to immune responses early after the initial encounter with antigen. The synthesis of IgM, compared with other isotypes, is much less dependent on the activity of T lymphocytes.

IgG is the most abundant Ig in serum and the principal antibody class of a secondary (anamnestic or memory) immune response. IgG molecules are heterodimeric monomers with two light (κ or λ) and two heavy (γ) chains and, except for IgG4, joined by interchain disulfide bridges. Because of its abundance, its capacity to fix complement, and the expression on phagocytes of Fcγ receptors, IgG is the most important antibody for systemic secondary immune responses. IgG is the only isotype that is actively transported across the placenta. These transported maternal IgG antibodies provide the neonate with an important level of antibody protection in the first 6 months of life, when its own antigen-driven antibody responses are first developing (Chapter 21).

IgA is the principal antibody in the body's secretions (Chapter 24). It is found in serum in monomeric form of two light and two heavy (α) chains or as a dimer joined by J chain. In secretions, it is usually present in dimeric form and is actively secreted across mucous membranes by attachment of a specialized secretory component (SC) that is recognized by the polyIg receptor on mucosal epithelial cells. Dimeric IgA is found in

high concentration in tears, saliva, and secretions of the respiratory, gastrointestinal, and genitourinary systems. It is relatively resistant to enzymatic digestion. It is particularly abundant in colostrum, where its concentration may be greater than 50 times that in serum, providing passive immunity to the gastrointestinal system of a nursing neonate. IgA does not fix complement by the antibody-dependent pathway and hence does not promote phagocytosis. Its role in host defenses lies in preventing a breach of the mucous membrane surface by microbes or their toxic products.

IgD and IgE are present in serum at concentrations much lower than that of IgG. The biologic role of serum IgD remains controversial.[48] B cells can express both membrane IgM and IgD by alternative splicing of the Ig heavy-chain gene or can secrete only IgD via an apparently atypical form of class switch recombination. These mechanisms do not require T-cell help.

Although IgE is the least abundant isotype in serum, it has dramatic biologic effects because it is responsible for immediate-type hypersensitivity reactions, including systemic anaphylaxis (Chapter 46). Such reactions reflect expression of high-affinity receptors for Fcε on the surfaces of mast cells and basophils. Cross-linking of IgE molecules on such cells by antigen induces their degranulation, with the immediate release of preformed potent biologic mediators and de novo synthesis and secretion of additional proinflammatory molecules. The protective role of IgE is in host defenses against parasitic infestation, particularly with helminths (Chapter 30).

Complement and Immune Complexes

The effector functions of IgG and IgM depend, in part, on their capacities to activate the complement system. Through a cascade of sequential substrate-enzyme interactions, the 11 principal components of the antibody-dependent complement cascade (C1q, C1r, C1s, and C2–C9) cause many of the principal consequences of an antigen-antibody interaction (Chapter 40). These include the establishment of pores in a target cell membrane by the terminal components (C5–C9) leading to osmotic lysis; the production of factors (principally C5a) with chemotactic activity for phagocytic myeloid cells; opsonization by C3b, promoting phagocytosis; and the ability to induce degranulation of mast cells (C3a, C4a, and C5a).

There are three distinct pathways to complement activation.[49] The pathway mediated by the binding of the first component (specifically C1q) to IgG or IgM has been termed the "classical" pathway (CP). The lectin pathway is similar to the CP but is activated by selected carbohydrate-binding proteins, the mannose (or mannan)-binding lectin (MBL), and ficolins, which recognize certain carbohydrate repeating structures on microorganisms. MBL and ficolins are plasma proteins that are homologous to C1q and contribute to innate immunity through their capacity to induce antibody- and C1q-independent activation of the CP. Finally, a large number of substances, including certain bacterial, fungal, and viral products, can directly activate the cascade through a distinct series of proteins also leading to activation of the central C3 component. Although bypassing C1, C4, and C2, this distinct pathway can achieve all the biologic consequences of C3 to C9 activation. Non–antibody-induced activation of C3 is referred to as the "alternative" pathway (AP) or "properdin" pathway. In addition, the central components of the cascade (e.g., C5a) can be directly produced by the action of serine proteases of the coagulation system.[49] Reflecting these separate pathways to activation and the fact that many types of leukocytes express receptors for activated complement components, the complement system is a major contributor to the efferent limbs of both innate and adaptive immune systems.

In addition to their roles in pathogen/antigen elimination, constituents of the complement system, together with antigen-antibody (immune) complexes, act at leukocyte surfaces to regulate immune functions. For example, interaction of immune complexes via FcγR on B cells decreases their responsiveness to stimulation. In contrast, complement activation on B-cell surfaces co-ligates their receptors with BCRs for antigen, rendering the cells more readily activated and resistant to apoptosis.

Essential for the proper function of the complement system is a series of downregulatory mechanisms that prevent unwanted activation of the system and that extinguish its activity when no longer needed. The regulatory pathways are mediated by a combination of both soluble complement-binding and digesting molecules and cell surface binding proteins.

APOPTOSIS AND IMMUNE HOMEOSTASIS

An immune response is commonly first viewed in a "positive" sense—that is, lymphocytes are activated, proliferate, differentiate, and carry out effector functions. However, it is equally important that this positive response be tightly regulated by mechanisms that operate to turn off the response and to eliminate cells no longer required (Chapter 17).[50,51] Under physiologic circumstances, once an immune response fades, commonly as a consequence of antigen depletion, two pathways to terminal lymphocyte differentiation become available: apoptosis or differentiation into memory cells. Memory cells are, of course, a key to the effectiveness of the adaptive immune system, because a second activating encounter with antigen (e.g., pathogen) is both more rapid and more productive. Isotype-switched high-affinity antibodies are rapidly produced, and/or clones of CTL effector cells proliferate. However, the majority of lymphocytes in an active response are not required for maintenance of immunologic memory, and the necessity for homeostasis leads to apoptosis of cells no longer required.

Apoptosis (or regulated cell death [RCD]) is a unique process of cellular death and widely conserved phylogenetically. It is distinguished from death by necrosis by cellular shrinking, DNA fragmentation, and breakdown of cells into "apoptotic bodies" containing nuclear fragments and intact organelles that can be eliminated by phagocytosis without release into the extracellular space of the majority of intracellular, especially nuclear, components. Necrosis can be genetically determined (regulated necrosis [RN]) or unregulated, reflecting some accidental or otherwise inevitable process. Apoptosis depends on the activation of cysteinyl proteases, termed *caspases*, which cleave proteins that regulate DNA repair and the establishment/maintenance of cellular architecture. In the absence of these apoptotic mechanisms, massive proliferation of cells in lymphoid tissues results. This is seen clinically as autoimmune lymphoproliferative syndrome (ALPS), which is characterized by lymphocytosis with lymphadenopathy and splenomegaly as well as autoimmunity and hypergammaglobulinemia (Chapter 51).[52]

MECHANISMS OF IMMUNOLOGIC DISEASES

Immunologic diseases can be classified on the basis of our understanding of normal immune physiology and its perturbations in disease states (Table 1.2). One type of immunologic

TABLE 1.2 Mechanisms of Immunologic Diseases

1. Functional deficiency of key immune system components
   a. Congenital
   b. Acquired
2. Malignant transformation of immune system cells
3. Immunologic dysregulation
4. Autoimmunity
5. Untoward consequences of physiologic immune function

disease results from failure or deficiency of a component of the immune system leading to failure of normal immune function (Chapters 32–42). Such disorders are usually identified by increased susceptibility to infection (Chapter 32). Failure of host defense can be congenital (e.g., X-linked agammaglobulinemia; Chapter 33) or acquired (e.g., acquired immunodeficiency syndrome [AIDS]; Chapter 41). It can be global (e.g., severe combined immunodeficiency [SCID]; Chapter 34) or, quite specific, involving only a single component of the immune system (e.g., selective IgA deficiency; Chapter 33).

A second type of immunologic disease is malignant transformation of immunologic cells (Chapters 77–81). Manifestations of leukocyte malignancies are protean, most commonly reflecting the secondary consequences of solid organ or bone marrow infiltration or replacement of normal cells by tumor cells, with resulting anemia and immunologic deficiency.

The remaining types of immunopathogenesis are more specific to the immune system. Dysregulation of an essentially intact immune system constitutes a third general type of immune disorder. Features of an optimal immune response include antigen recognition and elimination, with little adverse effect on the host. However, both initiation and termination of the response involve regulatory interactions that can go awry when the host is challenged by antigens of a particular structure or presented in a particular fashion. Diseases of immune dysregulation can result from genetic and environmental factors that act together to produce a pathologic immune response, such as acute allergic diseases (Chapters 43–50). Some forms of allergic disease are thought to be a consequence of insufficient exposure to nonpathogenic microbes and other potential allergens in early childhood, resulting in an increased susceptibility to allergy, atopy, and asthma once the immune system has matured. The so-called hygiene hypothesis suggests that mucosal tissue–colonizing organisms play key roles in the initial establishment of immune homeostasis.[53] The importance of establishing immune homeostasis early in life is also supported by studies demonstrating reduction in the likelihood of food allergy associated with feeding of the allergenic foods to infants at high risk for allergy (Chapter 49).[54–56]

A fourth type of immunologic disorder is the result of failure of a key feature of normal immune recognition, the molecular discrimination between self and nonself. Ambiguity in this discrimination can lead to autoimmune tissue damage (Chapters 51–76). Although such damage can be mediated by either antibodies or T cells, the common association of specific autoimmune diseases with inheritance of particular HLA alleles (Chapter 5) suggests that the pathogenesis of autoimmune diseases usually represents a failure of regulation of the anti-self inflammatory response by T cells.

The immunologic attack on self-tissues can be general, leading to systemic autoimmunity, such as systemic lupus erythematosus; or it can be localized, as in organ-specific autoimmune diseases. In the latter instances, the immune system attacks specific types of cells and usually particular cell surface molecules. In most cases, pathology is a consequence of target tissue destruction (e.g., multiple sclerosis, rheumatoid arthritis, or insulin-dependent diabetes mellitus). However, depending on the antigenic specificity of the abnormal immune response, autoimmunity can lead to receptor blockade (e.g., myasthenia gravis or insulin-resistant diabetes) or hormone receptor stimulation (e.g., Graves disease). It is thought by many immunologists that ambiguity in self/nonself discrimination is commonly triggered by an unresolved encounter with an infectious organism or other environmental agent that shares some structural features with self-tissue structures, although this remains a subject of controversy (Chapter 51).[57,58] Insight into mechanisms whereby specific HLA alleles predispose to development of autoimmunity and others may be protective are suggested by studies in HLA-transgenic mice, which suggest that alleles that predispose animals to a particular autoimmune disease may reflect a T-cell phenotype associated with secretion of proinflammatory cytokines. In contrast, protective alleles were associated with elaboration of tolerogenic cytokines by Tregs.[59]

A fifth form of immunologic disease occurs as a result of physiologic, rather than pathologic, immune functions. Inflammatory lesions in such diseases are the result of the normal function of the immune system. A typical example is contact dermatitis to such potent skin sensitizers as urushiol, the causative agent of poison ivy dermatitis (Chapter 48). These diseases can also have an iatrogenic etiology that can range from mild and self-limited (e.g., delayed hypersensitivity skin test reactions) to life threatening (e.g., graft-versus-host disease, organ graft rejection).

HOST IMMUNE DEFENSES SUMMARIZED

The first response upon initial contact with an invading pathogen depends on components of the innate immune system (Chapter 3). This response begins with recognition of PAMPs expressed by cells of the pathogen. These include lipoproteins, lipopolysaccharide, unmethylated CpG-DNA, and bacterial flagellin, among others. PAMPs bind to PRRs on or within effector cells of the host's innate immune system, including DCs, granulocytes, and ILCs.[2,3] The best characterized PRRs are the TLRs, first recognized as determinants of embryonic patterning in *Drosophila* and subsequently appreciated as components of host defenses in both insects and vertebrates. TLR subfamilies can be distinguished by expression either on the cell surface or in intracellular compartments. A second major family of PRRs comprises NLRs, which detect intracellular microbial products. Binding of TLRs or NLRs by PAMP ligands triggers intracellular signaling pathways via multiple "adapters," leading to a vigorous inflammatory response.

The innate immune response also includes the capacity of NK lymphocytes to identify and destroy, by direct cytotoxic mechanisms, cells lacking surface expression of MHC class I molecules, which marks them as potentially pathogenic.[4] In addition, an innate immune response involves elements of the humoral immune system that function independently of antibody, especially the activation of the complement cascade through

the lectin pathway and the AP, with consequent opsonization of particles and microbes to promote their phagocytosis and destruction.

The nature of the adaptive immune response to any particular pathogenic agent is determined largely by the context in which the pathogen is encountered. Regardless, effectiveness depends on the four principal properties of adaptive immunity: (i) a virtually unlimited capacity to bind macromolecules, particularly proteins, with exquisite specificity, reflecting generations of antigen-binding receptors by genetic recombination and, in the case of B cells, SHM; (ii) the capacity for self/nonself discrimination, consequences of a rigorous process involving positive and negative selection during thymocyte differentiation, as well as negative selection during B-cell differentiation; (iii) the property of immunologic memory, reflecting antigen-driven clonal proliferation of T cells and B cells, which results in increasingly rapid and effective responses on second and subsequent encounters with a particular antigen or pathogen; and (iv) mechanisms for pathogen destruction, including direct cellular cytotoxicity, release of inflammatory cytokines, opsonization with antibody and complement, and neutralization in solution by antigen precipitation or conformational alteration plus phagocytosis and intracellular digestion.

Although most acquired immune responses include multiple defense mechanisms, several generalizations may be conceptually useful. T cell–mediated (and NK cell–mediated) effector functions are particularly important in defenses against pathogens encountered intracellularly or at cell surfaces, such as intracellular viruses, intracellular bacteria, and tumor cells. These responses involve the production of inflammatory cytokines by CD4 Th1 cells, as well as the direct cytolytic activity of CD8 CTLs. In contrast, host defenses to most antigens encountered primarily in the extracellular milieu are largely dependent on humoral mechanisms (antibody and complement) for antigen neutralization, precipitation, or opsonization and subsequent destruction by phagocytes. Targets of antibody-mediated immunity include extracellular virus particles, bacteria, and toxins (or other foreign proteins). However, it is worth reiterating that induction of an effective antibody response (including isotype switching) and development of immunologic memory (resulting from B-cell clonal expansion and B memory cell differentiation) require antigen activation not only of specific B cells but also CD4 T cells, particularly of the Th2 type. In addition, antibacterial and antifungal responses that involve prominent responses by neutrophils require CD4 T cells of the Th17 type.

Finally, clinical "experiments of nature" have proven particularly instructive in our efforts to understand the role of specific components of the immune system in overall host defenses (Chapter 32).[60] The importance of T cell–mediated immunity in host defenses to intracellular parasites, fungi (Fig. 1.4), and viruses is emphasized by the remarkable susceptibility of patients with T cell deficiency to pathogenic organisms, such as *Pneumocystis jiroveci* and *Candida albicans*, and by the fact that using attenuated live virus vaccines in such patients can lead to devastating disseminated infections. Indeed, the relationship between susceptibility to particular potential pathogens and specific immunologic deficiencies is nicely illustrated by demonstrations that the pathogenesis of various familial forms of chronic mucocutaneous candidiasis reflect deficiency of IL-17–mediated immunity.[61]

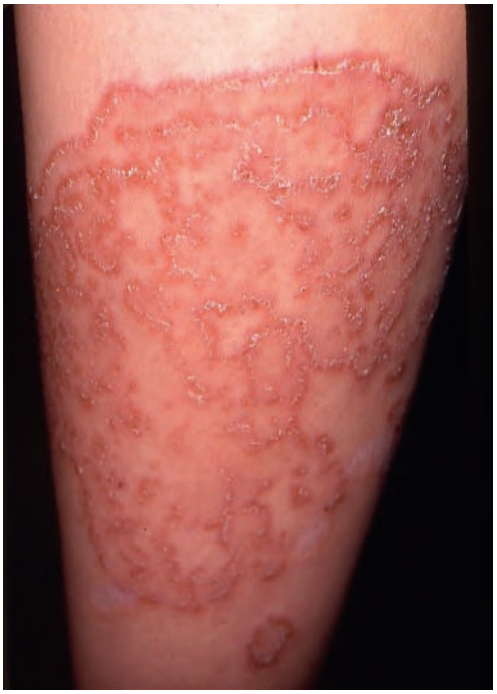


FIG. 1.4 Leg of a 16-year-old patient with chronic mucocutaneous candidiasis as a consequence of a congenital T-cell deficiency associated with hypoparathyroidism and adrenal insufficiency.

Patients with defects in antibody synthesis or phagocytic cell function are characteristically afflicted with recurrent infections with pyogenic bacteria, particularly gram-positive organisms. And patients with inherited defects in synthesis of terminal complement components have increased susceptibility to infection with species of *Neisseria*.

In recent years, immunology has entered the lay lexicon, largely as a result of the HIV and Covid-19 pandemics. People throughout the world are now aware of the tragic consequences of immune deficiency. However, the remarkable progress in understanding this diseases rested substantially on earlier studies of relatively rare patients with primary immunodeficiency syndromes and on genomic definition of their molecular basis. Similarly, cure of patients with primary immunodeficiencies by cellular reconstitution, particularly bone marrow or stem cell transplantation (Chapter 90),[62] presaged recent progress in correction of such diseases by gene replacement therapy (Chapter 91).[63] The "present" of clinical immunology is, indeed, bright, but its future potential to impact prevention and treatment of many challenging diseases, including cancer (Chapters 80 and 81), through specific analysis of genetic mutations and enhancement or suppression of antigen-specific immune responses with chimeric antigen receptor (CAR) T cells and checkpoint blockade (CPB) is even more exciting to contemplate.[64] A few approaches are broadly hinted at here, and it is hoped that readers will enjoy considering such "perspectives" throughout the book. Given the nature of the immune system, it is also hoped it will challenge readers to transform a particular author's views to new and different clinical settings.

ON THE HORIZON

Enhancement of Immune Responses

- Use of CRISPR/Cas9 technology in correction of gene defect in monogenic immunodeficiency diseases
- Enhancement of molecular pathway–specific cancer immunotherapy
- Prediction of molecular targets and improvements in adjuvants for vaccines

Suppression of Immune Responses

- Further development of protocols for prevention of childhood allergic diseases by appropriate environmental exposures in infancy
- Improved allergen-specific immunotherapy of allergic diseases
- Prevention of graft-versus-host disease in allogeneic bone marrow transplantation cases
- Induction of antigen-specific immune tolerance in humans
- Pharmacologic development of specific inhibitors of cytokines, chemokines, and their receptors

Immunodiagnostics and Immunopathogenesis

- Routine use of genomic analysis for development of personalized medicine applicable to immunologic diseases
- Development of novel diagnostic tools based on nanotechnology arrays
- Application of gastrointestinal microbiome analysis in understanding the pathogenesis of inflammatory bowel diseases, leading to novel therapeutic options
- Understanding the role of inflammation in the pathogenesis of leading causes of morbidity and mortality, including myocardial infarction, stroke, and Alzheimer disease

Studies in experimental animals, especially murine studies, have been critical to our understanding of molecular aspects of immune system function and have contributed importantly to our appreciation of how aberrations of such functions are involved in the pathogenesis of disease. Insights gained from use of transgenic mice (including murine expression of human genes) and constitutive or conditional gene-knockout mice are essential to a comprehensive view of the immune system at the advancing edge of its clinical application, implying that future progress in clinical immunology will equally depend on detailed analysis in such systems. However, it has become apparent that there are important differences in aspects of the human and rodent immune systems. Consequently, the carefully studied patient, particularly when coupled with the power of increasingly feasible genome and transcriptome sequencing, will remain the ultimate crucible for understanding human immunity and the roles of the immune system in the pathogenesis of and protection from disease.

REFERENCES

1. Hirano M, Das S, Guo P, Cooper MD. The evolution of adaptive immunity in vertebrates. *Adv Immunol*. 2011;109:125–157.
2. Buchmann K. Evolution of innate immunity: clues from invertebrates via fish to mammals. *Front Immunol*. 2014;5:459.
3. Kruse PH, Matta J, Ugolini S, Vivier E. Natural cytotoxicity receptors and their ligands. *Immunol Cell Biol*. 2014;92:221–229.
4. Liu J, Cao X. Cellular and molecular regulation in innate inflammatory responses. *Cell Mol Immunol*. 2016;13:711–721.
5. Croft M, Dubey C. Accessory molecule and costimulation requirements for CD4 T cell response. *Crit Rev Immunol*. 2017;37:261–290.
6. Lindquist RL, Niesner RA, Hauseer AE. In the right place, at the right time: spatiotemporal conditions determining plasma cell survival and function. *Front Immunol*. 2019;10:788. https://doi.org/10.3389/fimmu.2019.00788.
7. Juelke K, Romagnani C. Differentiation of human innate lymphoid cells (ILCs). *Curr Opin Immunol*. 2016;38:75–85.
8. Pende D, Falco M, Vitale M, et al. Killer Ig-like receptors (KIRs): their role in NK cell modulation and developments leading to their clinical exploitation. *Front Immunol*. 2019;10:1177. https://doi.org/10.3389/fimmu.2019.01179.
9. Humphrey MB, Nakamura MC. A comprehensive review of immunoreceptor regulation of osteoclasts. *Clin Rev Allergy Immunol*. 2016;51:48–58.
10. Pahl JHW, Cerwenka A, Ni J. Memory-like NK cells: remembering a previous activation by cytokines and NK cell receptors. *Front Immunol*. 2018;9:2796. https://doi.org/10.3389/fimmu.2018.02796.
11. Kang TH, Jung ST. Boosting therapeutic potency of antibodies by taming Fc domain functions. *Exp Mol Med*. 2019;51:138. https://doi.org/10.1038/s12276-019-0345-9.
12. Anderson DA, Murphy KM. Models of dendritic cell development correlate ontogeny with function. *Adv Immunol*. 2019;143:99–119.
13. Tesfaye DY, Gudjonsson A, Bogen B, Fossum E. Targeting conventional dendritic cells to fine-tune antibody responses. *Front Immunol*. 2019;10:1529. https://doi.org/10.3389/fimmu.2019.01529.
14. Le Nours J, Shahine A, Gras S. Molecular features of lipid-based antigen presentation by group 1 CD1 molecules. *Semin Cell Dev Biol*. 2018;84:48–57.
15. Harjunpää H, Asens ML, Guenther C, Fagerholm SC. Cell adhesion molecules and their roles and regulation in the immune and tumor microenvironment. *Front Immunol*. 2019;10:1078. https://doi.org/10.3389/fimmu.2019.01078.
16. Xu Z, Jin B. A novel interface consisting of homologous immunoglobulin superfamily members with multiple functions. *Cell Mol Immunol*. 2010;7:11–19.
17. Outters P, Jaeger S, Zaarour N, Perrier P. Long-range control of V(D)J recombination & allelic exclusion: modeling views. *Adv Immunol*. 2015;128:363–413.
18. Brady BL, Steinel NC, Bassing CH. Antigen receptor allelic exclusion: an update and reappraisal. *J Immunol*. 2010;185:3801–3808.
19. Chi X, Li Y, Qiu X. V(D)J recombination, somatic hypermutation and class switch recombination of immunoglobulins: mechanism and regulation. *Immunology*. 2020;160(3):233–247. https://doi.org/10.1111/imm.13176.
20. Zhang L, Dong X, Lee M, et al. Single-cell whole-genome sequencing reveals the functional landscape of somatic mutations in B lymphocytes across the human lifespan. *Proc Natl Acad Sci USA*. 2019;116:9014–9019.
21. Kurd N, Robey EA. T-cell selection in the thymus: a spatial and temporal perspective. *Immunol Rev*. 2016;271:114–126.
22. Kisielow P. How does the immune system learn to distinguish between good and evil? The first definitive studies of T cell central tolerance and positive selection. *Immunogenetics*. 2019;71:513–518.
23. Yi J, Kawabe T, Sprent J. New insights on T-cell self-tolerance. *Curr Opin Immunol*. 2019;28:14–20.
24. Methot SP, Di Noia M. Molecular mechanisms of somatic hypermutation and class switch recombination. *Adv Immunol*. 2017;133:37–87.
25. Antunes DA, Abella JR, Devaurs D, et al. Structure-based methods for binding mode and binding affinity prediction for peptide-MHC complexes. *Curr Top Med Chem*. 2018;18:2239–2255.
26. Hansen SG, Wu HL, Burwitz BJ, et al. Broadly targeted CD8+ T cell responses restricted by major histocompatibility complex E. *Science*. 2016;351:714–720.
27. Jurewicz MM, Stern LJ. Class II MHC antigen processing in immune tolerance and inflammation. *Immunogenetics*. 2019;71:171–187.
28. CohChancellor A, Gadola SD, Mansour S. The versatility of the CD1 lipid antigen presentation pathway. *Immunology*. 2018;154:196–203.
29. Roche PA, Furuta K. The ins and outs of MHC class II-mediated antigen processing and presentation. *Nat Rev Immunol*. 2015;15:203–216.
30. Meermeier EW, Harriff MJ, Karamooz E, Lewinsohn DM. MAIT cells and microbial immunity. *Immunol Cell Biol*. 2018;96:607–617.
31. Spaulding AR, Salgado-Pabon W, Kohler PL, et al. Staphylococcal and streptococcal superantigen exotoxins. *Clin Microbiol Rev*. 2013;26:422–447.
32. Regiero-Real N, Colom B, Bodkin JV, Nourshargh S. Endothelial cell junctional adhesion molecules: role and regulation of expression in inflammation. *Arterioscler Thromb Vasc Biol*. 2016;36:2048–2057.

33. Schulz O, Hammerschmidt SI, Moschovakis GL, Forster R. Chemokines and chemokine receptors in lymphoid tissue dynamics. *Annu Rev Immunol*. 2016;34:203–242.
34. Wingren AG, Parra E, Varga M, et al. T cell activation pathways: B7, LFA-3, and ICAM-1 shape unique T cell profiles. *Crit Rev Immunol*. 2017;37:463–487.
35. Cronkite DA, Strutt TM. The regulation of inflammation by innate and adaptive lymphocytes. *J Immunol Res*. 2018;2018:1467538. https://doi.org/10.1155/2018/1467538.
36. Chappert P, Schwartz RH. Induction of T cell anergy: integration of environmental cues and infectious tolerance. *Curr Opin Immunol*. 2010;22:552–559.
37. Fink K. Can we improve vaccine efficacy by targeting T and B cell repertoire convergence? *Front Immunol*. 2019;10:110. https://doi.org/10.3389/fimmu.2019.00110.
38. De Obaldia ME, Bhandoola A. Transcriptional regulation of innate and adaptive lymphocyte lineages. *Annu Rev Immunol*. 2015;33:607–642.
39. Kondo Y, Yokosawa M, Kaneko S, et al. Review: transcription regulation of CD4+ T cell differentiation in experimentally induced arthritis and rheumatoid arthritis. *Arthritis Rheumatol*. 2018;70:653–661.
40. Grande F, Occhiuzzi MA, Rizzuti B, et al. CCR5/CXCR4 dual antagonism for the improvement of HIV infection therapy. *Molecules*. 2019;2:24. https://doi.org/10.3390/molecules24030550.
41. Liu D, Cao S, Zhou Y, Xiong Y. Recent advances in endotoxin tolerance. *J Cell Biochem*. 2019;120:56–70.
42. Mazzzzoni A, Maggi L, Liotta F, et al. Biological and clinical significance of T helper 17 cell plasticity. *Immunology*. 2019;158:287–295.
43. Shevach EM, Thornton AM. tTregs, pTregs, and iTregs: similarities and differences. *Immunol Rev*. 2014;259:88–102.
44. Chien CH, Chiang BL. Regulatory T cells induced by B cells: a novel subpopulation of regulatory T cells. *J Biomed Sci*. 2017;24:86. https://doi.org/10.1186/s12929-017-0391-3.
45. Plitas G, Rudensky AY. Regulatory T cells: differentiation and function. *Cancer Immunol Res*. 2016;4:721–725.
46. Cosma GL, Eisenlohr LC. Impact of epitope density on CD8+ T cell development and function. *Mol Immunol*. 2019;113:120–125.
47. Sun Y, Huang T, Hammarström L, Zhao Y. The immunoglobulins: new insights, implications, and applications. *Annu Rev Anim Biosci*. 2020;8:145–169. https://doi.org/10.1146/annurev-animal-021419-083720.
48. Gutzeit C, Chen K, Cerutti A. The enigmatic function of IgD: some answers at last. *Eur J Immunol*. 2018;48:1101–1113.
49. Ling M, Murali M. Analysis of the complement system in the clinical immunology laboratory. *Clin Lab Med*. 2019;39:579–590.
50. Zheng L, Lenardo M. Restimulation-induced cell death: new medical and research perspectives. *Immunol Rev*. 2017;277:44–60.
51. Kane BA, Brant KJ, McNeil HP, Tedla NT. Termination of immune activation: an essential component of healthy host immune responses. *J Innate Immunol*. 2014;6:727–738.
52. Rieux-Laucat F. What's up in the ALPS. *Curr Opin Immunol*. 2017;49:79–86. https://doi.org/10.1016/j.coi.2017.10.001.
53. Lambrecht BN, Hammad H. The immunology of the allergy epidemic and the hygiene hypotheses. *Nat Immunol*. 2017;18:1076–1083.
54. Caffarelli C, Di Mauro D, Mastrorilli C, et al. Solid food introduction and the development of food allergies. *Nutrients*. 2018;10(11):1790. https://doi.org/10.3390/nu10111790.
55. Roberts G, Grimshaw K, Beyer K, et al. Can dietary strategies in early life prevent childhood food allergy? A report from two iFAAM workshops. *Clin Exp Immunol*. 2019;49:1567–1577.
56. Ierodiakonou D, Garcia-Larsen V, Logan A, et al. Timing of allergenic food introduction to the infant diet and risk of allergic or autoimmune disease. *JAMA*. 2016;316:1181–1192.
57. Theofilopoulos AN, Kono DH, Baccala R. The multiple pathways to autoimmunity. *Nat Immunol*. 2017;1l8:716–724.
58. Steed AL, Stappenbeck TS. Role of viruses and bacteria-virus interactions in autoimmunity. *Curr Opin Immunol*. 2014;31:102–107.
59. Ooi JD, Petersen J, Tan YH, et al. Dominant protection from HLA-linked autoimmunity by antigen-specific regulatory T cells. *Nature*. 2017;545:243–247.
60. Delmonte OM, Castognoli R, Calzoni E, Notarangelo LD. Inborn errors of immunity with immune dysregulation: from bench to bedside. *Front Pediatr*. 2019;7:353. https://doi.org/10.3389/fped.2019.00353.
61. Okada S, Puel A, Casanova JL, Kobayashi M. Chronic mucocutaneous candidiasis disease associated with inborn errors of IL-17 immunity. *Clin Transl Immunology*. 2016;5(12):e114. https://doi.org/10.1038/cti.2016.71.
62. Castagnoli R, Delmonte OM, Calzoni E, Notarangelo LD. Hematopoietic stem cell transplantation in primary immunodeficiency diseases: current status and future perspectives. *Front Pediatr*. 2019;7:295. https://doi.org/10.3389/fped.2019.00295.
63. Thrasher AJ, Williams DA. Evolving gene therapy in primary immunodeficiency. *Mol Ther*. 2017;25:1132–1141.
64. Grosser R, Cherkassky L, Chintala N, Adusumilli PS. Combination immunotherapy with CAR T cells and checkpoint blockade for the treatment of solid tumors. *Cancer Cell*. 2019;36:471–482.

2

Organization of the Immune System

Dorothy E. Lewis, Arthur W. Zieske, and Sarah E. Blutt

The human immune system consists of organs (*e.g.*, spleen, thymus, and lymph nodes) and movable cells (*e.g.*, from the bone marrow, blood, and lymphatics). This design allows central locations for initial production and differentiation of committed cells from naïve precursors, as in the fetal liver, the bone marrow and the thymus; and more dispersed sites used for selection and further differentiation of cells into mature effector cells, as in spleen, lymph nodes and intestinal Peyer patches. This arrangement also allows regulated immune responses at locations peripheral to primary lymphoid organs to provide local control of infectious processes. The mechanisms responsible for the ability of nonspecific leukocytes, innate lymphoid cells (ILCs), natural killer (NK) cells, and antigen-specific T and B lymphocytes to respond rapidly to challenge are discussed in later chapters. This chapter covers the basic features and the ontogeny of cells involved in the immune response. It also reviews the essential structure of lymphoid organs and sites of organized immune cells, including skin, large intestine, and adipose tissues. Immune cell development.

IMMUNE CELL DEVELOPMENT

Ontogeny of the Cells of the Immune System

In humans, in the first three weeks of embryogenesis, primitive hematopoiesis begins from the primitive streak and is limited to erythroid, macrophage, and megakaryocytic lineages. The yolk sac becomes the major source of erythropoietic islands physically associated with, but not inside, the embryo.[1] Specialized endothelial cells give rise to the first progenitor cells. The cells produced in primitive hematopoiesis are phenotypically and functionally different from those that develop later. Definitive hematopoiesis begins in the aorta–gonad–mesonephros (AGM) where the first hematopoietic stem cells (HSCs) are formed.[2] The placenta is also a source of fetal stem cells for both the AGM and the fetal liver.[3] Properties of HSC differ by site. For example, HSCs in the fetal liver are in cycle whereas those in adult bone marrow are not. The progenitor stem cells that first populate the embryonic liver begin blood cell production in the sixth week of gestation, or just after the organ can be identified. By the eleventh week, the liver is the major source of hematopoiesis and remains so until the sixth month of gestation.[1]

The first progenitor cells derived from HSCs are colony-forming cells that can differentiate into granulocytes, erythrocytes, monocytes, megakaryocytes, and lymphocytes.[4] Subsequent to skeleton formation, between the second and fourth months of gestation, white blood cell development shifts to the bone marrow. The transition from liver to bone marrow is completed in the sixth month of gestation. Cells that differentiate from early stem cells begin to populate the primary lymphoid organs, such as the thymus, by 7 to 8 weeks of gestation.[5,6] At 8 weeks gestation, T-cell precursors that have initiated T-cell receptor (TCR) rearrangement (Chapter 9) can be detected in thymus tissue. In the fetal liver, B-cell precursors initiate immunoglobulin (Ig) rearrangements by 7 to 8 weeks gestation (Chapter 7). Late in the first trimester, B-cell development spreads to the bone marrow where B-cell progenitors congregate in areas adjacent to the endosteum and differentiate in the direction of the central sinus. The association of B cells with stromal reticular cells is essential for eventual release of mature B cells into the central sinus. As in the case of T-cell development, a selection process occurs such that many B-cell progenitors die by apoptosis.

In adult humans, the bone marrow is the chief source of stem cells. However, stem cells with different characteristics and limited self-renewal can be induced into the peripheral blood via injection of granulocyte colony stimulating factor (G-CSF) (Chapter 14). Methods to increase HSC self-renewal and expansion are under intensive study, including a search for means to promote *in vitro* expansion.[7] There has been an explosion in information with regard to how HSC develop and maintain self-renewal *in vivo*, including how fetal HSCs migrate to a niche from which adult HSC are derived. A key idea is that there is a nonlinear continuum of development that allows individual HSCs to "sidestep" a designated pathway depending on cytokine cues in their niches.[3] In the bone marrow of aged humans, there is evidence for a myeloid predominance with a restricted diversity of HSCs.[8]

Tools Essential to an Understanding of Immune Cell Biology

Progress beyond morphologic categorization of hematopoietic cells was enhanced by the use of monoclonal antibodies that identify stage-specific leukocyte cell surface antigens and the use of flow cytometry. In the 1980s, the sheer number of monoclonal antibodies to human leukocyte antigens resulted in a complicated nomenclature. In response, leukocyte differentiation antigen workshops developed a more consistent naming system. The workshops grouped monoclonal antibodies that recognized a single molecule on leukocytes by the cluster pattern of cells with which they were identified, hence the term CD, or "cluster of differentiation" antigen (Table 2.1). No formal conferences have been held since 2010, but new CD antigens have been validated. As of 2020, 371 had been identified (https://www.uniprot.org/docs/cdlist.txt) (Appendix 1).

TABLE 2.1 Important Cell Surface Antigens on Hematopoietic Cells.

Cell type	Surface Antigens	Predominant Location
Hematopoietic Stem Cells		
Bone marrow HSC	$CD34^+$, Lin-, 90^+	Bone marrow
Peripheral blood HSC	$CD34^+$, Lin-, $CD38^+$, $CD71^+$	Blood
Myeloid Cells		
Monocytes	CD14, CD35 (CR1), CD64	Blood
Macrophages	(FcRrγ1)CD68, CD13	Tissue
Langerhans cells	CD64,CD35	Skin
Follicular dendritic cells	CD1a, CD207 (Langerin),	B-cell areas, lymph nodes
Interdigitating dendritic cells	CD35, CD64	T-cell areas, lymph nodes
Myeloid dendritic cells	CD21, CD35 FcγRIIb	Mainly tissues
Plasmacytoid dendritic cells	CD80, CD56, Class II CD83	
(IFN-α producing)	CD40	
	CD83, CD80, CD86, CD40	
	CD1a, CD11c	
	CD4, CCR5, CXCR4, CD123	
Granulocytes		
Neutrophils	CD16 (FcγRIII), CD35 CD88 (C5aR)	Blood, tissues
Eosinophils	CD32 (FcγRII)	Blood, tissues
Basophils	CD23, (FcεRII), CD32	Tissues, blood
Mast cells	FcεRI α	Tissues, blood
		Tissues
Lymphocytes		
T cells	CD7, CD3, CD4, CD8, CD28	Thymus, spleen, lymph nodes, MALT, blood
B cells	Surface Ig, class II,	Bone marrow, spleen, lymph nodes, MALT, blood
NKT cells	CD19, CD20, CD22, CD40	Spleen, lymph nodes, mucosal tissues, blood
Tregs	CD16, CD56, CD94	Blood, tissues
Th17	CD3, CD56, Vα24 TCR	Thymus, blood, tissues
Tfh	CD4, CD25, Foxp3, GARP	Intestine, blood, tissues
	CD4, CCR6, IL-17,	Germinal centers of lymph
	RoRγ	nodes
	CD4, ICOS, PD-1, Bcl-6	

HEMATOPOIESIS AND LYMPHOPOIESIS

All mature cells of the hematopoietic and lymphoid lineages are derived from pluripotent stem cells that produce progenitors for lineage-specific cells.[9] Hematopoietic progenitors mature into cells of the granulocytic, erythroid, monocytic–dendritic, and megakaryocytic lineages (GEMM colony-forming units, CFU-GEMM). Likewise, lymphoid progenitors mature into B lymphocytes, T lymphocytes, and innate lymphoid cells including NK cells (Fig. 2.1).

In the adult, hematopoiesis and lymphopoiesis occur in two distinct tissues. The development of hematopoietic lineage cells (*i.e.*, granulocytes, monocytes, dendritic cells, erythrocytes, and platelets) occurs in bone marrow (Table 2.2). B-lymphocyte development through the immature and transitional B cell stages also occurs in the bone marrow (Chapter 7). T cell progenitors arise in the bone marrow but then migrate to the thymus where they differentiate into γδ and αβ T cells, including regulatory αβ T cells (Chapter 9). Some NK cells develop from precursors in the thymus.[10] Other tissue specific NK cells may develop in the bone marrow, lymph nodes, or the uterus.[10]

TABLE 2.2 Normal Distribution of Hematopoietic Cells in the Bone Marrow

Cell Type	Approximate Proportion (%)
Stem cells	1
Megakaryocytes	1
Monocytes	2
Dendritic cells	2
Lymphocytes	15
Plasma cells	1
Myeloid precursors	4
Granulocytes	50–70
Red blood cell precursors	2
Immature and mature red blood cells	10–20

Characteristics of Hematopoietic Stem Cells

The pluripotent stem cell gives rise to all red and white blood cell populations. Human HSCs in the bone marrow are rare: 1 in 10,000 cells. They are found in two proposed niches of the bone marrow that are closest to the bone. One niche also contains osteoblasts (endosteal niche), and the other is associated with the sinusoidal endothelium (vascular niche). Many human HSCs closely associate with perivascular mesenchymal stem cells.[4] It is estimated that human long-term HSC divide once or twice per year. Quiescent HSCs tend to be found near arterioles in the endosteum. Actively dividing HSCs are more likely to be located near sinusoid regions close to central veins.

Many separation methods are used to identify stem cells and their differentiation potential. Early observations showed that HSCs had characteristic flow cytometric light-scattering properties (low side scatter, medium forward scatter), no lineage

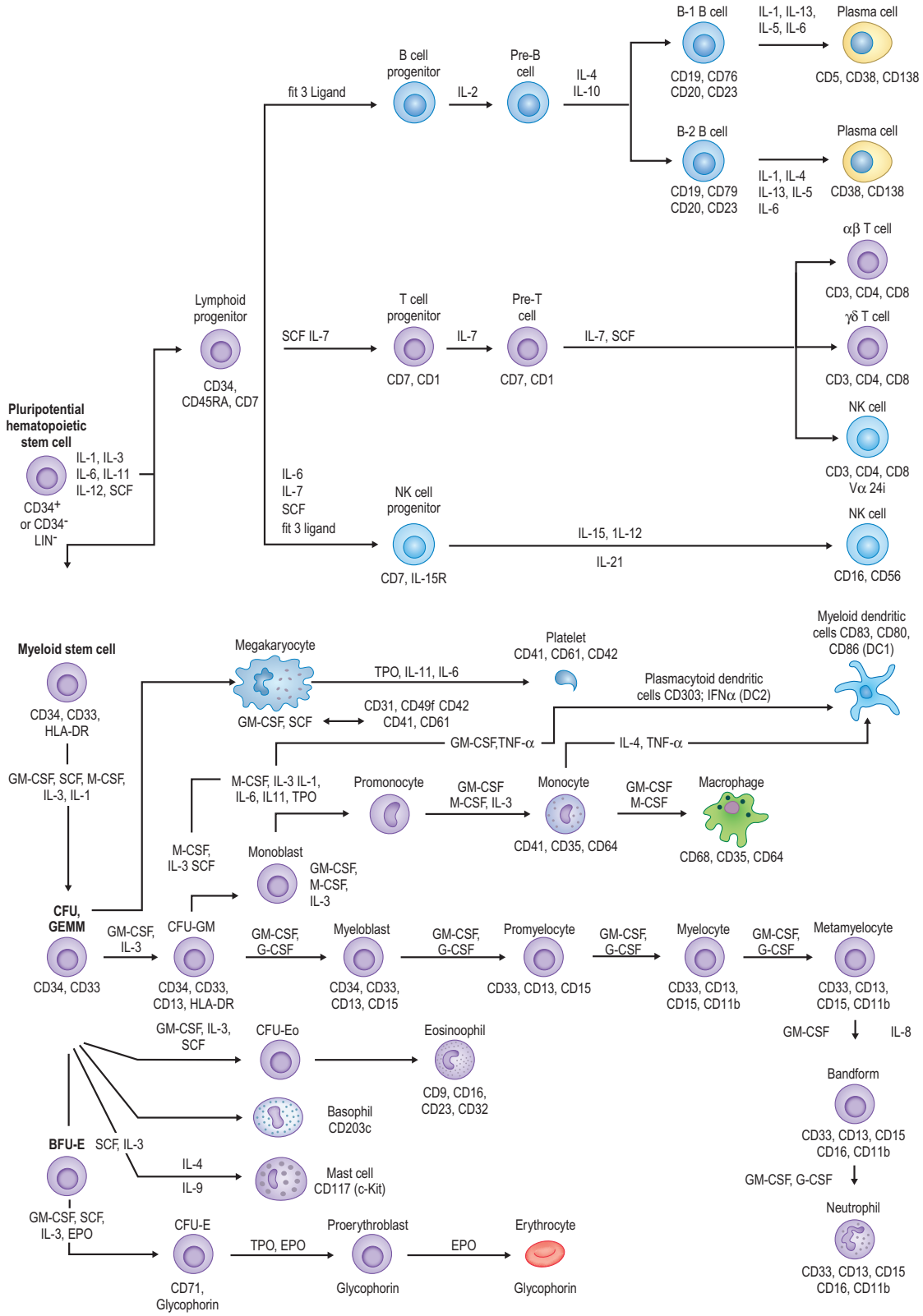


FIG. 2.1 Differentiation pathways for hematopoietic cells.

(LIN)-specific markers (*e.g.*, CD2, CD3, CD5, CD7, CD14, CD15, or CD16), and expressed CD34 on the cell surface. Definitive markers of human HSC are CD34 as well as CD90 and CD49f.[6] As HSC go out of quiescence to become lineage directed, they lose expression of both CD90 and CD49f before they become committed to the myeloid or lymphoid lineage. The lymphoid precursor expresses CD10, CD45Ra, whereas the myeloid precursor expresses CD135. Transcription factors are unique for each population. For the HSC, these include SOX8, SOX18, and NFIB. Out of quiescence, the HSC express MYC and IKZF1. Key signaling pathways include Notch and Wnt/β-catenin.[4]

A long-lived stem cell has capacity for self-renewal via asynchronous division that can be influenced by external factors, including infection.[11] Hematopoietic stem cells circulate in the peripheral blood with 10 to 100 times less frequency than their differentiated progeny. Mobilization of "stem cells" to the periphery is induced with G-CSF. Of these induced cells, although most are Lin^-, only 5% to 20% are true stem cells.[11] Most peripheral blood-induced stem cells express activation antigens (*e.g.*, transferrin receptor, CD71 and CD38). Peripheral blood HSCs engraft 2 to 3 days faster than conventional bone marrow HSCs, and thus can reduce bone marrow transplantation morbidities. However, peripheral blood HSCs are more differentiated than bone marrow HSCs with limited self-renewal.

Regulation of Hematopoietic and Lymphopoietic Cell Growth and Differentiation

Regulation of stem cell differentiation occurs through interactions with a variety of micro-environmental factors in the bone marrow or thymus. Cell surface receptors recognize either soluble ligands (*e.g.*, cytokines) released by other cells, or surface ligands (*e.g.*, cell interaction molecules) expressed on adjacent cells. These receptors can facilitate differentiation. Stem cells can be exposed to spatially and temporally regulated ligands or factors. The differential expression of receptors on the stem cells allows control of proliferation and differentiation along one of the hematopoietic or lymphoid lineages.[6]

Cytokines (Chapter 14) have pleiotropic effects on hematopoietic and lymphoid cell development. They affect both growth and maintenance of pluripotent stem cells, as well as the development and differentiation of specific cell lineages. The effect of the cytokine often differs, depending on whether the cell has previously been or is concurrently being stimulated by other cytokines. The stage of differentiation, as well as the presence or absence of the cytokine's receptor on the cell surface, also affects the cellular response. Although there are several cytokines, such as IL-6 and stem cell factor (SCF), that are classically involved in hematopoiesis, these cytokines have non-hematopoietic functions, as well.

Stromal cells located within the bone marrow and thymus regulate hematopoietic and lymphoid cell growth and differentiation by releasing cytokines. These include the interleukins IL-4, -6, -7, and -11; leukemia inhibitory factor (LIF); granulocyte–macrophage colony-stimulating factor (GM-CSF); granulocyte colony-stimulating factor (G-CSF); and SCF.[4,12] Stromal cells also participate in cell–cell interactions with progenitors that express fibroblast growth factor 1 (FGF-1) and FGF-2 that support HSC expansion. In addition, stromal cells form the intercellular matrix (*e.g.*, fibronectin and collagen) that binds to selectin and integrin receptors present on hematopoietic and lymphoid progenitors.[13]

Cytokines That Affect the Growth and Maintenance of Pluripotent and Multipotent Stem Cells

Pluripotent stem cells can reconstitute cells of the hematopoietic and lymphoid lineages. Maintenance of pluripotent capacity is mediated through factors that keep HSC quiescent. These include c-kit, N-cadherin, osteopontin, TGF-β and Wnt signaling. There are also factors that have a negative effect on quiescence, such as Hedgehog signaling and notch ligands, Delta and Jagged.[4] As progeny differentiate and the stem cell pool is depleted, a low level of stem cell proliferation is required. The entry of stem cells into the cell cycle and subsequent proliferation, as well as commitment to particular lineages, is controlled by cytokines and transcription factors. Data suggest that flt-3 ligand, c-kit ligand, and megakaryocyte growth and development factor (MGDF) all promote long-term stem cell expansion. The combination of c-kit ligand, IL-3 and -6 causes more rapid expansion, but does not allow long-term extension of precursor cells.[12]

Several cytokines, either alone or in combination, promote stem cell growth (Table 2.3).[5] Combinations of cytokines are more effective at inducing stem cell growth. For example, IL-1 promotes stem cell growth by inducing bone marrow stromal cells to release additional cytokines and by synergistically stimulating these cells in the presence of other cytokines. One of these other cytokines, IL-3, promotes the growth of hematopoietic progenitors. The effect is significantly enhanced by the addition of IL-6, IL-11, G-CSF, and SCF. IL-11, a stromal cell-derived cytokine, enhances IL-3-induced colony formation in 5-fluorouracil-resistant stem cells. Similarly, other cytokines secreted by stromal cells (*e.g.*, IL-6, G-CSF, and SCF) also exert their effects by shortening the G0 period in stem cells. In contrast, IL-3 acts on cells after they have left G0. Either by itself or in conjunction with IL-11 or SCF, IL-12 is unable to support the growth of primitive hematopoietic stem cells. However, IL-12 acts in synergy with IL-3 and IL-11, or IL-3 and SCF, to enhance stem cell survival and growth.

Depending on the circumstance, cytokines may either enhance or inhibit the growth or differentiation of hematopoietic and lymphoid cells. For example, IL-6 participates in the development of neutrophils, macrophages, platelets, T cells, and B cells. Thrombopoetin signaling promotes stem cell self-renewal to increase transplantation success.[14] The effects of individual cytokines can be altered when they function in combination. Together, GM-CSF and IL-3 promote development of granulocytes, macrophages, dendritic cells and erythrocytes. In the presence of IL-3, IL-6, and GM-CSF the LIF cytokine can enhance the growth and development of $CD34^+$ bone marrow progenitor cells along multiple lineages. However, in their absence, LIF has little effect. Similarly, while transforming growth factor-β (TGF-β) and IL-4 are potent inhibitors of hematopoietic progenitor cell growth, they enhance granulocyte development. Conversely, tumor necrosis factor-α (TNF-α) inhibits development of granulocytes, but potentiates the effects of IL-3 on hematopoietic progenitor cell proliferation.

Cytokines That Inhibit Hematopoietic Stem Cell Growth

Cytokines produced by mature cells may also downregulate hematopoietic stem cell growth. For example, macrophage inflammatory protein-1α (MIP-1α) can inhibit hematopoietic progenitor cell proliferation, interferon-γ (IFN-γ) and TGF-βpromote terminal differentiation, and TNF-α can induce apoptosis.

TABLE 2.3 Cytokines Important for Hematopoietic Cell Growth and Differentiation

Cytokines	Stem Cells	Thymocytes	B Cells	NK Cells
IL-1	Acts on stromal cells	Differentiation		
IL-2		Pleomorphic	Proliferation	Proliferation
IL-3	Proliferation			
IL-4		Pleomorphic	Promotes (low) Prevents (high)	Inhibits IL-2
IL-5			Proliferation/differentiation	
IL-6	Shortens G_0	Enhances stimulation	Maintains potential	Enhances IL-2
IL-7		Survival/proliferation	Proliferation of pro- and pre-B cells	Activation
IL-10			Survival	
IL-11 (Megakaryocyte development)	Shortens G_0		Maintains potential	
IL-12	Survival			Activation proliferation
IL-13			Activation/division of mature B cells	
IL-15			Proliferation	Development/survival
IL- 21			Proliferation	Expansion
SCF/c-*kit*	Survival	Atrophy	Maintains potential	Expansion
G-CS	Shortens G_0		Maintains potential	
FLt3 ligand	Growth factor		Increases proliferation	Expansion
SDF1-α		Proliferation/regeneration	Chemoattractant	
LIF	Proliferation	Atrophy		
Thrombopoietin	Expansion/regulates self-renewal			
TNF-α	Proliferation: inhibits granulocytes			
TGF-β	Inhibits growth enhanced granulocytes			
MIP-1α	Inhibits			
NGF			Proliferation/differentiation	Expansion

When pathologic conditions exist, these cytokines can have adverse effects on hematopoietic and lymphoid cell development, resulting in various deficiency states.

Cytokines Affecting Development and Differentiation of Specific Cell Lineages

The initial event in differentiation involves the commitment of pluripotent stem cells to a specific lineage. Cytokines are important for this process and appear to have lineage-specific effects that act at particular late stages of differentiation. For example, erythropoietin regulates the later stages of erythrocyte differentiation, whereas G-CSF induces granulocyte differentiation and macrophage colony-stimulating factor (M-CSF) is specific for macrophage maturation.[15]

KEY CONCEPTS

Cells of the Immune System

1. Pluripotent stem cells in bone marrow give rise to all lineages of the immune system as well as platelets and red blood cells.
2. Development and regulation of cells of the immune system is marked by the programmed appearance of specific cell surface molecules and responsiveness to selective cytokines. Most of these cell surface markers have been given a cluster of differentiation (CD) designation.
3. Mature cells of the immune system include antigen-presenting cells (APCs), phagocytic cells, including neutrophils, eosinophils, and basophils, and lymphocytes, including natural killer cells and T, B, or innate lymphoid cells (ILCs).
4. APCs include monocytes, macrophages, dendritic cells, endothelial cells, epithelial cells, and adipocytes. B lymphocytes can also function as potent APCs. APCs can direct the differentiation and function of both innate and acquired immune cells.
5. Polymorphonuclear (PMN) granulocytes are important in the early response to stress, tissue damage, and pathogens. These phagocytes include neutrophils, eosinophils, and basophils.
6. Lymphocyte lineages have discrete subpopulations with specialized functions. These include CD4 and CD8 T cells, T-helper (Th) subsets including T regs, Th9, Th17 and Tfh cells, innate lymphoid cells, and the B-1, conventional B-2, and marginal zone (MZ) B cells.

Mature Cells of the Immune System

The mature cells of the immune system arise mainly from progenitor cells in the bone marrow. They include both nonspecific and antigen-specific effector cells. The central player in both lines of defense is the antigen presenting cell (Chapter 6). In addition to their nonspecific effector functions, these cells are crucial for the development of specific immune responses. With maturation, these cells enter the blood (Table 2.4) where they circulate into the tissues and organs.

Antigen Presenting Cells

APC are found primarily in the solid lymphoid organs and skin (Chapter 23). The frequency of these in tissues varies between 0.1% and 1.0%. Specialized APC in B cell areas of lymph nodes and spleen are termed follicular dendritic cells (FDCs) because they have dendrites, not because they are related to other dendritic cell types. FDCs trap the antigen–antibody complexes that are important for the generation and maintenance of memory B cells. These cells do not express class II molecules, rather they have receptors for IgG and complement component C3b. These are FcγR (CD64) and CR1 (CD35), respectively.

TABLE 2.4 Normal Distribution of White Blood Cells in the Peripheral Blood of Adults and Children

	Approximate Percentage		Range Of Absolute Counts (no./μL)	
Cell type	Adults	Children (0–2 Years)	Adults	Children (0–2 Years)
Monocytes	4–13		400–1000	
Dendritic cells	0.5–1	ND[a]	30–170	ND
Granulocytes	35–73		2500–7500	1000–8500
Lymphocytes	15–52	34–75	1450–3600	3400–9000
As % of lymphocytes				
T cells	75–85	53–84	900–2500	2500–6200
CD4 cells	27–53	32–64	550–1500	1300–4300
CD8 cells	13–23	12–30	300–1000	500–2000
B cells	5–15	06–41	100–600	300–3000
NK cells	5–15	03–18	200–700	170–1100

[a]Not determined.

Child data adapted from Shearer W, Rosenblatt H, Gelman R, et al. Lymphocyte subsets in healthy children from birth through 18 years of age: the pediatric AIDS Clinical Trials Group P1009 study. *J Allergy Clin Immunol.* 2003;12:973–980.

Monocytes-Macrophages

Cells of the monocyte–macrophage lineage comprise approximately 10% of the white cells in the blood. They exist primarily as monocytes, which are large 10- to 18-μm cells with peanut-shaped, pale purple nuclei as determined by Wright's staining (see Table 2.4). The cytoplasm, which is 30% to 40% of the cell, is light blue and has azurophilic granules that resemble ground glass with intracytoplasmic lysosomes. These cells express MHC class II molecules, CD14 (the receptor for lipopolysaccharide), and distinct Fc receptors (FcR) (Chapter 8) for Ig. The latter include FcγRI (or CD64), which has a high affinity for IgG, and FcγRII (or CD32), which is of medium affinity and binds to aggregated IgG. FcγRIII (or CD16) has low affinity for IgG and is associated with antibody-dependent cellular cytotoxicity (ADCC). It is expressed on macrophages, but not on blood monocytes. Monocytes and macrophages also express CD89, the Fc receptor for IgA.[15]

Macrophages are more differentiated monocytes that reside in various tissues, including lung, liver, and brain.[16] Cells of the monocyte–macrophage lineage can be depleted or isolated because they adhere strongly to glass or plastic surfaces. However, this process activates them, which can confound functional studies. Many cells of this lineage phagocytose organisms or tumor cells *in vitro*. Cell surface receptors, including CD14, Fcγ receptors and CR1 (CD35), are important in opsonization and phagocytosis. Cells of this lineage also express MHC class II molecules, and some express the low-affinity receptor for IgE (CD23). Other cell surface molecules include myeloid antigens CD13 (aminopeptidase N) and CD15 (Gal (1–4) or [Fuc (1–3)] GlcNAc) and the adhesion molecules CD68 and CD29 or CD49d (VLA-4). Classical blood monocytes in humans (85%) express high levels of CD14, but no CD16. By contrast, nonclassical monocytes express less CD14, but more CD16. This later subset produces much more IL-12, TNF-α, and IL-1β. In addition to phagocytic and cytotoxic functions, these cells have receptors for various cytokines such as IL-4 and IFN-γ that can regulate their function. Activated macrophages are also a major source of cytokines, as well as complement proteins and prostaglandins.

Macrophages, along with DCs, have been shown to be more plastic in differentiation and function than previously realized. An alternative activation pathway induced by T helper cell 2 (Th2) (Chapter 11) cytokines, IL-4 and Il-13,[16] promotes suppressive anti-inflammatory properties that can reduce responses to cancer as well as maintaining adipose integrity. This macrophage subtype is seen among tumor-associated suppressive macrophages.

Monocytes and macrophages arise from colony-forming unit granulocyte–monocyte (CFU-GM) progenitors that differentiate first into monoblasts, then promonocytes, and finally monocytes.[15] Mature monocytes leave the bone marrow and circulate in the bloodstream until they enter tissues, where they develop into tissue macrophages (alveolar macrophages, Kupffer cells, intestinal gut macrophages and microglial cells). There is evidence suggesting that tissue macrophages actually originate from fetal macrophages and thus seed tissues early in fetal development where they are maintained by longevity and slow self-renewal.[17] Tissue resident macrophages and those derived from monocytes act together to combat infection.[18]

Several cytokines participate in the development of monocytes and granulocytes. For example, SCF, IL-3, IL-6, IL-11, and GM-CSF promote development of myeloid lineage cells from $CD34^+$ stem cells, especially in early stages of differentiation. Another cytokine, M-CSF, acts at the later stages of development and is lineage specific, inducing maturation into macrophages.[15]

Dendritic Cells

Dendritic cells are accessory cells that express high levels of MHC class II molecules and are potent inducers of primary T-cell responses. Except for the bone marrow, they are found in virtually all primary and secondary lymphoid tissues, as well as in skin, mucosa, and blood. DCs are also abundant in the thymus medulla for selection of thymocytes.

Dendritic cells, macrophages, and granulocytes are derived from $CD34^+$ MHC class II-negative precursors present in the bone marrow. GM-CSF and TNF-α are involved in development of DCs from their precursors in bone marrow.[19–21] DCs residing in peripheral sites such as the skin, intestinal lamina propria, lung, genitourinary tract, etc. are typically immature. These cells are more phagocytic and display lower levels of MHC class I, MHC class II, and co-stimulatory molecules. These immature dendritic cells take up antigens in tissues for subsequent presentation to T cells. As they migrate, they mature to become efficient APC.

The predominant APCs of the skin are the Langerhans cells[22] of the epidermis. These cells are characterized by rocket-shaped granules called Birbeck granules. Immature tissue DCs in peripheral tissues engulf and process antigen and home to T-cell areas in the draining lymph nodes or spleen.[23] Mature DCs can directly present processed antigens to resting T cells to induce proliferation and differentiation, a key functional difference between mature DCs and macrophages. The effector cells produced after this presentation then home to the site of the antigenic assault.

TNF-α maintains viability of Langerhans cells in the skin and stimulates their migration. In Peyer patches (Chapter 24), immature dendritic cells are located in the dome region underneath the follicle-associated epithelium (FAE) where they actively endocytose antigens taken up by M cells in the FAE. More mature interdigitating DCs are found in T cell regions. These cells, like their counterparts in the lungs, induce Th2 immune responses. There are at least three types of DC. cDC1 derive from bone marrow,

are found in lymphoid tissues, and express CD141, CLEC-9a and XCR-1. These are similar to murine CD8 α/α DCs in function. cDC2 express CD1c and CD171a. The third type, plasmacytoid DC, are high producers of IFN-α. They express CD123, CD303, CD304 and low levels of CD11c.[21]

All three of the DC types can be derived from either myeloid or lymphoid lineages and are extremely plastic.[19,20] Antigen-presenting cells, especially dendritic cells, are largely influenced by stimulation with Toll-like ligands found on a variety of stimuli, which then direct differentiation and function of innate and acquired immune cells.

Polymorphonuclear Granulocytes

Polymorphonuclear (PMN) granulocytes arise from progenitors that mature in the bone marrow. They are essentially end-stage cells. Their life span after release varies from a few days to 5 to 6 days and is regulated by responses required. PMN constitute 65% to 75% of the white blood cells in the peripheral blood, are 10 to 20 μm in diameter, and are characterized by a multilobed pyknotic nucleus characteristic of cells undergoing apoptosis (see Table 2.4).[24]

PMN cells are also found in tissues. They use diapedesis to gain access from the blood. Granulocytes act as early soldiers in the response to stress, tissue damage, or pathogen invasion. Because of their function in phagocytosis and killing, they possess granules whose unique staining characteristics categorize the cells as neutrophils (Chapter 39), basophils (Chapter 44), or eosinophils (Chapter 45).

Neutrophils. Most circulating granulocytes are neutrophils (90%). Their granules are azurophilic and contain acid hydrolase, myeloperoxidase, and lysozymes. These granules fuse with ingested organisms to form phagolysosomes, which eventually kill the invading organism. In some cases, there is extracellular release of granules after activation via the Fc receptors. Neutrophils express a number of myeloid antigens, including CD13, CD15, CD16 (FcγRIII), and CD89 (FcαR). In response to bacterial infection, there is typically an increase in the number of circulating granulocytes. This often includes the release of immature granulocytes, called band or stab cells, from the bone marrow. In a mild infection, both the number and function of neutrophils are increased. This is associated with a delay in apoptosis. With a more severe infection, there may actually be impairment of function owing to the release of immature cells. Neutrophils form extracellular traps (NETs) that capture microbes and use autophagy to digest them intracellularly.[25] Out of control NET formation (Netosis) occurs in sepsis to enhance tissue destruction. Neutrophils may also act as APC.[26]

Neutrophils derive from CFU-GM progenitor cells and differentiate within a 10- to 14-days period. These progenitors give rise to myeloblasts, which in turn differentiate into promyelocytes, myelocytes, and finally mature neutrophils. The cytokines SCF, IL-3, IL-6, IL-11, and GM-CSF promote the growth and development of neutrophil precursors, whereas certain cytokines are important for differentiation of CFU-GM progenitors into mature neutrophils.[27] For example, G-CSF induces maturation of neutrophil precursors into mature neutrophils. IL-4 enhances neutrophil differentiation induced by G-CSF, while at the same time inhibiting development of macrophages induced by IL-3 and M-CSF.

Eosinophils. Eosinophils typically comprise 2% to 5% of the white cells in the blood. They exhibit a unique form of diurnal variation with peak production at night. This is perhaps due to lower glucocorticoid levels at night. Eosinophils are capable of phagocytosis followed by killing, although this is not their main function. The granules in eosinophils are much larger than in neutrophils and are actually membrane-bound organelles. The crystalloid core of the granules contains a large amount of major basic protein (MBP). MBP can neutralize heparin and is toxic. During degranulation, the granules fuse to the plasma membrane, and their contents are released into the extracellular space. Organisms that are too large to be phagocytosed, such as parasites, can be exposed to cell toxins by this mechanism. MBP can damage schistosomes *in vivo*, but damage is minimized because the MBP is confined to a small extracellular space (Chapter 29). Eosinophils also release products that counteract the effects of mast cell mediators. Whether eosinophils are absolutely required for helminth control is controversial, and more study is needed (Chapter 30).

Eosinophils derive from a progenitor (CFU-Eo) that progresses through developmental stages similar to those of neutrophils.[28] These stages begin with an eosinophilic myeloblast, followed by an eosinophilic promyelocyte, a myelocyte, and finally a mature eosinophil. Three cytokines are important in the development of eosinophils: GM-CSF, IL-3, and IL-5. GM-CSF and IL-3 promote eosinophil growth and differentiation; SCF also has an effect on eosinophil function. The chemokine eotaxin (CCL11) (Chapter 15) also promotes eosinophilia. IL-5 has more lineage-specific effects on eosinophil differentiation, although it also affects some subsets of T and B cells. IL-5 is also essential for eosinophil survival and maturation. These cells are recruited into inflammatory tissue via chemokines that target CCR3. Eosinophils are involved in the pathophysiology of asthma (Chapter 43) by contributing to airway dysfunction and tissue remodeling. IL-5 is a current target to control eosinophilia.

Basophils and mast cells. Basophils represent less than 1% of the cells in the peripheral circulation, and have characteristic large, deep-purple granules. Mast cells are found in proximity to blood vessels and are much larger than peripheral blood basophils. The granules are less abundant, and nucleus is more prominent. There are two different types of mast cells, designated mucosal or connective tissue depending on their location.[29] Mucosal mast cells require T cells for their proliferation, whereas connective tissue mast cells do not. Both types have granules that contain effector molecules. After degranulation, which is effected by cross-linkage of cell surface IgE bound to cells via the high-affinity receptor for IgE, the basophils–mast cells release heparin, histamine, and other effector substances to mediate an immediate allergic attack (Chapter 46).

Since basophils and mast cells share a number of phenotypic and functional features, they have been considered to share a common precursor. They both contain basophilic-staining cytoplasmic granules, express the high-affinity IgE receptor (FcεRI), and release a number of similar chemical mediators that participate in immune and inflammatory responses, particularly anaphylaxis. They both have been implicated in allergic inflammation and in fibrosis. However, basophils and mast cells have distinct morphologic and functional characteristics that suggest they are distinct lineages of cells, rather than cells at different stages within the same lineage. Analysis of human transcription factors places basophils closer to eosinophils than to mast cells.[30]

Basophils mature from a progenitor (CFU-BM) into basophilic myeloblasts, then basophilic promyelocytes, myelocytes, and finally mature basophils. Less is known about the stages of

mast cell development, although they are probably derived from the same CFU-BM progenitor as basophils.[31,32]

In humans, SCF induces the most consistent effects on the growth and differentiation of both basophils and mast cells. Both IL-3 and SCF are important for intestinal mast cell differentiation. IL-6 can also increase mast cell numbers. This probably explains why T cells are needed for their development.[29] While both IL-4 and -9 stimulate mast cell development in mice; in humans, only IL-9 acts in synergy with SCF to enhance mast cell growth.[30] Additional cytokines that affect basophil growth include nerve growth factor and GM-CSF or TGF-β, and IL-5 for basophil differentiation.

Platelets and Erythrocytes

Hematopoietic stem cells also give rise to platelets and erythrocytes. Platelets are necessary for blood clot formation and mediate a number of immune functions. Mature RBCs are necessary for oxygen delivery to tissues.[31] Platelets derive from CFU-GEMM progenitors, which in turn differentiate into burst-forming units for megakaryocytes (BFU-MEG). The BFU-MEG then differentiate into CFU-MEG, promegakaryoblasts, megakaryoblasts, megakaryocytes, and finally platelets.[33] Several cytokines, particularly thrombospondin, IL-1, IL-3, GM-CSF, IL-6, IL-11, and LIF, affect the growth and differentiation of platelets. Platelets are a main producer of exosomes, small 80 micron membranous vesicles with cell surface markers characteristic of their cell source, as well as mRNA encoding for cell-related molecules. Exosomes travel in the blood and the lymph and may be a significant way that cells function at a distance. Other immune cell types, including dendritic cells can transfer information via exosomes.

While erythrocytes also derive from CFU-GEMM progenitors, their progenitors are burst-forming units for erythrocytes (BFU-E), which in turn differentiate into CFU-E, pronormoblasts, basophilic normoblasts, polychromatophilic normoblasts, orthochromic normoblasts, reticulocytes, and finally erythrocytes. Again, several cytokines, notably GM-CSF, SCF, IL-9, thrombospondin, and erythropoietin, regulate erythrocyte development.

Lymphocytes

Lymphocytes, the central cell type of the specific immune system, represent about 25% of white cells in the blood (see Table 2.4). Small lymphocytes range between 7 and 10 μm in diameter. They are characterized by a nucleus that stains dark purple with Wright stain, and by a small cytoplasm. Large granular lymphocytes range between 10 and 12 μm in diameter and contain more cytoplasm and scattered granules. The three types of lymphocytes that circulate in the peripheral blood—T, B, and ILCs, including NK cells cells—constitute approximately 80%, 10%, and 10% of the total blood lymphocyte population, respectively (Chapters 7, 9, and 12). In the thymus, most of the lymphocytes (90%) are T cells. However, in the spleen and lymph nodes, only about 30% to 40% are T cells. The preponderant lymphocytes in these locations are B cells (60% to 70%).[34,35]

T cells. T lymphocytes arise from lymphocyte progenitors in the bone marrow that are committed to the T-cell lineage even before they move to the thymus. Early in embryogenesis, T-cell precursors migrate to the thymus in waves.[36,37] Associated with this migration is the developing ability of thymic education elements, epithelial cells, and DCs to select appropriate T cells.[37] In the thymus, T cells rearrange their specific antigen receptors (TCR) and then express CD3 along with the TCR on their surface (Chapter 9).

Resting T cells in the blood typically range between 7 and 10 μm in diameter and are agranular, except for the presence of a structure termed a Gall body, which is not found in B cells (see Table 2.4). The Gall body is a cluster of primary lysosomes associated with a lipid droplet. A minority of T cells in the blood (about 20%) are also of the large granular type, meaning that they are 10 to 12 μm in diameter and contain primarily lysosomes that are dispersed in the cytoplasm. Golgi apparatus also are found.

The preponderant form of the TCR, found on about 95% of circulating T cells, consists of α and β chains ($\alpha\beta TCR^+$).[38] Some $CD3^+$ cells do not express either CD4 or CD8 (double-negative or DN) and are characterized by having an alternative TCR composed of γ and δ chains ($\gamma\delta TCR^+$). Further differentiation in the thymus occurs from $CD3^+$ cells that express both CD4 and CD8 (double-positive or DP) T cells expressing either CD4 or CD8 but not both. These mature cells then circulate in the peripheral blood at a ratio of about 2:1 (CD4:CD8) and populate the lymph nodes, spleen, and other secondary lymphoid tissues.

T-cell progenitors, which are $CD7^+$, arise in the bone marrow from a multipotential lymphoid stem cell. After migration to the thymus, the $CD7^+$ progenitors give rise to a population of $CD34^+$, $CD3^-$, $CD4^-$, and $CD8^-$ T-cell precursors. These cells undergo further differentiation into mature T cells. Cytokines produced by thymic epithelial cells (*e.g.*, IL-1 and soluble CD23) promote differentiation into $CD2^+$, $CD3^+$ thymocytes (see Table 2.3). IL-7 induces the proliferation of $CD3^+$ DN ($CD4^-$ $CD8^-$) thymocytes, even in the absence of co-mitogenic stimulation. IL-7 is absolutely required for human T-cell development.[39]

IL-2 and -4 demonstrate complex effects on thymocyte development. Both can promote development of pro-thymocytes, as well as antagonizing their development. IL-6 acts as a co-stimulator of IL-1- or -2-induced proliferation of DN thymocytes and can stimulate the proliferation of mature, cortisone-resistant thymocytes alone. Once T cells leave the thymus, a variety of cytokines affect their growth and differentiation.

T cell subsets. T cells can be divided into subsets based on surface expression of CD4 and CD8, as well as by function in an immune response. CD4 and CD8 T cells were originally characterized by expression of the respective antigen and association with functional ability. For example, human T cells expressing CD4 provide help for antibody synthesis, whereas cells expressing CD8 develop into cytotoxic T cells. The distinction is better described as which antigen-presenting molecule is used for TCR interaction. Thus, CD4 T cells recognize antigen in the context of MHC class II molecules, and CD8 T cells recognize antigen presented by class I molecules (Chapter 6).

Memory T cells are divided based on expression of CD45R0, CCR7, CD28 and CD95 that categorize functions of cells as stem cell memory, central memory, transitional memory, effector memory and terminal effector cells. Early memory T cells have high lymph node homing and proliferation potential. Later stage T cells home to the periphery, are effector cells, and do not proliferate.[40,41]

T-helper (Th) cells mature in response to foreign antigens. Their function is dependent on the production of cytokine patterns, which characterize them as Th type 1 (Th1), Th2 or Th17.[39] The precursor Th cell first differentiates into a Th0 cell producing interferon-γ(IFN-γ) and IL-4. The cytokine environment subsequently determines whether Th1 or Th2 cells

predominate. Th1 cells produce primarily IFN-γ, IL-2, and TNF-α and are important in cell-mediated immunity to intracellular pathogens (Chapter 26), such as the tubercle bacillus. Th1 cells primarily use the T-bet transcription factor. Th2 cells produce predominantly IL-4, -5, -6, -10, and -13, as well as IL-2, and predominate in immediate or allergic type 1 hypersensitivity (Chapter 46) and primarily use the Gata-3 transcription factor. IL-33 also enhances production of Th2 cytokines.[39]

Other populations of CD4 T cells can develop and rely on IL-23 or -12 action upon the cells. If T cells are exposed to IFN-γ, they upregulate both IL-12R and -23R, which then produce either conventional Th1 cells or another subset, Th17, which produces IL-17 and is important for controlling immune cell activation in the GI tract (Chapter 24). Overactive function of this subset has been associated with autoimmunity. The Th17 population preferentially uses the ROR γ transcription factor.

T follicular helper cells are those classically determined to help B cell responses in germinal centers. They are $CD4^+$, $ICOS^+$, $PD\text{-}1^+$ and express the transcription factor Bcl-6. It is likely that there are other epigenetically altered T cells that allow diversity of function during an immune response.[41] Th9 cells express the transcription factor, PU.1 and secrete IL-9 and are produced in response to TGF-B and IL-4. These cells are increased in various immune-related diseases.[42]

A minor subpopulation (<5%) of $CD3^+$ cells in the peripheral blood express γδ TCR molecules. Most of these cells do not express CD4 or CD8. However, some intraepithelial lymphocytes that express γδ TCR also express CD8 αα homodimers in place of conventional CD8 αβ heterodimers (see Fig. 2.10). These cells, which are thymus independent, are involved in the initial response to bacterial antigens presented in mucosal epithelium. Another minor subpopulation of T cells, NKT cells (Chapter 12), can be $CD4^+$ or $CD8^+$ and express a single Vα chain, Vα24, which recognizes glycolipids in the context of CD1a rather than a classical MHC molecule (Chapter 6). NKT cells express MIP-1 α and β, have a Th1 bias, but lack Il-10 production. [43,44] The final subset is regulatory T cells (Treg) (Chapter 13), which occur naturally and can be induced *in vitro*. They are $CD4^+$ and express high levels of CD25 and the transcription factor Foxp3. Other definitive markers remain elusive.[45] Tregs are reduced in autoimmunity in adipose tissue during obesity and increased in cancer to aid in immunosuppression.

B cells and plasma cells. B cells represent 5% to 10% of the lymphocytes in the blood (see Table 2.4; Chapter 7). They are typically 7 to 10 μm in diameter and lack Gall bodies and granules. B cells express cell membrane immunoglobulin (mIg), the majority expressing both IgM and IgD.[46] The cytoplasm is characterized by scattered ribosomes and isolated rough endoplasmic reticulum (RER). The Golgi is not prominent unless the cells are activated. A small minority of B cells express either surface IgG or IgA. A number of other cell surface molecules are found on B cells, including CD19, CD20, CD23, CD40, CD72, CD79a and b, MHC class II, Fcγ RII receptors (CD32) and complement receptors C3b (CR1a; CD35) and C3d (CR2a; CD21). Similar to T cells, which surround the TCR with activation effector molecules, B cell mIg associates with CD19, CD21, and CD81 to enhance B cell activation (Chapter 7).

Upon activation and cross-linking of surface Ig by specific antigen, B cells undergo proliferation and differentiation to produce plasma cells. Plasma cells are non-dividing, specialized cells terminally differentiated from B cells, the function of which is to secrete Ig. They lose expression of mIg and MHC class II molecules. Plasma cells (10–15 μm) are not normally found in the blood. They display an eccentric nucleus and a basophilic cytoplasm with a well-developed Golgi. The plasma cell displays parallel arrays of expanded rough endoplasmic reticulum (RER) that contains Ig. B cell proliferation and differentiation processes take place in the germinal centers of the lymph nodes.

Several cytokines influence the development of B lymphocytes. *In vitro* studies of cytokines involved in the development of early B-cell progenitors show that combinations of SCF (but not IL-3) with IL-6, IL-11 or G-CSF can maintain B-lymphoid potential.[47] Stromal cell-dependent differentiation of fetal pro-B cells occurs in conjunction with Flk-2/flt-3 ligand and IL-7 and on several transcription factors, including PU.1, IKAROS, E2A, EBF, PAX5 and IRF8. Unlike in mice, in humans IL-7, is not absolutely required for B cell development.[48]

IL-4 has a variety of important effects on B-cell growth and differentiation. Low doses of IL-4 induce pre-B cells to differentiate into B cells expressing surface membrane IgM, whereas higher doses of IL-4 inhibit differentiation of B cells. In mature B cells, IL-4 increases expression of MHC class II, CD23, and CD40 molecules; promotes activation and progression to the G_1 stage of the cell cycle; enhances proliferation after stimulation through the Ig receptor; and induces immunoglobulin class switch in human to IgG4 and IgE (IgG1 and IgE in mouse). IL-13, which is closely related to IL-4 and has many similar effects on B cells.

Other cytokines, such as IL-2, -5, -6, -11, and nerve growth factor (NGF), act on mature B cells and can either enhance their proliferation or promote their differentiation into immunoglobulin-secreting cells. In addition, IL-10 enhances the viability of B cells *in vitro*, increases MHC class II expression, and augments the proliferation and differentiation of B cells after stimulation through the Ig receptor or CD40. TGF-β1 is a significant switch factor for IgA. This cytokine induces human B cells triggered by mitogen to switch to both IgA1 and IgA2.

SDF-1 (stromal cell-derived factor) attracts early-stage B-cell precursors and is a likely mechanism whereby B cells form islands in the bone marrow. There are at least two major populations of B cells: B-1 in the follicular mantle and peritoneal cavity, and conventional B-2 cells, found in lymphoid follicles. The B-1 lineage predominates early in gestation and produces natural antibodies of the IgM isotype.[47] Local expression of IgA plasma cell precursors in the ileum important for bacterial containment.[48]

Innate Lymphoid Cells

Natural killer cells. Lymphocytes that do not express a T cell receptor but have cytolytic or non-cytolytic function typical of T cells are called innate lymphoid cells see (Chapter 3). The first described were cytolytic natural killer cells, which comprise about 10% to 15% of circulating lymphocytes (see Table 2.4; Chapter 12). These cells are usually larger than typical lymphocytes (10 to 12 μm) with less nuclear material and more cytoplasm small lymphocytes. They possess electron-dense peroxidase-negative granules and a developed Golgi apparatus.

Functional NK cells are found in the fetal liver as early as 6 weeks gestation. Fetal NK cells express cytoplasmic CD3 proteins, but no TCR rearrangements. Evidence suggests that an Fcγ receptor-positive cell that does not express lineage-specific markers (LIN^-) exists in the fetal mouse thymus where it normally gives rise to T cells. However, if removed from the thymus, the cells develop into $CD3^-$ NK cells. Such $CD3^-$ cells with

variable CD16 expression exist in human thymus and can be induced to proliferate, express NK-associated antigens, and exhibit NK cell function *in vitro*. These cells also express substantial levels of CD3δ and CD3ε in the cytoplasm.[49]

Mature NK cells in the blood do not express conventional antigen receptors, such as TCR or Ig, and the genes for these receptors remain un-rearranged. Some express FcγRIII (CD16), and others express CD56, an adhesion molecule. More than 90% of these cells express CD11b but not CD27. In the tissues, subsets of human NK cells express variable levels of CD11b and CD27, which defines their function (tolerant, cytotoxic or regulatory). NK cells, like T cells, also express the CD2 molecule. NK cells express the β chain of the IL-2 receptor, CD122, which allows resting NK cells to respond directly to IL-2.

The function of some NK cells is to provide nonspecific cytotoxic activity towards virally infected cells and tumor cells (Chapter 25). NK cells also can kill specifically when they are provided an antibody. This death delivery mechanism, known as antibody-dependent cellular cytotoxicity (ADCC), occurs via binding of the antibody to the Fcγ receptor CD16. After activation, NK cells produce cytokines, such as IFN-γ, that affect proliferation and differentiation of other cell types, especially DCs. Some of the recognition molecules on human NK cells are activating, some are inhibiting, and some act as receptors for MHC class I molecules.

The ontogeny of NK cells is now better understood. Although they express a number of membrane antigens in common with T cells and share functional properties with some T cell subsets, suggesting a common origin, NK cells are found in fetuses before the development of T cells or the thymus. In addition, NK cells develop normally in nude, athymic mice. NK cells probably develop extrathymically, and data suggest that they can develop from stem cells in lymph nodes. NK cells arise from triple-negative ($CD3^-CD4^-CD8^-$) precursors that are $CD56^+$, but do not express CD34 or CD5. T cells, on the other hand, develop from "triple-negative" precursors that are $CD34^+CD5^+CD56^+$. It is likely that T and NK cells arise from a common "triple-negative" precursor with the phenotype $CD7^+CD34^+CD5^+CD56^+$.

The cytokine receptor that determines lineage specificity is the α chain of the IL-2 receptor, CD25. Once CD25 is upregulated, the cell is destined to become a T cell. The cytokines most important in the early development of NK cells are IL-15 and IL-7. Flt ligand and c-kit also facilitate NK cell expansion. Several cytokines promote the growth and differentiation of mature NK cells. IL-2 induces proliferation and activation of NK cells. This probably occurs via the IL-2 receptor β chain (CD122) as NK cells do not express CD25. IL-2 also induces the growth of NK cells from precursors in bone marrow cultures. Both IL-7 and IL-12 activate NK cells. Although IL-4 inhibits the effects of IL-2 or IL-7 on NK cells, it acts synergistically with IL-12 to induce proliferation of $CD56^+$ cells. IL-6, despite having no effect by itself, enhances NK cell activity in thymocytes cultured with IL-2. Finally, IL-15 is also involved in signaling NK cells for survival.[50] Subsets of human NK cells develop based on responsiveness to TGF-β and IL-10 (tolerant), IL-12 and IGF-1 (cytotoxic) and TGF-β, IL-7 and IL-15.[50] There is now compelling evidence that NK cells can form a type of memory, especially after intense challenge.[51]

Non-cytotoxic innate immune cells. Non-cytotoxic innate immune cells, which are similar in function to T helper subsets, are divided into three main groups: ILC1, ILC2, and IL 3. These groups are defined by the cytokines they produce.[52,53] IlC-1 cells are non-cytotoxic Lin^- cells that produce INF-γ and TNF-α. ILC2's produce Th2 cytokines, such as Il-4, Il-5, Il-9 and Il-13; and some produce amphiregulin. They can be driven by IL-33.53 ILC3's produce IL-17A, IL-17F, IL-22, GM-CSF, and TNF-α. ILC3 are the most heterogeneous subset. They express CCR6 and CD117 and can be divided based on expression of the NCR Nkp44. The role of these cells in normal host function and responses to chronic inflammatory stimuli and cancer indicates that their roles are multifaceted.

KEY CONCEPTS

Tissues of the Immune System

1. Stem cells proliferate and mature into effector cells in the primary lymphoid organs, which include bone marrow and thymus.
2. Mature immune cells reside, undergo additional maturation, and generate immune responses in the secondary lymphoid organs.
3. Spleen and lymph nodes comprise the systemic immune system, which functions to protect the body from antigens in the lymphatic drainage and in the blood stream.
4. The mucosal immune system (respiratory, gastrointestinal, and genital) and the skin and adipose tissues have unique features that differentiate the immune system at these sites from the systemic immune sites, including the mucosal associated lymphoid tissue (MALT).
5. Commensal organisms at mucosal surfaces are an important component of the immune response at these sites.

MAJOR LYMPHOID ORGANS

The primary lymphoid organs are sites where lymphocytes differentiate from stem cells and proliferate and mature into effector cells. From birth to old age, these functions are carried out only in the bone marrow and the thymus.

Bone Marrow

The bone marrow provides the environment necessary for the development of most of the white blood cells of the body (Fig. 2.2). At birth, most bone cavities are filled with actively dividing blood-forming elements known as "red" marrow. By 3 to 4 years, however, the tibia and femur become filled with fat cells, limiting their role in hematopoietic development. The ribs, sternum, iliac crest and vertebrae remain 30% to 50% cellular and produce hematopoietic cells throughout life.[1]

Main components of the bone marrow include blood vessels, cells, and extracellular matrix. The production of cells from HSC occurs in areas separated by vascular sinuses. The walls of the surrounding sinus contain a layer of endothelial cells with endocytic and adhesive properties. These specialized endothelial cells of the sinuses probably produce type IV collagen and laminin for structural support via CXCL-12 (SDF-1) interactions. These cells also elaborate colony-stimulating factors and IL-6. The outer wall of the sinus is irregularly covered with reticular cells, which branch into areas where cells develop and provide anchors by producing reticular fibers. Megakaryocytes lie against this wall, touching the endothelial cells.

A functional unit of marrow, called a spheroid, contains adipocytes, stromal cell types and macrophages. These reticular cell networks compartmentalize the developing progenitor cells into separate microenvironments called *hematons*. Osteoblasts and osteoclasts regulate production of progenitor cell expansion.[4]

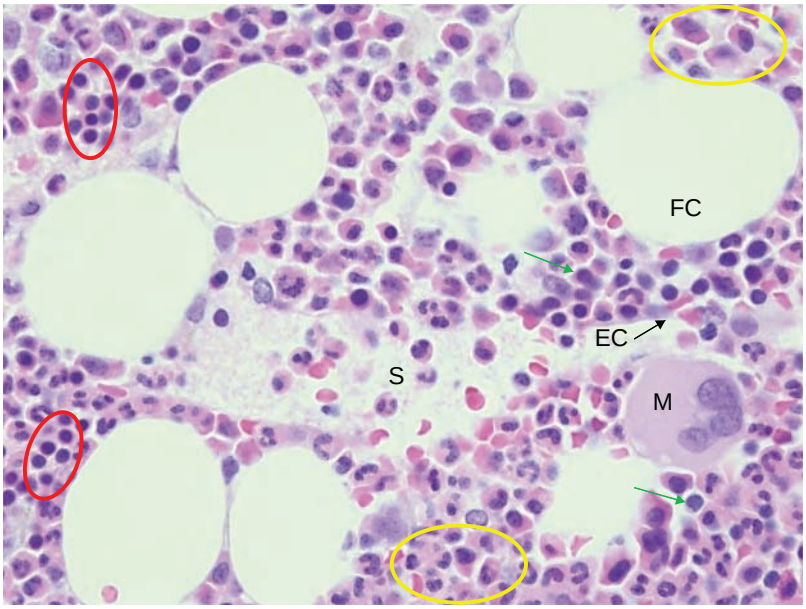


FIG. 2.2 Bone marrow with erythroid precursors *(circled red)*, myeloid precursors *(circled yellow)*, megakaryocytes *(M)*, sinusoids *(S)*, and fat cell *(FC)*.

The kind of distribution of stem and progenitor cells across the radial axis of the bone suggests that the HSC are next to the bone surface, whereas the more mature progenitor cells are nearer the central venous sinus, which facilitates release of mature cells. The production of new progenitor cells from stem cells occurs as a result of interactions between stem cells and stromal cells. Given the right stimulus, most of the progeny proliferate and differentiate further, which may result in migration from the bone marrow. In migrating, the cells become detached from stromal elements and progress toward the central sinus.

Control of hematopoiesis is regulated by both positive and negative cytokines and by up- and downregulation of various adhesion molecules (Chapter 16) in committed progenitor cells. The molecules involved include the fibronectin receptor, glycoproteins IIb and IIIa, ICAM-1 (CD54), LFA-1 (CD11, CD18), LFA-3 (CD58), CD2, and CD44. Adhesion molecules on stromal cell surfaces include fibronectin, laminin, ICAM-1 (CD54), types I, III, and IV collagen, and N-CAM. The most clearly established role for adhesion molecules involves fibronectin, which allows erythroid precursors to bind to stromal cells and thus facilitates progression from erythroblast to reticulocyte. Molecular signals important for the HSC niche include N-cadherin that regulates osteoblastic interactions with HSC, Wnt/B catenin signals important for self-renewal of HSC, VEGF important for coupling osteoblasts with vascular endothelial cells and PDE2, an inflammatory mediator that can increase HSC numbers

Accessory cell populations in bone marrow regulate many aspects of hematopoiesis, both positively and negatively. The upregulation of growth of the earliest progenitor cells is mediated by cytokines. For example, macrophages produce IL-1, which then induces stromal cells to express growth factors such as GM-CSF, IL-6, and IL-11. However, downregulation can occur at any stage. For example, T cells regulate hematopoiesis by producing factors that act on early erythroid progenitor cells, BFU-E. CFU-E, which are later progenitors, are fully differentiated by erythropoietin. By contrast, activated T cells produce factors that suppress BFU-E and CFU-E *in vitro*.

Cells in the bone marrow were originally characterized by morphology. The predominant types are those of the myeloid lineage, which account for about 50% to 70% of the cells. Red blood cell precursors represent from 15% to 40% of the total cells. Other lineages exist in lower proportions (<5%). With the advent of cell surface antigen markers and flow cytometry, a more precise delineation could be made (Fig. 2.3). Thus, we now know that of the mature leukocytes in the bone marrow, approximately 70% are $CD3^+$, $CD14^+$, $CD20^+$, or $CD11b^+$.

Both memory T and B cells return to the bone marrow after generation. These are designated as Lin^+. Of the Lin^- cells, about 6% are $CD33^+$ and primarily of myeloid lineage. A Lin^-CD71^+ population comprises about 18% of the total and is mostly of the red blood cell lineage.

Thymus

The thymus is located below the sternum and in the mediastinum. This bi-lobed organ develops from the third and fourth pharyngeal pouches and is endodermal in origin. It is organized into a loose lobular structure. Areas in each lobe consist of a cortex of rapidly dividing cells and a medulla that contains fewer, but more mature, T cells (Figs. 2.4 and 2.5). This arrangement has long suggested a scenario for differentiation where cells progress from the cortex to the medulla.

Non-lymphocyte cells play very important site-specific roles in development of T cells. Epithelial cells are scattered throughout the thymus. Depending on their location, they are known as nurse cells, cortical epithelial cells, or medullary epithelial cells. Macrophage-type cells and interdigitating cells that are bone-marrow derived are located at the junction between cortex and medulla and are involved in T-cell selection.

Enlarged, activated T-cell precursors from the bone marrow begin by colonizing the subcapsular region of each lobe. These are actively proliferating and can self-renew. Selection begins when their progeny encounter MHC class II molecule-bearing cortical epithelial cells. A further education process probably occurs by interaction with macrophage-like cells found at the cortico-medullary junction and in the medulla.

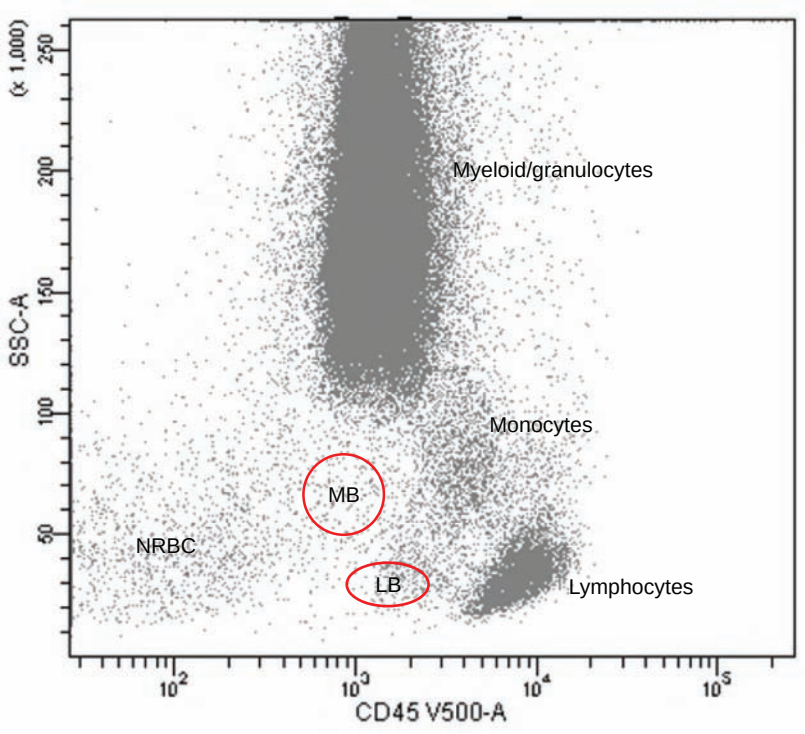


FIG. 2.3 Flow cytometry cytogram of normal human bone marrow based on CD45 expression and light side scatter. This technique separates out the majority of marrow cells, including nucleated red blood cells *(NRBC)*, myeloblasts *(MB)*, and lymphoblasts *(LB)*.

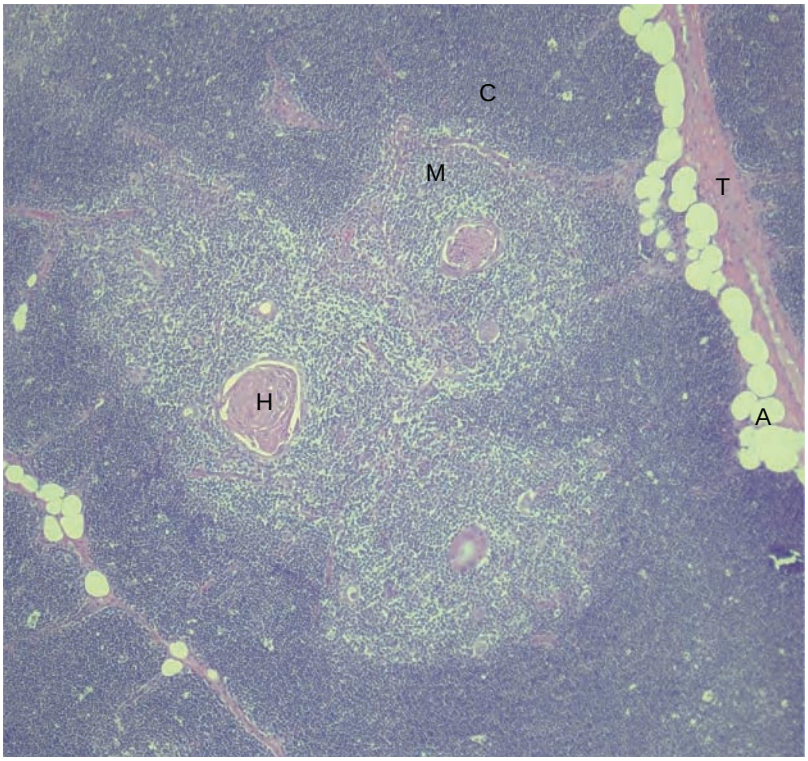


FIG. 2.4 Thymus showing medulla *(M)*, cortex *(C)*, Hassel corpuscle *(H)*, trabecular *(T)*, and mature adipose *(A)*.

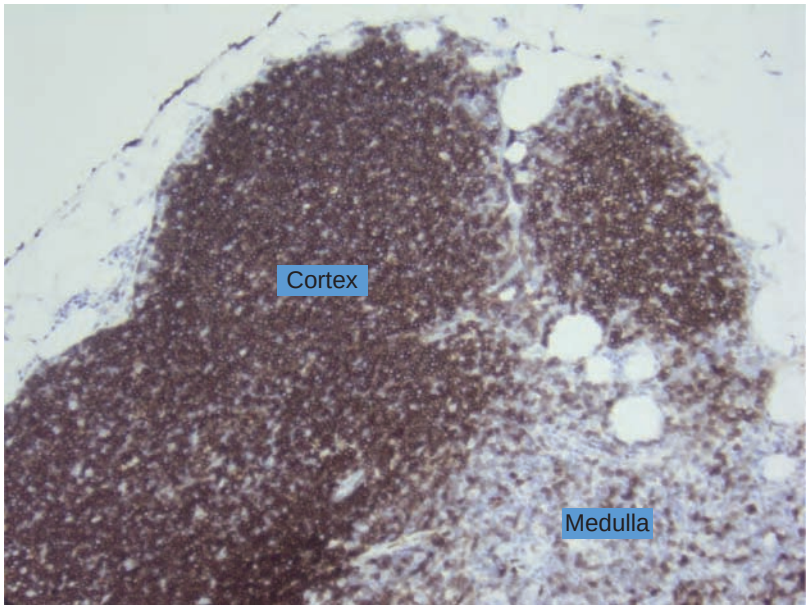


FIG. 2.5 Thymus CD1a immunostain. Cortical thymocytes are extensively positive; medullary thymocytes are focally positive.

Thymus nurse cells, found in the cortex, were originally thought to contribute to the thymic education of T cells. Because large numbers of thymic cells (50 to 200) can be found inside each nurse cell, it was believed that these structures provided an environment where selection and expansion could occur. There is now evidence that secondary rearrangement of Vα can also occur in these structures.[54]

A structure known as Hassall corpuscle, which consists of concentric whorls of epithelial cells, is found in the medulla. Its function remains unclear. The Hassall medullary epithelial cells contain secretory granules, and this network of cells may be active in the production of thymic hormones. For example, thymic stromal lymphopoietin (TSLP) plays a role in production of dendritic cells that select T regs in the thymus.[55] In the fetus, these bundles of cells are widely scattered but become larger as the thymus matures. The center cells eventually become keratinized and die.

The thymic differentiation process (Chapter 9) involves rearrangement of functional TCR, surface expression of CD3, and both positive and negative selection that allows only a small percentage of T cells to survive. Pre-T cells in the thymus express CD2, CD5, and CD7, as well as activation antigens such as CD38 and the transferrin receptor (CD71). Pre-T cells express intracytoplasmic CD3 and exhibit rearrangements in the TCR-β chain. Successful rearrangement of TCR-α allows the cell to progress to the next stage of development with functional TCR and CD3 on the cell surface.[36]

Most cells in the thymus (85%) express both CD4 and CD8 on their surface, termed the "double-positive" stage, as well as CD1 and CD69, an activation marker. CD69 is expressed until the cell reaches the single-positive stage, where it expresses either CD4 or CD8, but not both. T cells are $CD45RO^+$ at the double-positive stage into the single-positive stage. Prior to leaving the thymus, CD45RO is downregulated and CD45RA appears. Mature thymocytes lose CD1 expression and either CD4 or CD8 expression. Most of these mature cells are also negative for activation molecules (CD38 and CD71). However, they acquire an adhesion molecule, CD44, which is necessary for homing. Upon completion of this process of thymus selection and education, mature CD4 or CD8 T cells leave the thymus and enter the peripheral circulation via the post-capillary venules at the cortico-medullary junction.

After birth and during childhood, the thymus continues to grow and to select and educate T cells. This process promotes the development of a robust and diversified repertoire. Prior to puberty, however, the thymus begins to involute. The rapidly dividing cortex is the first to atrophy, leaving medullary areas intact. The sensitivity of cortical thymocytes to hormone-induced death probably accounts for the involution, although human thymocytes are less sensitive to glucocorticosteroids than are murine thymocytes. However, an increase in steroids reduces immature thymocyte numbers and enhances thymus involution. Recent evidence suggests that active TCR rearrangements, and hence T cell development, continue in the adult thymus, albeit at a lower level than during childhood. There is an age-associated decline in new T cell production, such that by age 75 the ability to make new T cells in humans is severely reduced.

Peripheral Development of Hematopoietic and Lymphoid Cells

Although most of the key steps during the growth and development of hematopoietic and lymphoid cells occur in the bone marrow and thymus, additional maturation steps occur after the cells leave those tissues. For example, monocytes and dendritic cell precursors migrate from blood vessels into tissues where they mature into macrophages and dendritic cells, respectively. There is recent evidence for a tissue-associated macrophage that is fetal in origin. Mast cells and eosinophils also undergo further differentiation in resident tissues. After leaving the bone marrow and thymus, B and T cells undergo further maturation and memory cell development in secondary lymphoid organs. There is strong evidence that some T cells, particularly γδ T cells residing in mucosal epithelium, do not develop in the thymus.

SECONDARY LYMPHOID ORGANS

Secondary lymphoid organs are sites where mature lymphocytes reside and where immune responses are generated. Secondary lymphoid organs belong to either the systemic or mucosal immune systems. The systemic immune system includes the spleen and lymph nodes and functions to protect the body from antigens in the lymphatic drainage and circulating in the bloodstream. The mucosal immune system responds to antigens that enter through mucosal epithelium and plays an important role in the inductive phase of the immune response. Unique features differentiate the mucosal immune system from the systemic immune system (Chapter 24). These include efferent, but not afferent, lymphatics, a specialized FAE involved in antigen sampling at the mucosal surface (Fig. 2.6), specialized dendritic cells that rapidly process and present antigens to initiate antigen-specific immune responses, unique distribution and subsets, and an environment that promotes class switching to IgA.

Systemic Immune System

Spleen

The human spleen is surrounded by a capsule of fibrous tissue with many trabeculae traversing from the capsule into the tissue of the spleen. These trabeculae branch and anastomose, forming a complex framework of lobules. Splenic blood vessels enter and exit through the hilum of the spleen and branch into smaller vessels within the trabeculae. Splenic tissue is supported by a fine network of reticular cells and fibers, called the reticulum, which connects and supports the trabeculae, blood vessels, and capsule.

The lobules of the spleen can be functionally divided into two compartments, the red pulp and the white pulp. The largest compartment is the red pulp, which contains numerous venous sinuses situated between arteries and veins. Blood is filtered through these sinuses, which contain many macrophages that phagocytose senescent red and white blood cells, bacteria, and other particulate material. Other leukocytes are found in the red pulp, including neutrophils, eosinophils, and lymphocytes, particularly plasma cells.[56]

The white pulp consists of lymphoid tissue surrounding central arterioles, which are branches of trabecular arteries. The human spleen is structurally different from rodent spleens in that there is no central organization of follicles and the central artery.[56] Rather, a T cell-predominant area is found immediately surrounding a central arteriole, the so-called periarteriolar lymphoid sheath (PALS). The PALS contains both CD4 and CD8 T cells. It is punctuated at intervals by B-cell-predominant areas, termed follicles or so-called malpighian corpuscles. These B-cell-predominant areas include both primary and secondary follicles. Primary follicles consist of only a mantle zone, without germinal centers, whereas secondary follicles contain an inner germinal center in addition to the outer mantle zone (Fig. 2.7). Within the mantle zone are predominantly resting B cells, which express surface IgM/IgD and CD23 (FcεRII). It is within germinal centers that immunoglobulin class switch, affinity maturation through somatic mutation, and the development of memory B cells occurs. Germinal centers are more prevalent at younger ages and diminish with aging. CD4 T cells play a key role in B-cell responses through CD40L and other interactions. The signaling that occurs through this interaction is central to B-cell activation and class switching. In addition to activated B cells and CD4 T cells, the germinal center contains FDCs and macrophages.

At the interface between white pulp and red pulp is an anatomical location known as the marginal zone, which receives blood from branches of central arterioles opening into this region. The marginal zone contains T cells, as well as subsets of macrophages and B cells. Marginal zone B cells (MZB) are distinct from follicular B cells. They express surface IgM, but only low levels of IgD and no CD23. The initial encounter of T cells and B cells with antigen occurs in the marginal zone after blood enters through branches of the central arteriole. Antigen presentation is enhanced by MZB cells, which are important in T-cell-independent responses.

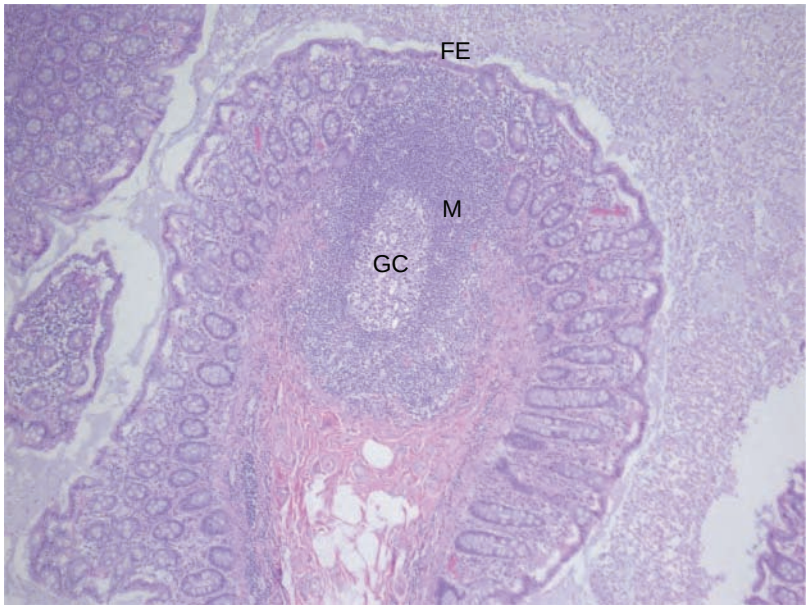


FIG. 2.6 Colon. Shown is a mucosal lymphoid follicle containing a germinal center *(GC)* and mantle layer (M). Follicle associated epithelium is labeled *(FE)*.

Lymph Nodes and Lymphatics

Lymph nodes occur as chains or groups located along lymphatic vessels. Lymph nodes exist in two major groups: those that drain the skin and superficial tissues (*e.g.*, cervical, axillary, or inguinal lymph nodes), and those that drain the mucosal and deep tissues of the body (*e.g.*, mesenteric, mediastinal, and periaortic lymph nodes). Lymph nodes are oval structures surrounded by adipose tissue with an indentation at the region of a hilus, where blood vessels enter and leave the node (Fig. 2.8). A lymph node is surrounded by a fibrous capsule contiguous with trabeculae traversing the node (Fig. 2.9). Blood vessels and nerves, which enter through the hilum, branch through these trabeculae to the various parts of the node. Immediately beneath the capsule is a subcapsular (marginal) sinus. Afferent lymph vessels enter into this sinus opposite the hilum. Dendritic cells process antigen encountered in the skin and migrate into lymph nodes from afferent lymphatics through the subcapsular sinus and into the lymph node. Lymph nodes vary in size, from barely visible in an unstimulated state to several centimeters in size when undergoing an active immune response.

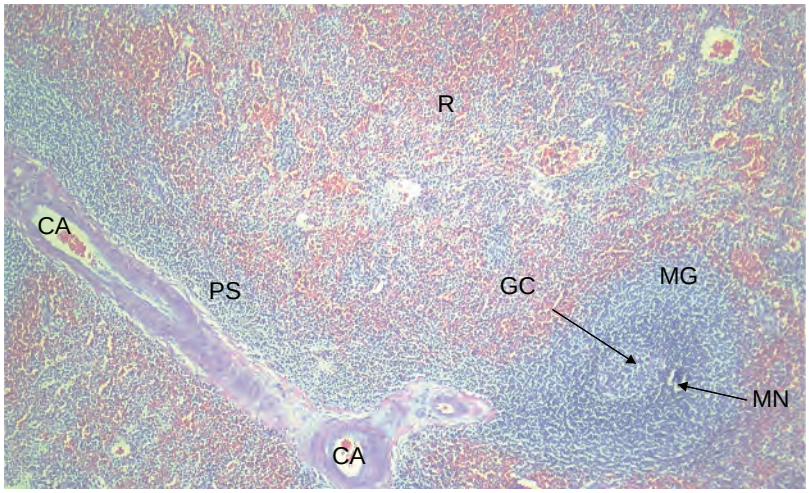


FIG. 2.7 Spleen showing trabecular containing central arteriole *(CA)* with associated periarteriolar lymphoid sheath *(PS)* and follicles containing germinal centers *(GC)*, mantle layer *(MN)*, and marginal zone layer *(MG)*. *R*, Red pulp.

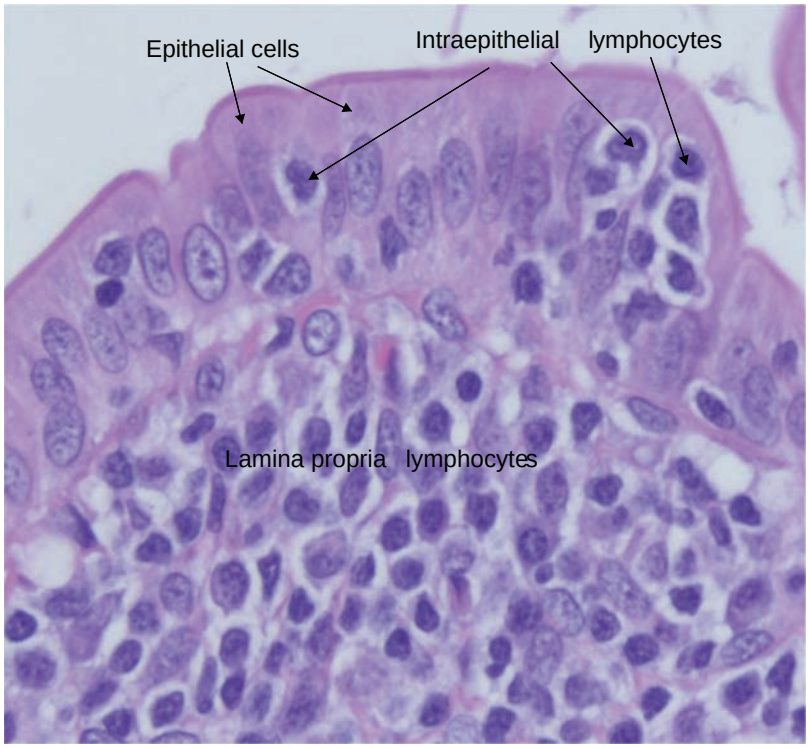


FIG. 2.8 Ileum with intraepithelial lymphocytes and lymphocytes within the lamina propria.

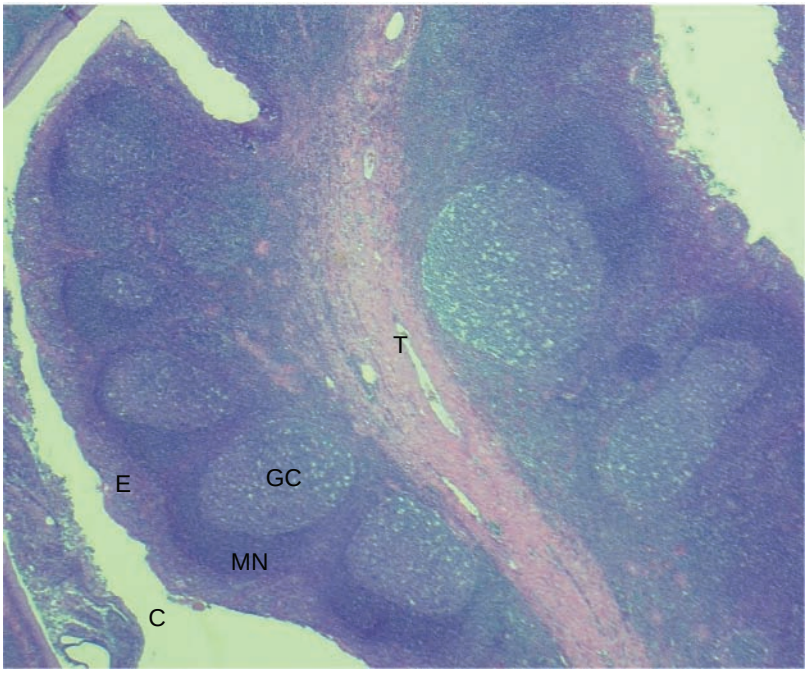


FIG. 2.9 Tonsil structure showing epithelium *(E)*, follicles containing germinal centers *(GC)* and mantle layer *(ML)*, trabecular *(T)*, and crypts *(C)*.

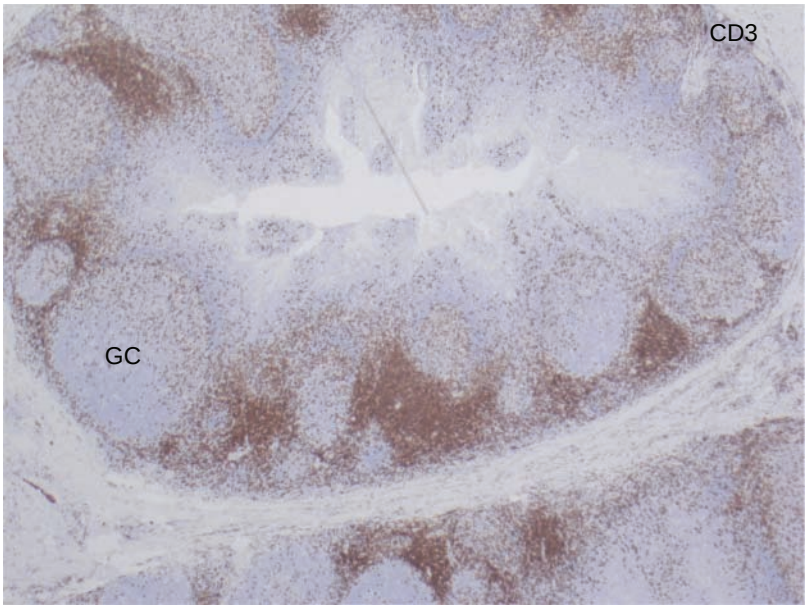


FIG. 2.10 Tonsil CD3 immunostain highlighting T cells predominantly within the interfollicular regions and scattered within germinal centers *(GC)*.

A lymph node is divided into two major regions, the cortex and the medulla. The cortex contains numerous primary and secondary lymphoid follicles, each approximately 0.5 mm in diameter, similar to those in the spleen. Surrounding the lymphoid follicles in the cortex is the paracortical region, which contains mostly T cells along with some macrophages and dendritic cells. Both CD4 and CD8 T cells are present, as are macrophages and B cells (Figs. 2.10–2.13). Accessory cells, including interdigitating dendritic cells, present peptide antigens in association with MHC molecules to the TCR on T cells to activate the T cells (see Chapter 10). Additional accessory molecules (*e.g.*, B7 [CD80] or LFA-3 [CD58]) on the accessory cell, and their ligands (CD28 or CD2, respectively) on the T cell, provide important co-stimulatory signals required for activation of the T cell. Other surface antigens, particularly adhesion molecules such as LFA-1 (CD18) and ICAM-1 (CD54), stabilize cellular interactions, as well as providing additional signals between cells.

In the center of the lymph node, beneath the cortex, lies the medulla, which is divided into medullary cords. Surrounding the medullary cords are medullary sinuses that drain into the hilum. B and T cells migrate from the follicles and paracortical region to the medulla. Medullary cords contain T cells, B cells,

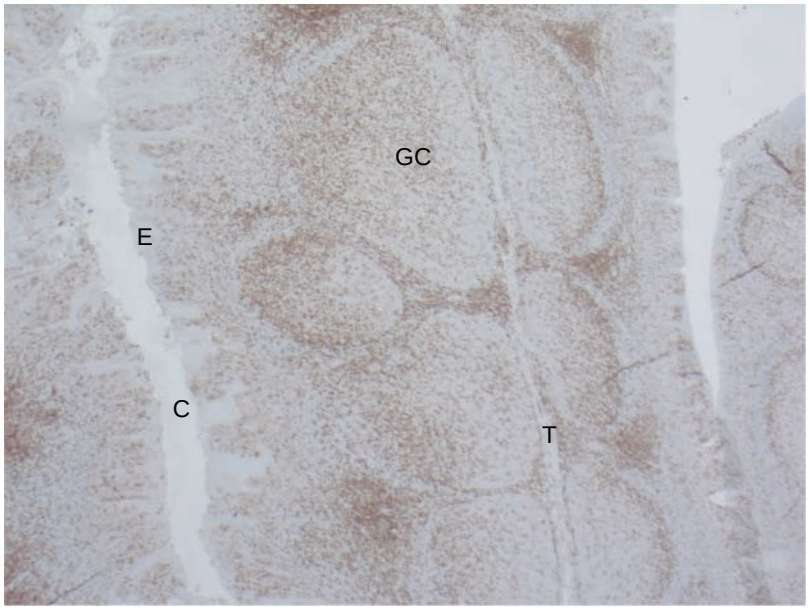


FIG. 2.11 Tonsil CD4 immunostain highlighting follicular helper T cells within germinal centers *(GC)*, and within the interfollicular regions. *C*, Crypt; *E*, epithelium; *T*, trabecular.

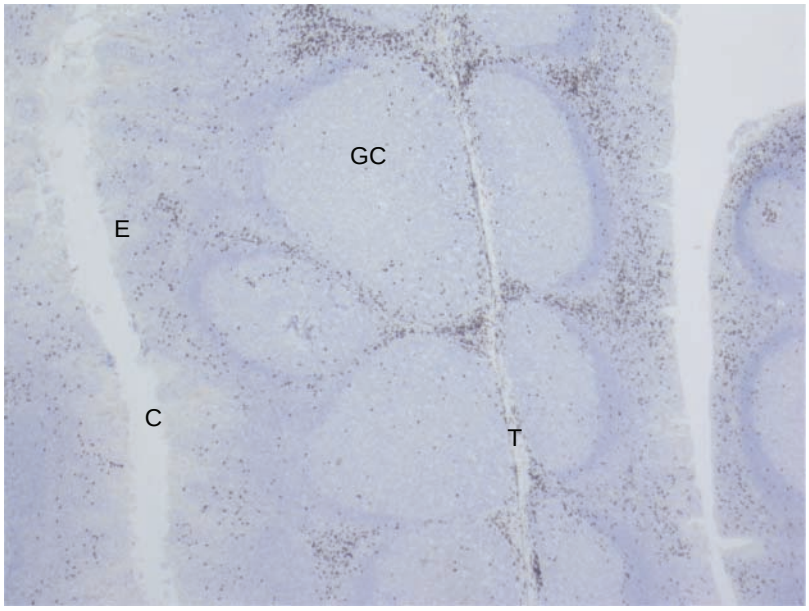


FIG. 2.12 Tonsil CD8 immunostain highlighting T cells predominantly within the interfollicular regions. *C*, Crypt; *E*, epithelium; *GC*, germinal center; *T*, trabecular.

and macrophages, as well as a large number of plasma cells that produce immunoglobulin, which drains into medullary sinuses that empty into the hilum. Efferent lymphatic vessels leave the hilum carrying lipids and antibodies together with mature B and T cells that migrate to other tissues and act as memory B and T cells. The lymphatic vessel system serves to carry lymphocytes derived from various tissue spaces through the network of lymph nodes and eventually to the thoracic duct.

Lymphatic capillaries are lined with lymphatic epithelial cells that serve as valves to move lymph fluid, cells, and nutrients around the body. These epithelial cells express high levels of TLR4, which means they can be activated after LPS delivery to increase lymphaniogenesis.[57] Lymph from the nodes is drawn into the left subclavian vein and back into the circulation. Cancer cells found in lymph nodes may take advantage of this system to seed the body. This system of transport develops early in gestation with both lymphatic muscle cells for propulsion and valves that regulate unidirectional lymph flow. Lymph also serves as a carrier of lipids and the endothelial cells that line the lymphatics are responsive to metabolic signals.[57] Lymph also is recognized as a major carrier of exosomes that may facilitate communication between lymphoid organs.[58]

Gastrointestinal Tract

The organized mucosa associated lymphoid tissue (MALT) of the gastrointestinal system is termed the gut-associated

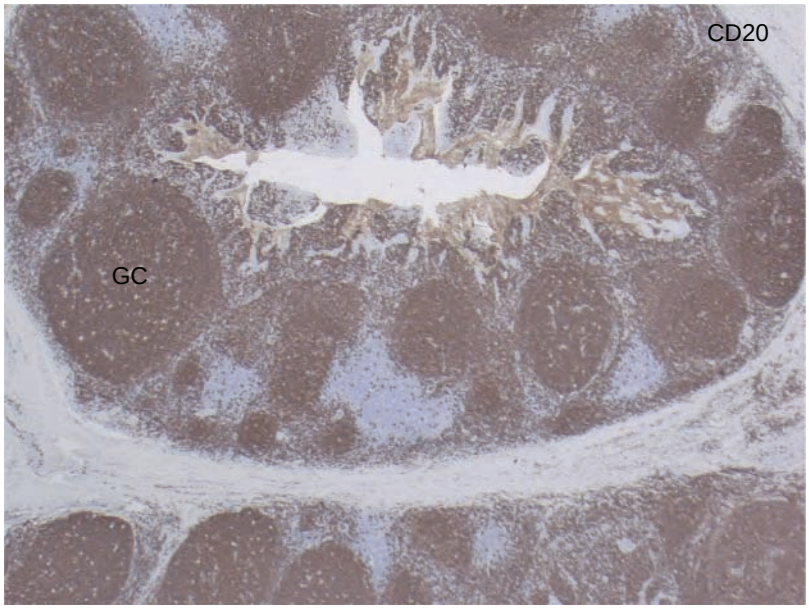


FIG. 2.13 Tonsil CD20 immunostain highlighting B cells predominantly within germinal centers *(GC)*, and scattered within the interfollicular regions.

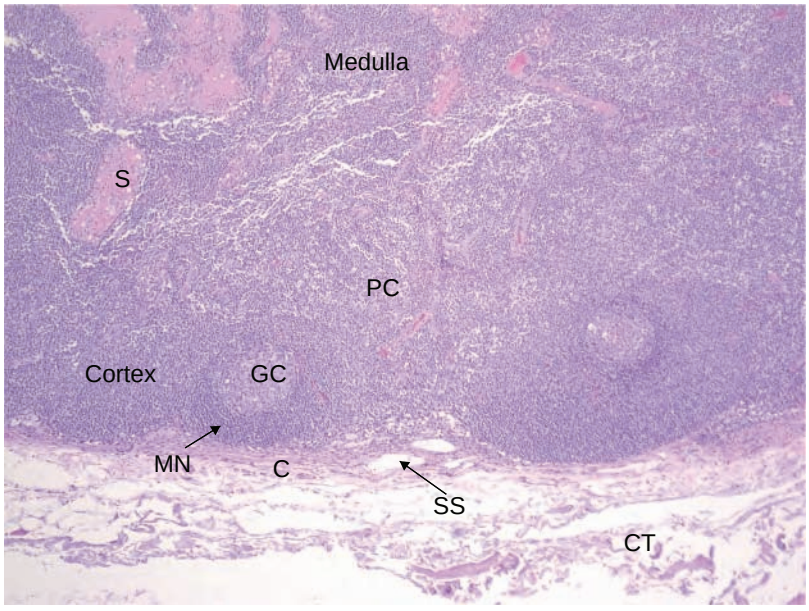


FIG. 2.14 Lymph node showing capsule *(C)* with outer connective tissue *(CT)*, cortex containing follicles with germinal centers *(GC)*, and mantle layer *(MN)*, paracortex *(PC)*, sinusoids *(S)*, and subcapsular sinuses *(SS)*.

lymphoreticular tissue (GALT). It is composed of Peyer patches, cecal and rectal patches, and isolated lymphoid follicles. Isolated lymphoid follicles and cecal and rectal patches are found throughout the lamina propria and are similar to an individual follicle of a Peyer patch. Peyer patches consist of variably sized aggregates of closely associated lymphoid follicles located in the intestinal lamina propria, occurring predominantly in the ileum (Fig. 2.14). These structures arise during fetal life; their full development, with follicles containing germinal centers, does not occur until several weeks after birth, presumably in response to antigenic stimulation. Their number and size increase until puberty and decline thereafter, similar to the thymus.

Virtually all Peyer patch follicles have germinal centers that contain activated B cells, FDCs, CD4 T cells, and tingible-body macrophages (so called because of their appearance after they have phagocytosed cellular debris). Many of the B cells within Peyer patch germinal centers express surface IgA, and it is believed that this is where IgA class switch occurs. An inter-follicular region contains predominantly CD4 and CD8 T cells, as well as dendritic cells, macrophages, and some B cells.

The diffuse tissue of the gastrointestinal tract consists of two components: the lamina propria and intraepithelial lymphocytes (IEL) (Fig. 2.15). The lamina propria is located immediately beneath the epithelium. It contains large numbers of B

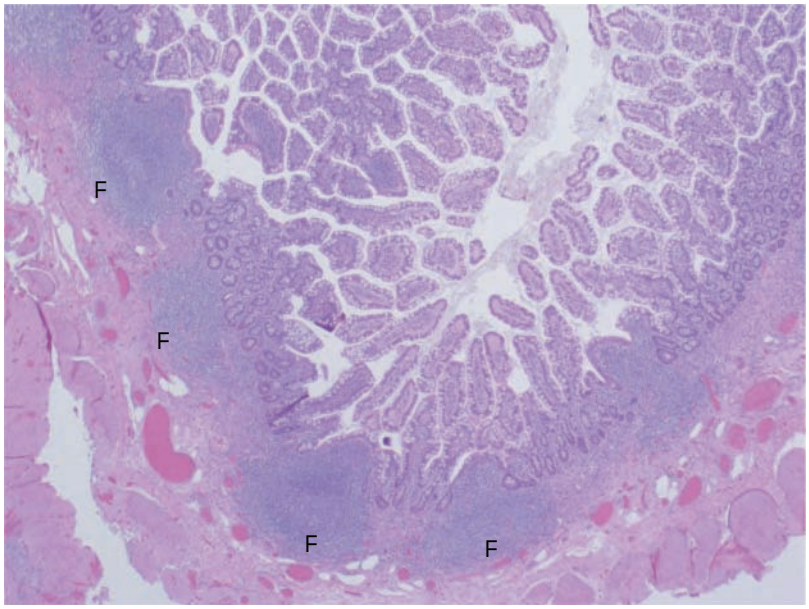


FIG. 2.15 Ileum showing Peyer patches comprised of lymphoid follicles *(F)*.

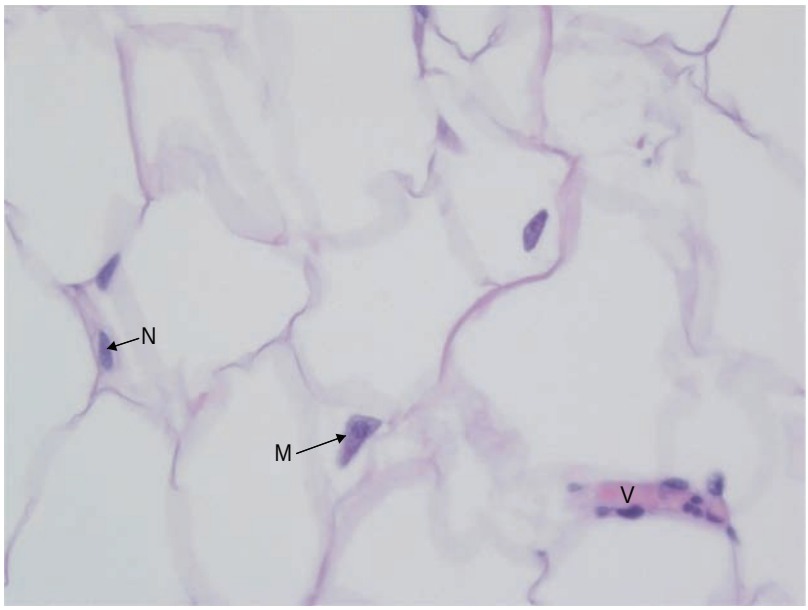


FIG. 2.16 Adipose. The nucleus of a mature fat cell *(N)*, a blood vessel *(V)*, and a macrophage *(M)* are shown.

lymphocytes and plasma cells. A key effector function of the lamina propria is the secretion of antibodies, primarily IgA. IgM represents only 10% to 18% and IgG 3% to 5% of all Ig produced. Two IgA subclasses occur, IgA1 and IgA2. IgA1 represents greater than 90% of IgA in the respiratory tract and greater than 60% in the lamina propria of the small intestine.[51] IgA2 increases in the lower ileum and becomes predominant in the colon and rectum.

Adipose Tissue

In light of the obesity epidemic, adipose tissue has received renewed scrutiny, which has led to the realization that immune cells play central roles in adipose homeostasis and in obesity-associated chronic inflammation (Fig. 2.16). Macrophages are a central component, switching from M2 type to M1 type during obesity. In lean adipose tissue there are numerous T regs, ILC2s and few CD8 T cells, which reverses during obesity when inflammation increases.[59,60]

Epithelial Innate Immunity and Commensal Organisms

It has become clear that epithelial cells have autologous immune capacity and that this property is conserved among epithelial tissues in the body. This immune capacity functions to reinforce the epithelial barrier, stimulate crosstalk within the epithelium, recruit underlying immune cells within the stromal connective tissue, and initiate repair of epithelial damage. In particular, the communication between the epithelium and underlying immune cells, including T cells, B cells, dendritic cells, and

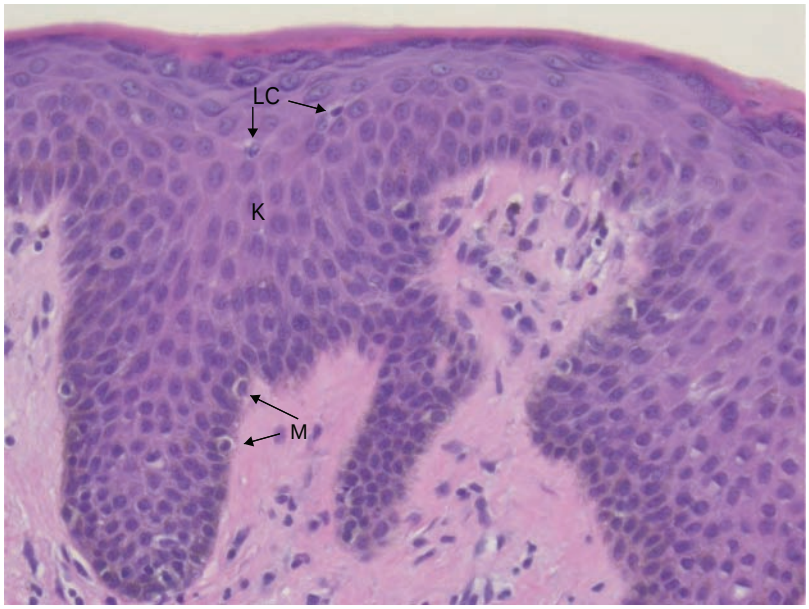


FIG. 2.17 Skin. Keratinocytes *(K)*, Langerhans cells *(LC)*, and melanocytes *(M)* are shown.

macrophages, facilitates the coordination of a tissue and inflammatory response that ultimately can protect the epithelium from external damage (Fig. 2.17).

In some sense, the epithelium is a bidirectional immune mediator of antigens on one side and host on the other. The epithelium has multiple sensors that alert the cells to environmental insults that range from the biologic (bacteria and viruses) to the non-living challenges (pollutants and allergens). These sensors include pattern associated molecular pattern (PAMP) receptors such as toll-like receptors (TLRs) and nucleotide binding and oligomerization domain (NOD) receptors (see Chapter 3).[61,62] Binding of these receptors triggers the epithelial cell production of cytokines and chemokines, hormones, anti-microbial peptides, MHCII expression and antigen presentation, mucous, carbohydrate moieties, and the transport of IgA.[61] Many of these products play an important role in signaling both neighboring cells as well as the underlying stromal immune system to offer a rapid and immediate response to the triggering entity. Additionally, there is some evidence that the induction of the innate immune response of the epithelium can epigenetically rewire the epithelium resulting in the enhancement of epithelial repair upon seeing the insult additional times

All epithelial surfaces live in symbiosis with over 1000 different species of viruses, bacteria, protozoa, and fungi that together outnumber human cells by a factor of 10. Collectively called the commensal microbiota (Chapter 22), these organisms are essential to the development, maturation, organization, and regulation of both the epithelium and the underlying immune system at mucosal surfaces.[63] The commensal microbiota generates a large number of metabolites that have significant effects on both the cells and their functional state and can affect regulation of differentiation, production of cytokines and anti-microbial products, and epithelial barrier function.[64]

Both the epithelium and mucosal immune cells express receptors for these metabolites. Examples of these metabolites include single chain fatty acids (SCFAs), pyruvate, and lactate.[64] These essential interactions activate and prime both innate and specific immune responses such as production of IgA, induction of regulatory T cells, and stimulation of anti-inflammatory cytokines.[64] Microbial metabolites induce metabolite reactive mucosal-associated T cells (MAIT) that are located through all mucosal surfaces.[65] Once activated, the T cell can respond to local cytokines produced by the epithelium as well as other environmental cues. Thus, the types and quantities of microorganisms present at a mucosal surface are an important component of the mucosal immune response.[66–68]

New Modalities to Study Immune System Development and Function

Two new modalities have recently altered the landscape of knowledge on the immune system: single cell sequencing and induced pluripotent stem cell (iPSC)-derived immune cells. Advances in next generation sequencing have overcome a critical barrier in characterization of the immune system and have allowed the immune system to be investigated at the single cell level. Many discoveries using single-cell transcriptomics combined with fluorescent activated cell sorting (FACS) are now resolving immune cell heterogeneity, and what was once believed to be well-defined immune populations are now seen as a compilation of distinct cellular subsets that segregate by function yet express overlapping phenotypic markers. This new approach is indicating that immune cell function is very dependent on environment and tissue context. Using iPSC derived technology, various hematopoietic and immune cells can be successfully generated from a pluripotent precursor. iPSC cultures can provide a constant and continuous production of immune cells from the same genetic individual. These cells are easily amenable to genetic manipulation and, upon stimulation, can produce many immune effectors such as cytokines and chemokines. Applications towards customized immune therapies using iPSC derived hematopoietic cells are in their infancy. Both single-cell genomics and iPSC-derived immune cells help define key processes in immune cell differentiation and maturation, understand hematopoiesis, and ultimately contribute to predicting immune cell function. These approaches hold promise to advance the fundamental knowledge of the human immune system.

ON THE HORIZON

1. Understanding how stem cells self-renew will be a key to exploiting them for gene therapy.
2. Exploiting innate and acquired immune cell function will require an understanding of subpopulations of cells and the manner in which they are induced.
3. Generation of new T and B cells later in life might prolong quality of life in the aged.
4. The role of adipose tissue in HSC development in the bone marrow and control of inflammation in obesity will be fundamental to controlling the obesity epidemic.
5. Exploiting interactions between the mucosal immune system and commensal populations is likely to improve health, prevent inflammation, and allow less antibiotic use.
6. The role of lymphatics in transportation of lymphocytes and exosomes will be crucial for spreading immunologic information; inflammation and cancer spread might be reduced by controlling inflammation of the lymphatics.

ACKNOWLEDGMENTS

We thank Dr. Gregory R. Harriman at BioAdvance for his previous efforts on this chapter; Eleanor Chapman, Anna Wirt, Terry Saulsberry, Yvette Wyckoff and Pamela Love for help with the manuscript; and Dr. Jerry McGhee for critical review of the first edition.

REFERENCES

1. Mahony CB, Bertrand JY. How HSCs colonize and expand in the fetal niche of the vertebrate embryo: an evolutionary perspective. *Front Cell Dev Biol*. 2019;7:34.
2. Kumar A, D'Souza SS, Thakur AS. Understanding the journey of human hematopoietic stem cell development. *Stem Cells Int.*. 2019;2019:2141475.
3. Ciriza J, Thompson H, Petrosian R, et al. The migration of hematopoietic progenitors from the fetal liver to the fetal bone marrow: lessons learned and possible clinical applications. *Exp Hematol*. 2013;41:411–423.
4. Brown G, Ceredig R. Modeling the hematopoietic landscape. *Front Cell Dev Biol*. 2019;7:104.
5. Smith JN, Calvi LM. Concise review: current concepts in bone marrow microenvironmental regulation of hematopoietic stem and progenitor cells. *Stem Cells*. 2013;31:1044–1450.
6. Copley MR, Eaves CJ. Developmental changes in hematopoietic stem cell properties. *Exp Mol Med*. 2013;45:e55.
7. Wilkinson AC, Ishida R, Kikuchi M, et al. Long-term ex vivo haematopoietic-stem-cell expansion allows nonconditioned transplantation. *Nature.*. 2019;571:117–121.
8. Pang WW, Price EA, Sahoo D, et al. Human bone marrow hematopoietic stem cells are increased in frequency and myeloid-biased with age. *Proc Natl Acad Sci U S A*. 2011;108:20012–20017.
9. Metcalf D. Hematopoietic cytokines. *Blood*. 2008;111:485–491.
10. Yu J, Freud AG, Caligiuri MA. Location and cellular stages of natural killer cell development. *Trends Immunol*. 2013;34:573–582.
11. Wilson A, Laurenti E, Oser G, et al. Hematopoietic stem cells reversibly switch from dormancy to self-renewal during homeostasis and repair. *Cell*. 2008;135:1118–1129.
12. Zhang CC, Lodish HF. Cytokines regulating hematopoietic stem cell function. *Curr Opin Hematol*. 2008;15:307–311.
13. Kondo M, Wagers AJ, Manz MG, et al. Biology of hematopoietic stem cells and progenitors: implications for clinical application. *Annu Rev Immunol*. 2003;21:759–806.
14. Seita J, Ema H, Ooehara J, et al. Lnk negatively regulates self-renewal of hematopoietic stem cells by modifying thrombopoietin-mediated signal transduction. *Proc Natl Acad Sci U S A*. 2007;104:2349–2354.
15. Varol C, Mildner A, Jung S. Macrophages: development and tissue specialization. *Annu Rev Immunol*. 2015;33:643–675.
16. Bouhlel MA, Derudas B, Rigamonti E, et al. PPARgamma activation primes human monocytes into alternative M2 macrophages with anti-inflammatory properties. *Cell Metab*. 2007;6:137.
17. Hoeffel G, Ginhoux F. Ontogeny of tissue-resident macrophages. *Front Immunol.*. 2015;6:486.
18. Watanabe S, Alexander M, Misharin AV, et al. The role of macrophages in the resolution of inflammation. *J Clin Invest*. 2019;129:2619–2628.
19. Karsunky H, Merad M, Cozzio A, et al. Flt3 ligand regulates dendritic cell development from Flt3+ lymphoid and myeloid-committed progenitors to Flt3+ dendritic cells in vivo. *J Exp Med*. 2003;198:305–313.
20. Chicha L, Jarrossay D, Manz MG. Clonal type I interferon-producing and dendritic cell precursors are contained in both human lymphoid and myeloid progenitor populations. *J Exp Med*. 2004;200:1519–1524.
21. Patente TA, Pinho MP, Oliveira AA, et al. Human dendritic cells: their heterogeneity and clinical application potential in cancer immunotherapy. *Front Immunol*. 2019;9:3176.
22. Mende I, Karsunky H, Weissman IL, et al. Flk2+ myeloid progenitors are the main source of Langerhans cells. *Blood*. 2006;107:1383–1390.
23. Randolph GJ, Angeli V, Swartz MA. Dendritic-cell trafficking to lymph nodes through lymphatic vessels. *Nat Rev Immunol*. 2005;5:617–628.
24. Friedman AD. Transcriptional regulation of granulocyte and monocyte development. *Oncogene*. 2002;21:3377–3390.
25. Delgado-Rizo V, Martínez-Guzmán MA, Iñiguez-Gutierrez L, et al. Neutrophil extracellular traps and its implications in inflammation: an overview. *Front Immunol*. 2017;8:81.
26. Lin A, Loré K. Granulocytes: new members of the antigen-presenting cell family. *Front Immunol*. 2017;8:1781.
27. Lieber JG, Webb S, Suratt BT, et al. The in vitro production and characterization of neutrophils from embryonic stem cells. *Blood*. 2004;103:852–859.
28. Ramirez GA, Yacoub MR, Ripa M, et al. Eosinophils from physiology to disease: a comprehensive review. *Biomed Res Int*. 2018;2018:9095275.
29. Kambe N, Hiramatsu H, Shimonaka M, et al. Development of both human connective tissue-type and mucosal-type mast cells in mice from hematopoietic stem cells with identical distribution pattern to human body. *Blood*. 2004;103:860–867.
30. Chirumbolo S. State-of-the-art review about basophil research in immunology and allergy: is the time right to treat these cells with the respect they deserve? *Blood Transfus*. 2012;10:148–164.
31. Karasuyama H, Miyake K, Yoshikawa S, et al. Multifaceted roles of basophils in health and disease. *Allergy Clin Immunol*. 2018;142:370–380.
32. Hattangadi SM, Wong P, Zhang L, et al. From stem cell to red cell: regulation of erythropoiesis at multiple levels by multiple proteins, RNAs, and chromatin modifications. *Blood*. 2011;118:6258–6268.
33. Woolthuis CM, Park CY. Hematopoietic stem/progenitor cell commitment to the megakaryocyte lineage. *Blood*. 2016;127:1242–1248.
34. Pelayo R, Welner R, Perry SS, et al. Lymphoid progenitors and primary routes to becoming cells of the immune system. *Curr Opin Immunol*. 2005;17:100–107.
35. Bloom B, Spits H. Development of human lymphoid cells. *Annu Rev Immunol*. 2006;24:287–320.
36. Weerkamp F, Pike-Overzet K, Staal FJ. T-sing progenitors to commit. *Trends Immunol*. 2006;27:125–131.
37. Haddad R, Guimiot F, Six E, Jourquin F, et al. Dynamics of thymus-colonizing cells during human development. *Immunity*. 2006;24:217–230.
38. Spits H. Development of alphabeta T cells in the human thymus. *Nat Rev Immunol*. 2002;2:760–772.
39. Takahama Y. Journey through the thymus: stromal guides for T cell development and selection. *Nat Rev Immunol*. 2006;6:127–135.
40. Mahnke YD, Brodie TM, Sallusto F, et al. The who's who of T cell differentiation: human memory T cell subsets. *Eur J Immunol*. 2013;43:2797–2809.
41. Kumar BV, Connors TJ, Farber DL. Human T cell development, localization, and function throughout life. *Immunity*. 2018;48:202–213.
42. Chen J, Guan L, Tang L, et al. T helper 9 cells: a new player in immune-related diseases. *DNA Cell Biol*. 2019;38:1040–1047.

43. Snyder-Cappione JE, Tincati C, Eccles-James IG, et al. A comprehensive ex vivo functional analysis of human NKT cells reveals production of MIP1-α and MIP1-β, a lack of IL-17, and a Th1-bias in males. *PLoS One*. 2010;5:e15412.
44. Kumar A, Suryadevara N, Hill TM, et al. Natural killer T cells: an ecological evolutionary developmental biology perspective. *Front Immunol*. 2017;8:1858.
45. Sakaguchi S, Mikami N, Wing JB, et al. Regulatory T cells and human disease. *Annu Rev Immunol*. 2020;38:541–566.
46. Tobón GJ, Izquierdo JH, Cañas CA. B lymphocytes: development, tolerance, and their role in autoimmunity-focus on systemic lupus erythematosus. *Autoimmune Dis*. 2013;2013:827254.
47. Martin VG, Wu YB, Townsend CL, et al. Transitional B cells in early human B cell development—time to revisit the paradigm? *Front Immunol*. 2016;7:546.
48. Barone F, Vossenkamper A, Boursier L, et al. IgA-producing plasma cells originate from germinal centers that are induced by B cell receptor engagement in humans. *Gastroenterology*. 2011;140:947–956.
49. Farag SS, Caligiuri MA. Human natural killer cell development and biology. *Blood Rev*. 2006;20:123–137.
50. Fu B, Tian Z, Wei H. Subsets of human natural killer cells and their regulatory effects. *Immunology*. 2014;141:483–489.
51. Pahl JHW, Cerwenka A, Ni J. Memory-like NK cells: remembering a previous activation by cytokines and NK cell receptors. *Front Immunol*. 2018;9:2796.
52. Artis D, Spits H. The biology of innate lymphoid cells. *Nature.*. 2015;517:293–301.
53. Chan BCL, Lam CWK, Tam LS, et al. IL33: roles in allergic inflammation and therapeutic perspectives. *Front Immunol*. 2019;10:364.
54. Guyden JC, Martinez M, Chilukuri RV, et al. Thymic nurse cells participate in heterotypic internalization and repertoire selection of immature thymocytes; their removal from the thymus of autoimmune animals may be important to disease etiology. *Curr Mol Med*. 2015;15:828–835.
55. Hanabuchi S, Watanabe N, Liu YJ. TSLP and immune homeostasis. *Allergol Int*. 2012;61:19–25.
56. Steiniger BS. Human spleen microanatomy: why mice do not suffice. *Immunology*. 2015;145:334–346.
57. Wong BW, Zecchin A, García-Caballero M, et al. Emerging concepts in organ-specific lymphatic vessels and metabolic regulation of lymphatic development. *Dev Cell*. 2018;45:289–301.
58. Srinivasan S, Vannberg FO, Dixon JB. Lymphatic transport of exosomes as a rapid route of information dissemination to the lymph node. *Sci Rep*. 2016;6:24436.
59. Ferrante Jr. AW. The immune cells in adipose tissue. *Diabetes Obes Metab*. 2013;15:34–38.
60. Liu R, Nikolajczyk BS. Tissue immune cells fuel obesity-associated inflammation in adipose tissue and beyond. *Front Immunol*. 2019;10:1587.
61. Larsen SB, Cowley CJ, Fuchs E. Epithelial cells: liaisons of immunity. *Curr Opin Immunol*. 2020;62:45–53.
62. Newberry RD, Lorenz RG. Organizing a mucosal defense. *Immunol Rev*. 2005;206:6–21.
63. Ross KF, Herzberg MC. Autonomous immunity in mucosal epithelial cells: fortifying the barrier against infection. *Microbes Infect*. 2016;18:387–398.
64. Spasova DS, Surh CD. Blowing on embers: commensal microbiota and our immune system. *Front Immunol*. 2014;5:318.
65. Buck MD, Sowell RT, Kaech SM, et al. Metabolic instruction of immunity. *Cell*. 2017;169:570–586.
66. Hevia A, Delgado S, Sánchez B, et al. Molecular players involved in the interaction between beneficial bacteria and the immune system. *Front Microbiol*. 2015;6:1285.
67. Levy M, Thaiss CA, Elinav E. Metabolites: messengers between the microbiota and the immune system. *Genes Dev*. 2016;30:1589–1597.
68. Kurashima Y, Goto Y, Kiyono H. Mucosal innate immune cells regulate both gut homeostasis and intestinal inflammation. *Eur J Immunol*. 2013;43:3108–3115.

Innate Immunity

Douglas R. McDonald and Ofer Levy

Innate immunity is the first line of host defense against infection. All living organisms are continually exposed to microbes. For example, the human gut is colonized by trillions of commensal bacteria, fungi, and viruses (Chapter 22). The innate immune system must accommodate commensal microbes yet recognize and respond to pathogens. Potentially life-threatening infections can result from naturally occurring defects in the innate immune response (Chapter 3).

A defining characteristic of innate immunity is its existence before microbial exposure. Innate immune responses are induced rapidly by microbes and precede the development of adaptive immune responses. The adaptive immune system is characterized by the tremendous diversity of its receptors and its antigen ligands. The innate immune system responds to a more limited set of antigens that are typically essential and invariant structural components specific to microbes. These microbial components are known as *pathogen-associated molecular patterns* (PAMPs). They include microbial cell wall components and nucleic acids. PAMPs are recognized by pattern recognition receptors (PRRs), and they are highly potent and effective in initiating inflammatory responses.

"Trained immunity" refers to the phenomenon of enhanced innate immune responses following microbial exposure.[1] This increase in host resistance to reinfection can provide "cross-protection" against other infectious agents. For example, macrophages and natural killer (NK) cells can expand and contract their cell populations, upregulate genes involved in pathogen recognition and presentation, and secrete cytokines that augment the antimicrobial activity of bystander cells. Thus, there is an appreciation that the adaptive and innate immune systems have certain similar characteristics.

KEY CONCEPTS

The Innate Immune System

- Composed of barriers to the environment (e.g., skin), antimicrobial peptides and proteins, cells (e.g., neutrophils), and soluble factors (e.g., cytokines, chemokines and complement).
- Provides the initial immune response to microbes and primes the adaptive immune system.
- Differentiates between pathogens and commensals.
- Pathogen detection is mediated by germline-encoded pathogen recognition receptors (PRRs) that recognize invariant microbial structures known as pathogen-associated molecular patterns (PAMPS).
- Has a form of memory termed "trained immunity," whereby activation can modulate subsequent innate immune responses to unrelated stimuli or infections.

BARRIERS TO INFECTION

Skin and Mucosa

The epithelial layers of the skin (Chapter 23) and the linings of the gastrointestinal (GI) (Chapter 24), genitourinary (GU), and respiratory tracts provide a mechanical barrier to microbial entry and thus play an essential role in host defense. The stratum corneum of the skin is the first barrier encountered by microbes. The skin is persistently colonized with numerous microbes. Thus, an intact physical barrier is essential to prevent activation of the immune system under nonpathological conditions. Key cellular components of the skin's immune barrier include keratinocytes, dendritic cells (DCs), macrophages, T lymphocytes, and mast cells. These cells express a wide variety of pathogen recognition receptors and secrete a broad range of cytokines, chemokines, and antimicrobial proteins and peptides (APPs) that mediate inflammatory responses to infection. Genetic disorders of the skin that compromise skin integrity, such as epidermolysis bullosa (Chapter 63), can result in life-threatening infections.

Skin disorders that impair barrier function, such as atopic dermatitis (AD) (Chapter 48) or eczema are common. Filaggrin (FLG) is a key structural component of the outermost layer of the epidermis. Loss of function variants in filaggrin (R510X, 2282del4) is estimated to be present in up to 50% of patients with AD. FLG mutations are a risk factor for the development of early-onset AD and thus for sensitization to food and environmental allergens via increased permeability of allergens, leading to allergic rhinitis and asthma (Chapter 43) (the atopic march). Eczematous skin can lead to reduced expression of APPs and increased susceptibility to cutaneous bacterial (e.g., *Staphylococcus*, *Streptococcus*) and viral (e.g., herpes) infections.

The luminal surfaces of the intestines are sites of continual exposure to numerous microbes. Intestinal epithelial cells (IECs) (Chapter 22) protect against infection by forming a physical barrier through tight junctions and by producing mucus (goblet cells) and APPs. IECs express apical junction complexes, including E-cadherin, ZO-1, claudin, and occludin, which function to form a tight monolayer that prevents penetration by bacteria.[2] A breakdown in epithelial gut homeostasis can lead to inflammatory bowel diseases (e.g., Crohn disease, ulcerative colitis) (Chapter 75) and increased susceptibility to bacterial infection.

Influenza viruses and respiratory syncytial viruses replicate in airway epithelial cells, leading to cell death and inflammation. The subsequent impaired barrier function of the airways can lead to increased susceptibility to secondary invasive bacterial infections by *Streptococcus pneumoniae* and other pyogenic

bacteria. Inflammatory bowel diseases also result in impaired barrier functions of the small and large intestines, which can be associated with increased translocation of bacteria across gut mucosa, potentially leading to serious infection.

CLINICAL RELEVANCE

Innate Immunity: Barriers

- Barrier function is an underappreciated component of the innate immune system.
- Barriers include the epithelial layers of the skin and the gastrointestinal, respiratory, and genitourinary tracts.
- Defects in barriers (e.g., epidermolysis bullosa and atopic dermatitis) increase the risk of infection.
- Production of antimicrobial peptides and proteins at barrier sites helps prevent microbial invasion.

Antimicrobial Proteins and Peptides

Among the APPs produced by the skin and by the epithelia of the gastrointestinal, genitourinary, and respiratory tracts are *bactericidal/permeability-increasing protein* (BPI), *defensins* (β-strand peptides connected by disulfide bonds), and *cathelicidins* (linear α-helical peptides) (Table 3.1).[3] Most APPs have a net positive charge, which enhances their affinity for negatively charged microbial cell membranes. Binding of APPs to microbes can permeabilize microbial membranes, leading to their destruction.

BPI is a ~55-kilodalton (kDa) cationic and hydrophobic protein with high affinity for the lipid A region of lipopolysaccharide (endotoxin). It is found in neutrophil primary (azurophilic) granules and is also inducible in epithelial cells. BPI shields against gram-negative bacteria via its microbicidal, opsonic, and endotoxin-neutralizing properties.[4] Neutralization of endotoxin may serve to limit inflammatory responses to gram-negative bacteria.

Some APPs, such as lysozyme (Lz), have enzymatic activities, which cleave peptidoglycans found in bacterial cell walls. Other APPs bind to and compete for nutrients, a form of so-called nutritional immunity. Lactoferrin (Lf), for example, binds iron, a nutrient essential to bacterial survival.[4]

Defensins are classified by the linking pattern of cysteines and their sizes. Alpha-defensins are expressed in neutrophils and Paneth cells of the small intestine (Chapter 22), whereas β-defensins are expressed by mucosal surface epithelia, including those of skin, eyes, and the oral, urogenital, and respiratory tracts. Defensins exhibit broad specificity of antimicrobial activity against bacteria, mycobacteria, fungi, parasites, and viruses (Table 3.2). They have also been shown to enhance antigen uptake and processing, and to stimulate the chemotaxis of monocytes, macrophages, and mast cells.[5] The expression of several of the defensins is constitutive. For others, inflammatory stimuli (bacterial products, proinflammatory cytokines) will increase defensin expression (human neutrophil proteins 1–3 and human β-defensin-2). Given the increasing incidence of antibiotic-resistant bacteria, there is great interest in the potential uses of APPs as treatment for bacterial infections and infections with multidrug-resistant organisms.[6] Exogenous administration of APPs may be most effective in those with reduced APP expression; for example, in the setting of neutropenia.[7]

TABLE 3.1 Epithelial Antimicrobial Proteins and Peptides

Antimicrobial Peptide	Source	Target Organism
Dermicidin	Eccrine sweat glands	Broad spectrum
Psoriasin	Keratinocytes, sebocytes	G^-
RNase 7	Keratinocytes	Broad spectrum
RNase 5/angiogenin	Keratinocytes	*Candida albicans*
Cathelicidin (LL-37)	Keratinocytes, sebocytes	G^+, G^-
BPI	Epithelia-oral, GI, urogenital tract	G^-, (G^+, fungi)
hBD-1	Keratinocytes, sebocytes	G^-
hBD-2	Keratinocytes, sebocytes	G^-
hBD-3	Keratinocytes	Broad spectrum
hBD-4	Keratinocytes	G^+, G^-
SLPI	Keratinocytes	Broad spectrum
Elafin	Keratinocytes	Broad spectrum
Adrenomedullin	Keratinocytes, hair follicles, eccrine/apocrine sweat glands, sebocytes	G^+, G^-
MIP-3α/CCL20	Keratinocytes	Broad spectrum
Lysozyme	Keratinocytes, sebocytes, hair bulb cells	G^+, G^-
Lactoferrin	Milk, saliva, tears, nasal secretions, neutrophils	Broad spectrum

BPI, Bactericidal/permeability-increasing protein; *CCL*, chemokine ligand; G^+, gram-positive; G^-, gram-negative; *GI*, gastrointestinal; *hBD*, human β-defensin; *MIP*, macrophage inflammatory protein; *RNase*, ribonuclease; *SLPI*, secretory leukocyte peptidase inhibitor.

TABLE 3.2 Neutrophil-Derived Antimicrobial Proteins and Peptides (APPs)

Neutrophil APP	Granule Type	Target Organism
Lysozyme	Azurophil, specific	G^+, G^-
Azurocidin	Azurophil, secretory	G^+, G^-, *Candida albicans*
Elastase	Azurophil	G^+, G^-
Cathepsin G	Azurophil	G^+, G^-
Proteinase 3	Azurophil	G^+, G^-
BPI	Azurophil	G^-, (G^+, fungi)
α-Defensins (HNP-1 to -4)	Azurophil	G^+, G^-, fungi, viruses
Cathelicidin (hCAP-18)	Specific	G^+, G^-, mycobacteria
Lactoferrin	Specific	G^+, G^-, fungi, viruses
SLPI	Specific	G^+, G^-, *Aspergillus fumigatus, C. albicans*
NGAL	Specific	G^+, G^-, fungi
Lysozyme	Azurophil, specific	G^+, G^-
Azurocidin	Azurophil, secretory	G^+, G^-, *C. albicans*
Elastase	Azurophil	G^+, G^-
Cathepsin G	Azurophil	G^+, G^-

BPI, Bactericidal/permeability-increasing protein; G^+, gram-positive; G^-, gram-negative; *hCAP*, human cathelicidin antimicrobial protein; *HNP*, human neutrophil peptide; *NGAL*, neutrophil gelatinase-associated lipocalin; *SLPI*, secretory leukocyte peptidase inhibitor.

KEY CONCEPTS

Humoral Innate Immunity

- Chemokines and cytokines are essential mediators.
- Cytokine influences are both redundant and pleotropic.
- Due to the risk of collateral tissue damage, cytokine synthesis is tightly controlled.
- Acute phase reactants are induced by cytokines and promote opsonization of microbes.
  - The acute phase reactant C-reactive protein (CRP) is induced by IL-6.
  - The CRP plasma level can be used to monitor infections and inflammation.
- The complement system protects against a wide variety of microbes.

HUMORAL INNATE IMMUNITY

The Acute Phase Response

A variety of soluble proteins found in plasma help recognize PAMPs and function as mediators of innate immunity. The cytokines tumor necrosis factor-α (TNF-α) and interleukin-1β (IL-1β) (Chapter 14) induce production of acute phase reactants in hepatocytes, including members of the pentraxin family (e.g., serum amyloid A [SAA], serum amyloid P [SAP], and C-reactive protein [CRP]). These pentraxins bind to components of the bacterial cell wall. TNF-α and IL-1β also induce production of IL-6 (Chapter 14) from mononuclear phagocytes, endothelial cells, and fibroblasts. IL-6 is another potent inducer of acute phase reactants, including CRP and fibrinogen. An *opsonin* is a molecule, including an antibody, that can enhance phagocytosis by marking an antigen for an immune response or marking dead cells for recycling. CRP, SAA, and SAP function as opsonins. They can bind phosphorylcholine and phosphatidylethanolamine expressed on bacteria and apoptotic cells, thereby enhancing phagocytosis by marking these bacteria and apoptotic cells as targets for macrophages.

Lipopolysaccharide-binding protein (LBP) is an acute phase reactant synthesized by the liver in response to gram-negative bacterial infections. LBP binds to LPS and subsequently forms a complex with CD14, TLR4, and MD-2, which functions as a high-affinity receptor for LPS.

Mannose-binding lectin (MBL) is a member of the calcium-dependent (C-type) lectins (collectins) produced by the liver in response to infection. MBL binds to carbohydrates with terminal mannose and fucose residues that are expressed on microbial cell surfaces.[8] MBL can bind to the C1q receptor on macrophages to enhance phagocytosis and can activate the complement system (Chapter 40) via the lectin pathway (discussed below).

Surfactant protein A and surfactant protein D are collectins expressed in the lung and can bind a variety of microbes and inhibit their growth.[6] They also function as opsonins that promote ingestion by alveolar macrophages.

Finally, ficolins are plasma proteins capable of binding to several types of bacteria and can activate complement.

The Complement System

The complement system comprises a collection of plasma proteins activated by microbes (Chapter 40). It helps mediate microbial destruction and inflammation.[9] Complement activation can occur via three pathways: the classical pathway (CP), the alternative pathway (AP), and the lectin pathway (LP).

In the classical pathway, complement C1 detects immunoglobulin M (IgM), IgG1, or IgG3 bound to the surface of a microbe. C1 is composed of the C1q, C1r, and C1s subunits. These form multimeric complexes that recognize IgM or IgG bound to microbial surfaces. C1r and C1s are serine proteases. Activated C1s generates a C3 convertase composed of C4b and C2b (C4b2b) bound to the microbial surface. C3 convertase cleaves C3, generating C3b. C3b binds covalently to C4b2b, generating C5 convertase. C5 convertase then activates the late steps of complement activation, leading to assembly of the membrane attack complex (MAC) and subsequent cytolysis (Fig. 3.1).

The alternative pathway is initiated by small amounts of C3b, which are spontaneously generated in plasma. C3b that remains unbound to a cell surface is rapidly hydrolyzed and inactivated. C3b bound to a microbe becomes a binding site for factor B. Bound factor B is cleaved by factor D, generating factor Bb that binds covalently to C3b, forming the AP C3 convertase, which activates the later steps of complement activation, as in the CP (see Fig. 3.1).

The lectin pathway is activated by MBL or ficolins binding to microbial surfaces. MBL then binds to MBL-associated serine proteases (MASPs)-1, -2, and -3. MASP-2 cleaves C4 and C2 to activate the complement cascade, as in the CP (see Fig. 3.1).

Complement components also function as opsonins. Complement-coated microbes can be phagocytosed via complement receptors on phagocytes. Complement receptor type 1 (CR1) is a high-affinity receptor for the C3b and C4b fragments of complement and mediates the internalization of C3b- and C4b-coated particles. On erythrocytes, CR1 mediates clearance of immune complexes from the circulation. Complement type 2 receptor (CR2, also known as CD21) is expressed on B cells (Chapter 7) and follicular dendritic cells (FDCs) (Chapter 6). It binds C3 proteolytic fragments, including C3d, C3dg, and iC3b. CR2 augments humoral immune responses by enhancing B-cell activation by antigen and by promoting trapping of antigen–antibody complexes in germinal centers.[10] Epstein-Barr virus (EBV) can use CR2 as a receptor, thereby allowing EBV to enter B cells. Complement receptor 3 (CR3) is composed of a heterodimer of CD18 and CD11b and is expressed in polymorphonuclear neutrophils (PMNs), monocytes, and macrophages. CR3 binds to iC3b bound to the surface of microbes, leading to phagocytosis and destruction of the pathogen. Activation of complement via the AP can greatly enhance monocyte-generated TNF-α elicited by gram-positive bacteria, such as group B streptococcus.

KEY CONCEPTS

Humoral Innate Immunity: Complement

- Defects in early complement cascade components paradoxically promote both immune deficiency (invasive infections with encapsulated bacteria) and autoimmunity (lupus-like syndromes).
- Defects in late complement cascade components (C5–9) promote susceptibility to meningitis due to *Neisseria meningitidis*.
- Defects in C1 inhibitor protein (or function) underlie hereditary angioedema.
- Defects in factor H are associated with membranoproliferative glomerulonephritis, hemolytic–uremic syndrome, and age-related macular degeneration.
- Defects in mannose-binding lectin (MBL) create susceptibility to bacterial infection in individuals with comorbid conditions (e.g., chemotherapy, cystic fibrosis).

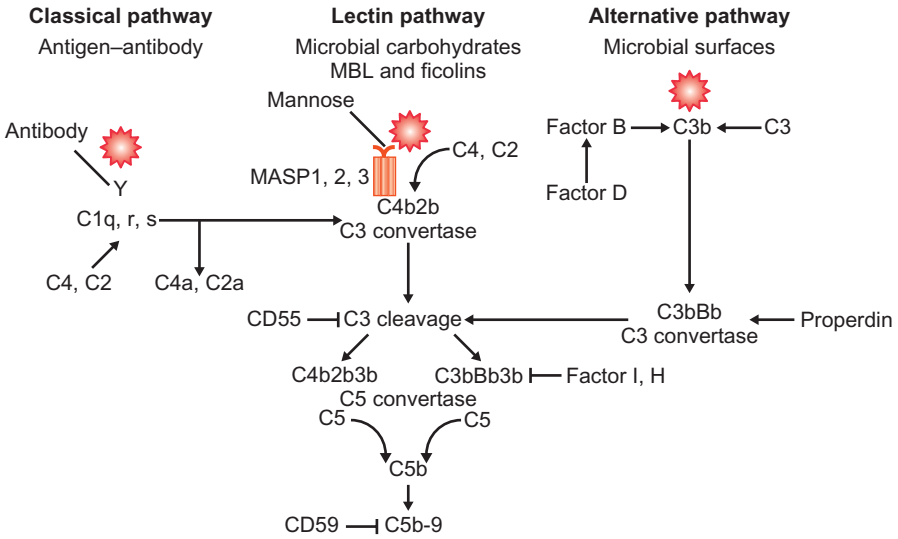


FIG. 3.1 Complement Activation Pathways. The classical complement cascade is activated by antibody bound to microbial surfaces, which is a binding site for the C1 complex. The alternative pathway is activated by the binding of spontaneously generated C3b to microbial surfaces. Microbe-bound C3b binds factor B, which is converted to factor Bb, forming a C3 convertase. The lectin pathway is activated by the binding of mannose-binding lectin (MBL) to mannose residues on microbial surfaces. MBL binds MBL-associated serine proteases, which bind and cleave C4 and C2, forming a C3 convertase.

There are multiple regulatory proteins within the complement pathways (see Fig. 3.1). C1 esterase inhibitor (C1INH) binds to, and inhibits, the enzymatic functions of C1r and C1s within the CP.[11] Properdin stabilizes C3bBb complex, increasing the life span of the AP C3 convertase. Conversely, factor H inhibits the formation of and degrades C3bBb complexes. Factor I inactivates C3b. CD55 (decay accelerating factor) and CD59 are cell surface, glycophosphatidylinositol (GPI)-linked proteins that block complement-mediated cytolysis by inhibiting formation of C3bBb complex and binding of C9 to C5b678 complex, respectively. Paroxysmal nocturnal hemoglobinuria, an acquired defect in the *PIGA* gene that causes a deficiency of GPI-linked proteins, is the result of absent cell surface expression of CD55 and CD59 that leads to hemolytic anemia caused by complement-mediated lysis of red blood cells (RBCs).

Complement Deficiency Diseases

Deficiencies of early components of the complement pathway are associated with invasive bacterial infections caused by encapsulated organisms (Chapter 40). Lack of early components of the complement pathway are also associated with rheumatic disorders, including a lupus-like syndrome that may be caused by impaired immune complex clearance, impaired clearance of apoptotic cells, and loss of complement-dependent B-cell tolerance (Chapter 52). Deficiency of factor I is also associated with increased incidence of invasive infection with encapsulated bacteria (Chapter 27), as well as glomerulonephritis (Chapter 68) and autoimmune disease (Chapter 40).

Deficiency of C1INH protein and function, either hereditary or acquired, leads to hereditary angioedema (HAE) or acquired angioedema (AAE) (Chapter 40). C1INH inhibits C1, factors XIa and XIIa, and kallikrein. Dysregulation of these cascades leads to generation of vasoactive products that result in angioedema. Deficiencies of late components of complement, including C5 through C9, as well as factors B, D, and properdin create susceptibility to meningococcal infections.[12] Deficiency of factor H function is associated with membranoproliferative glomerulonephritis (Chapter 68), hemolytic–uremic syndrome, and age-related macular degeneration (AMD) (Chapter 74). Deficiency of MBL is associated with increased susceptibility to bacterial infections in infancy and particularly in individuals with other comorbid conditions, such as cystic fibrosis.

KEY CONCEPTS

Cellular Innate Immunity

- While short lived, polymorphonuclear leukocytes (neutrophils) are the most abundant and earliest cells of the innate immune system to respond to infection.
- Several days later in an infection, monocytes and macrophages predominate.
- Activated neutrophils, monocytes, and macrophages kill phagocytosed bacteria with reactive oxygen intermediates and antimicrobial peptides and proteins (APPs).
- Dendritic cells take up and present foreign antigens and thus link the innate to the adaptive immune system.
- Natural killer (NK) cells can kill infected or malignant cells without prior activation.
- Mast cells are found at the interface between the host and the environment, are first responders to microbes, and recruit other inflammatory cells.

CELLULAR INNATE IMMUNITY

Polymorphonuclear Leukocytes

PMNs are the most abundant leukocyte (Chapter 39). They have a short life span of ~6 hours in circulation, and in the healthy adult, ~10^9 PMNs are produced per hour. PMNs are readily identified by light microscopy by segmented nuclei divided into 3 to 5 lobules. Their cytoplasm contains four types of granules: azurophilic (or primary), specific (or secondary),

gelatinase, and secretory. PMN granules contain a wide variety of APPs with a broad spectrum of antimicrobial activities (see Table 3.2). Azurophilic granules contain enzymes, such as proteinase 3, cathepsin G, and elastase, as well as α-defensins and BPI. Specific granules contain lactoferrin and the proforms of cathelicidin peptides. Gelatinase granules are rich in gelatinase and are a marker of terminal neutrophil differentiation. Secretory granules contain a variety of receptors that are inserted into the cell membrane upon activation. Exocytosis of these receptors convert PMNs into cells more responsive to inflammatory stimuli. PMNs are the earliest responders to infection. Those not recruited to sites of infection undergo apoptosis and are cleared by the reticuloendothelial system. Individuals with severely low numbers of neutrophils (<500 cells/μL) are susceptible to overwhelming bacterial infections.

KEY CONCEPTS

Defects in Neutrophil Number or Function

- Severe neutropenia (<500/μL) creates susceptibility to overwhelming bacterial infection, regardless of whether it is primary or secondary.
- Defective production of reactive oxygen intermediates leads to chronic granulomatous disease, which is marked by:
  - Susceptibility to invasive bacterial and fungal infections, and
  - Impaired wound healing.
- Myeloperoxidase deficiency is asymptomatic in the majority of affected individuals.
  - Candidal infections (mucocutaneous and invasive) have been reported.

Monocytes and Macrophages

Mononuclear phagocytes include monocytes and macrophages. Monocytes originate in bone marrow and migrate into the peripheral circulation. $CD14^+$ monocytes are efficient in phagocytosis and in the production of reactive oxygen intermediates (ROIs) and proinflammatory cytokines in response to a wide variety of microbial stimuli. A subset of monocytes with low CD14 expression ($CD14^{dim}$), but expressing CD16, is associated with vascular endothelia and appears to be specialized for response to viruses and nucleic acid–containing immune complexes. This subset may also be involved in the pathogenesis of autoimmune disorders.[13] $CD14^+$ monocytes enter tissues where they mature into macrophages.

Distinct macrophage populations in different tissues are given specific names, including Küpffer cells in the liver, alveolar macrophages in the lungs, osteoclasts in bone, and microglia within the brain. Macrophages differ from PMNs in that they are not terminally differentiated and can proliferate at sites of infection. They are longer-lived than PMNs and are the predominant innate immune cell several days after an infection.

Macrophages display considerable plasticity in their functions, depending on the cytokine milieu. Classically activated macrophages (M1) are induced by Th1 cytokines, like interferon-γ (IFN-γ), TNF-α, and bacterial products, leading to enhanced killing of phagocytosed microbes. M1 macrophages secrete higher amounts of TNF-α, IL-1α, IL-1β, IL-6, IL-12, and IL-23, and lower amounts of IL-10. M1 macrophages can cause ROS-induced tissue damage and impair tissue regeneration and wound repair. In contrast, alternatively activated macrophages (M2) are induced by Th2 cytokines (e.g., IL-4, IL-13) and have anti-inflammatory functions and promote healing. M2 macrophages have been shown to inhibit T-cell activation through production of IL-10 and transforming growth factor-β (TGF-β).[14]

Recently, animal models evaluating SARS-CoV infection in macaques has demonstrated that antibodies against the S protein of SARS-CoV can activate FcγR on M2 macrophages in the lung, triggering an exaggerated inflammatory response from classically anti-inflammatory M2 macrophages. This hyperinflammatory response is characterized by the production of large quantities of IL-6 and IL-8 (CXCL-8), and recruitment of inflammatory cells to the lungs, leading to acute lung injury (ALI), diffuse alveolar damage (DAD), and death.[15] Thus, in certain inflammatory conditions, even anti-inflammatory M2 macrophages can contribute to tissue damage.

Neutrophil and Macrophage Microbicidal Mechanisms

Microbicidal Molecules

Activated PMNs and macrophages kill phagocytosed bacteria by releasing microbicidal molecules both extracellularly and within phagolysosomes. Microbes are detected by pattern recognition receptors, as well as by Fc and complement C3 receptors. Bacteria are internalized into phagosomes. Phagosomes fuse with lysosomes containing proteolytic enzymes (e.g., elastase, cathepsin G) to form phagolysosomes.

Reactive Oxygen Intermediates

Activated PMNs and macrophages produce ROIs, which are toxic to microbes. ROIs are produced by phagocyte-derived nicotinamide adenine dinucleotide phosphate (NADPH) oxidase, a multisubunit enzyme. NADPH oxidase consists of five subunits, p22phox, p40phox, p47phox, p67phox, and gp91phox. This phagocyte oxidase is activated following engulfment of opsonized bacteria (oxidative burst). Genetic defects in components of the NADPH oxidase complex create susceptibility to invasive infections with bacteria and fungi (chronic granulomatous disease), as well as impaired wound healing (Chapter 39).

A variety of stimuli activate the phagocyte oxidase complex, including the complement fragment C5a; formylated peptides, such as FMLP (*N*-formyl-methionine-leucine-phenylalanine); LTB4 (leukotriene B4); PAF (platelet-activating factor); and pattern recognition receptors, such as TLR4. Upon cellular activation, p40phox, p47phox, and p67phox are phosphorylated and recruited to cellular membranes, where they associate with membrane-bound gp91phox and p22phox (flavocytochrome b558) and GTP-bound Rac1 (monocytes) or Rac2 (PMNs). The activated enzyme generates superoxide radicals, which are then converted to hydrogen peroxide by superoxide dismutase. Hydrogen peroxide is combined with halide ions by myeloperoxidase to generate hypohalous acids, which are toxic to bacteria.

The phagocyte oxidase complex also generates an environment within the phagolysosome conducive to proteolytic enzyme activation. The oxidase functions as an electron pump that generates an electrochemical gradient across the phagolysosomal membranes, which is compensated by the movement of ions into the vacuole. This results in the increase in vacuolar pH and osmolarity required for activation of the antimicrobial proteases elastase and cathepsin G.[16]

Neutrophil Extracellular Traps

In addition to phagocytosis of microbes, release of neutrophil extracellular traps (NETs) is another mechanism employed by neutrophils to neutralize and kill a variety of pathogens, including

bacteria, fungi, parasites and viruses.[17] NETs are extracellular, web-like structures composed of decondensed chromatin imbedded with a variety of cytosolic and granule proteins. These proteins include neutrophil elastase, myeloperoxidase, calprotectin, cathelicidins and defensins. The composition of NETs may vary, depending upon the stimulus. NET release is triggered by a cell death process, termed NETosis.[17] Generation of reactive oxygen species (ROS) is critical in generation of NETs and neutrophils from patients with chronic granulomatous disease (CGD), who are unable to generate ROS, are unable to produce NETs, likely contributing their susceptibility to invasive bacterial and fungal infections.

A variety of signaling pathways trigger NETosis, including TLR2, TLR4, TLR7, TLR8, Dectin2, complement receptor 3, Siglec-14, FcγRIIIb and receptor for advanced glycation endproducts (RAGE).[17] NETosis is typically triggered by larger, difficult to phagocytose, pathogens, such as fungal hyphae. NETosis is a tightly regulated process and excessive NET formation has been shown to induce lung damage in pulmonary fungal infections. In addition, it is hypothesized that NETs may contribute to the pathogenesis of autoimmune diseases, such as rheumatoid arthritis (Chapter 53) and systemic lupus erythematosus (Chapter 52).

Reactive Nitrogen Intermediates

Macrophages produce reactive nitrogen intermediates in response to microbes. Nitric oxide (NO) is produced by inducible nitric oxide synthetase (iNOS). Expression of iNOS is induced by activation of Toll-like receptors (TLRs), and expression is augmented further by IFN-γ.[18] iNOS catalyzes the conversion of arginine to citrulline, releasing diffusible nitric oxide gas. Within phagolysosomes, NO combines with hydrogen peroxide or superoxide to produce peroxynitrite radicals, which contribute to microbial killing. Although ROIs and NO are effective antimicrobial agents, they are nonspecific and are also capable of inducing damage to host tissues.

Dendritic Cells

Dendritic cells (Chapter 6) have long membranous extensions for surveying the local environment and are highly phagocytic. They link innate to adaptive immune responses after activation by microbes. DCs express a variety of PRRs, which allow them to respond to microbes by phagocytosis and cytokine secretion. Activated DCs rapidly take up antigens and then home to draining lymph nodes where they localize to T-cell zones. During their migration to lymph nodes, DCs mature and become efficient antigen-presenting cells (APCs). Once in the lymph node, DCs express high levels of costimulatory molecules, such as B7 and IL-12p70, and present antigen to naïve T cells, inducing their differentiation into effector T cells (Th1 T cells). Plasmacytoid dendritic cells (pDCs) are specialized for response to viral infection and secrete large amounts of type 1 IFNs.

One subset of DC characterized by $CD11c^{high}CD103^{+}$ expression in the lamina propria of the small intestine (Chapter 24) facilitates the differentiation of regulatory T cells (Chapter 13) in a retinoic acid– and TGF-β–dependent manner. Such DC subsets may play a role in the development of tolerance to commensal bacteria (Chapter 22).

Natural Killer Cells

NK cells are derived from common lymphoid progenitor cells and constitute 5% to 20% of mononuclear cells in the periphery (Chapter 12). They do not express somatically rearranged antigen receptors. Target cells are identified using germline DNA-encoded receptors. NK cells are divided into two subsets, $CD56^{bright}CD16^{-}$ and $CD56^{dim}CD16^{+}$, which have different functions. $CD56^{dim}$ NK cells account for roughly 90% of NK cells in the periphery and express the low-affinity Fcγ receptor (CD16), which mediates antibody-dependent, cell-mediated cytotoxicity (Chapter 8). $CD56^{bright}$ NK cells are poorly cytotoxic but produce large amounts of cytokine and represent the majority of NK cells in peripheral lymphoid organs. NK cells are a major source of IFN-γ, which augments the microbicidal functions of macrophages. Conversely, NK cells are primed by IL-15 derived from DCs and IL-12 or IL-18 derived from macrophages, demonstrating the regulatory interactions that occur between NK cells and other cells of the immune system.

NK-cell function is regulated by a delicate balance between signals generated by inhibitory and activating receptors. NK cells possess the ability to recognize and kill infected or malignantly transformed cells, while leaving healthy host cells unharmed. Inhibitory receptors on NK cells recognize class I major histocompatibility complex (MHC) molecules expressed on most healthy cells in the body, preventing NK-cell activation.

NK inhibitory receptors include three families of receptors: heterodimers composed of CD94 and NKG2A, the Ig-like transcripts (i.e., ILT-2), and the killer cell Ig-like receptor (KIR) family (Chapter 12). The immunoreceptor tyrosine-based inhibition motifs (ITIMs) in their cytoplasmic tails recruit phosphatases (Src homology region 2 [SH2] domain-containing phosphatase-1 [SHP-1], SHP-2, and SHIP [SH2-containing inositol polyphosphate 5-phosphatase]), which oppose the effects of kinases activated by activating receptors. When NK cells encounter host cells expressing MHC class I molecules, protein tyrosine phosphatases are activated, reducing signaling downstream of activating receptors and opposing NK-cell activation (Fig. 3.2).

NK cells also possess activating receptors. CD16 mediates antibody-dependent, cell-mediated cytotoxicity and natural cytotoxicity receptors (e.g., NKp46, NKp30, NKp44, NKG2D, CD94/NKG2C, 2B4). Activating receptors are linked to molecules (e.g., CD3-ζ, FcR-γ, or DAP12) that contain immunoreceptor tyrosine-based activation motifs (ITAMs). Upon ligand binding, tyrosine residues within ITAMs are phosphorylated by Src family kinases. The tyrosine phosphorylated ITAMs serve as binding sites for the activation of other protein tyrosine kinases, such as Syk and ZAP-70, which activate downstream effector molecules in a signaling cascade. Infection of host cells with some viruses can lead to reduced MHC class I expression (Chapter 5), thereby reducing viral antigen presentation to T cells. Reduced expression of MHC class I by infected cells promotes activation of NK cells, which then kill infected cells. Concomitantly, ligands for activating receptors are expressed by the infected cell, also leading to NK-cell activation and killing of the infected cell.

NK cells play an important role in immunosurveillance against tumors (Chapter 80). In humans, NK-cell receptors that mediate tumor recognition include NKp46, NKp30, NKp44, DNAM-1 (DNAX accessory molecule-1), and NKG2D. Ligands expressed on target cells include MHC I-related chain (MIC)-A, MICB, unique long 16-binding proteins (ULBP), poliovirus receptor (PVR), and nectin-2. DNAM-1 specific ligands include PVR and nectin-2, which are expressed in cell lines that include carcinomas, melanomas, and neuroblastomas. Nectin expression is not specific to tumors, since nectins are expressed on normal

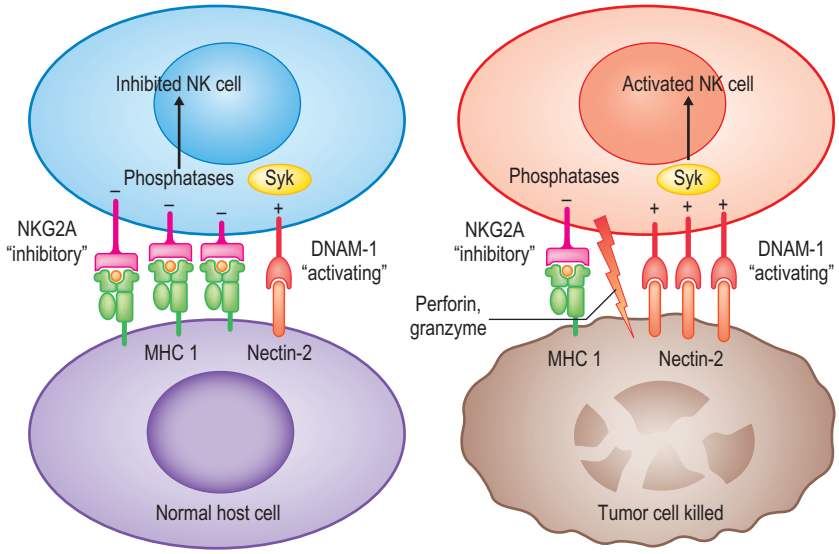


FIG. 3.2 Regulation of Natural Killer *(NK)* Cell Function. Upon encountering normal host cells, inhibitory receptors on NK cells that contain immunoreceptor tyrosine-based inhibitory motifs (ITIMs) preferentially activate phosphatases (e.g., SHP-1/2, SHIP) that send inhibitory signals, inhibiting NK-cell function. NK cells that encounter virally infected cells or tumor cells receive signals through activating receptors that contain immunoreceptor tyrosine-based activation motifs (ITAMs) that activate tyrosine kinases (e.g., Syk), leading to NK-cell activation, release of perforin and granzyme, and target cell death.

cells. DNAM-1–nectin interactions on normal cells do not result in NK-cell lysis because normal cells are protected by MHC class I expression. Conditions favoring NK cell–mediated lysis include tumors in which nectins are overexpressed and/or MHC class I expression is reduced, favoring NK-cell activation (see Fig. 3.2).

Natural killer T (NKT) cells are a small, but highly variable, population of thymus-derived T cells that express NK-cell markers and a restricted repertoire of T-cell receptors (TCRs) that recognize lipids bound to the MHC-like molecule CD1d (Chapter 5).[19] Type 1 NKT cells (also referred to as invariant NKT [iNKT] cells) express the invariant Vα24 and Jα28 TCR α chain, whereas type 2 NKT cells have a more diverse TCR repertoire. Mature human NKT cells can be further divided into three groups, $CD4^+CD8^-$, $CD4^-CD8^-$, and $CD4^-CD8^+$ subsets. The most completely characterized NKT antigen is the lipid α-galactosylceremide (α-GalCer), which is often used to activate NKT cells experimentally.

Identification of natural NKT ligands has proven difficult. NKT cells express perforin and granulysin and are capable of cytotoxic activity. NKT cells are also able to influence innate and adaptive immune responses through release of large amounts of cytokines, including IFN-γ, TNF-α, IL-4, IL-13, IL-10, and granulocyte macrophage–colony-stimulating factor (GM-CSF). In general, NKTs found in blood can produce large amounts of cytokines, whereas NKTs in the thymus are poor cytokine producers.

Decreased NKT cell frequency and/or function may increase susceptibility to some autoimmune diseases, including type 1 diabetes (Chapter 71) and multiple sclerosis (Chapter 66). Mice with NKT defects are susceptible to tumors and adoptive transfer of normal NKTs can provide protection against tumors (Chapter 80). NKT cells may also contribute to the pathogenesis of the airway hyperresponsiveness (AHR) in asthma (Chapter 43), which is dependent on IL-4 and IL-13 production in the airways. iNKT cells are necessary for AHR in several murine models of asthma, since NKT-deficient mice fail to develop AHR following allergen challenge, ozone challenge, or viral infection.[20] iNKT cell deficiency was associated with severe varicella infection, demonstrating a role for iNKT cells in innate antiviral immunity.

Intraepithelial Lymphocytes, B-1 and MZ B Cells, Innate Lymphoid Cells, and Mast Cells

Intraepithelial Lymphocytes

Barrier epithelia of skin and the GI tract contain unique types of lymphocytes, including intraepithelial T lymphocytes (IELs) and B-1 B cells (Chapter 7), which respond to commonly encountered microbes. Because of their more limited diversity of receptors, IELs can be considered part of the innate immune system. The main immune cell populations within the epidermal layer include keratinocytes, melanocytes, a type of DC known as the Langerhans cell, and IELs (Chapter 23).

Keratinocytes and melanocytes express a variety of PRRs, enabling detection of microbes, resulting in secretion of cytokines that can contribute to innate immune responses through recruitment and activation of phagocytes. Langerhans cells form an elaborate network of dendritic processes that allow them to capture antigens that gain access to skin. Following activation by microbes, Langerhans cells migrate to draining lymph nodes and express chemokine receptor-7 (CCR7) (Chapter 15), which allows them to migrate to the T-cell zones within lymph nodes in response to the chemokine ligands CCL 19 and CCL 21 and present antigen to T cells.[21]

Intraepidermal T lymphocytes constitute roughly 2% of lymphocytes within the skin. This lymphocyte subset expresses a more restricted set of antigen receptors, which include both αβ and γδ TCRs (Chapter 4), similar to IELs found in the intestines. These specialized T cells appear to be committed to recognizing microbial peptide antigens commonly found at epithelial surfaces and thus function as components of the innate immune system.

IELs are a significant component of the GI immune system and reside at the basolateral side of the intestinal epithelial cell

layer (Chapters 2 & 24). IELs are among the first immune cells to encounter pathogens that have breached intestinal epithelia. IELs consist of CD8 T cells (Chapter 12), as well as memory-effector T cells bearing αβ or γδ TCRs. IELs contain a greater proportion of TCRγδ$^+$ cells than is found in the peripheral circulation.[22]

CD4$^+$TCRαβ$^+$ and CD8$^+$αβTCRαβ$^+$ IELs are MHC class II and class I restricted (Chapter 5), respectively. These IELs have likely undergone thymic selection (Chapter 9) and subsequently homed to the gut after antigenic stimulation. As such, these IELs are likely specific for foreign antigens.[22] They have a memory phenotype and an oligoclonal TCR repertoire. IELs in the small intestine frequently express CD8$^+$αα (CD4$^+$CD8$^+$αα or CD8$^+$αβCD8$^+$αα), a characteristic of activated mucosal T cells within the gut microenvironment. Upon antigenic stimulation, CD8$^+$αβTCRαβ$^+$ IELs are cytolytic and kill via granzymes and perforin or through engagement of Fas (Chapter 17).[22]

TCRγδ$^+$ IELs emigrate from the thymus and subsequently take up residence in the intestinal epithelium (Chapter 24). They constitute approximately 10% of intestinal IELs in humans, and the majority expresses CD8αα. TCRγδ$^+$ IELs recognize nonclassic MHC molecules (Chapter 5), such as thymus leukemia antigen or MHC class I–like molecules, MICA (MHC I–related chain [MIC]-A) and MICB, and may help modulate inflammatory immune responses. These IELs can be cytolytic and express FasL. TCRγδ+ IELs can produce keratinocyte growth factor and promote intestinal epithelial integrity.[22]

B1 and Marginal Zone B Cells

The B1 and marginal zone (MZ) subsets of B lymphocytes have been characterized as innate-like B cells (Chapter 7). They express antigen receptors enriched for germline sequence. These cell types have been mostly studied in mice. Their identity in humans is less clear. B1 cells and MZ B cells can function as APCs, but unlike conventional B cells, B1 cells and MZ B cells do not develop into memory B cells. B1 cells and MZ B cells share characteristics: (i) they are the main source of natural antibodies; (ii) they express high surface levels of IgM and low surface levels of IgD; and (iii) they are rapidly activated by microbes through pattern recognition receptors to produce large amounts of natural antibodies. B1 cells and MZ B cells produce IL-10 upon activation, which may downregulate immune responses. The natural antibodies produced by B1 cells and MZ B cells function as the first line of defense against invading microbes.

Innate Lymphoid Cells

Innate lymphoid cells (ILCs) are a heterogeneous population of cells. ILCs do not express rearranged antigen-specific receptors. This lymphoid subset includes killer ILCs (e.g., NK cells) and helper ILCs. Helper ILCs are further classified as ILC1, ILC2, and ILC3.

ILC1 cells express T-bet, similar to NK cells (Chapter 12), and produce IFN-γ but lack cytolytic activity. They are found mainly within tissues and are barely detectable in peripheral blood. Oncogene-induced murine cancer model studies suggest that ILC1 cells may play a role in cancer immunosurveillance.

ILC2 development is dependent on expression of the transcription factor GATA-3 and produces the cytokines IL-5 and IL-13. ILC2 cells were first identified in mice as a source of T-helper (Th2) cytokines (IL-4, IL-5, IL-13). ILC2 cells may play a role in anti-helminthic immunity, immune surveillance, immune regulation, and wound healing. ILC2 cells have been shown to accumulate in the skin of patients with AD and within nasal polyps of patients with chronic rhinosinusitis. ILC2 cells produce IL-4, IL-5, and IL-13 in response to epithelial-derived IL-33, IL-25, and thymic stromal lymphopoietin (TSLP). The production of Th2 cytokines by ILC2 cells may represent an early step in the development of atopic disorders.

ILC3 cells express the retinoic acid receptor–related orphan receptor γt (RORγt) and produce IL-17 and IL-22 (Fig. 3.3). ILC3 cells include fetal lymphoid tissue inducer (LTi) cells, which drive secondary lymphoid organ development during embryogenesis. LTi cells can induce upregulation of adhesion molecules on stromal cells and the release of chemokines involved in the recruitment of T and B lymphocytes and DCs to lymph nodes, leading to the subsequent differentiation of naïve T cells into effector T cells and B-cell activation and the production of antibody-secreting cells. Postnatal ILC3 cells influence tissue homeostasis and host defense against extracellular organisms. ILC3-produced IL-22 induces expression of APPs by

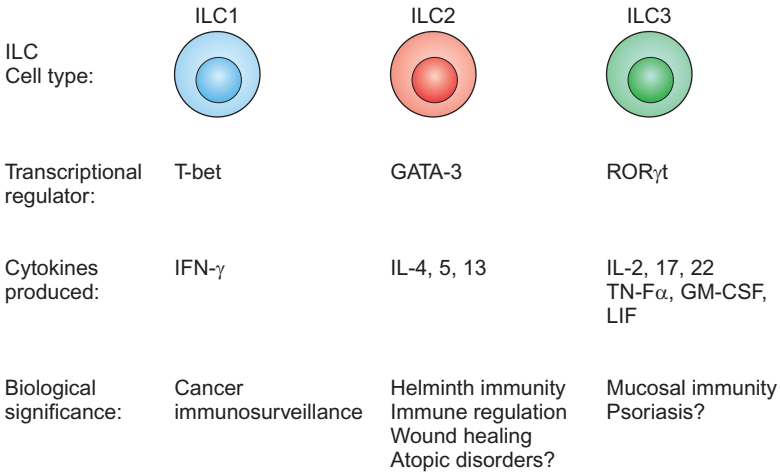


FIG. 3.3 Developmental Regulation and Functions of Innate Lymphoid Cells. The development of innate lymphoid cells (ILCs) is regulated by master transcriptional regulators, including T-bet, GATA-3, and RORγt. Three subsets are recognized: ILC1, ILC2, and ILC3. ILC play roles in tumor immunosurveillance, immune regulation, wound healing, mucosal immunity, atopic disorders, psoriasis, and mucosal immunity.

intestinal epithelial cells. Production of IL-17 by ILC3 cells likely contributes to host defense against *Candida*. ILC3 cells can also produce IL-2, GM-CSF, TNF-α, and leukemia inhibitory factor (LIF). Production of GM-CSF by splenic ILC3 cells is believed to promote survival and activation of splenic neutrophils. A significant population of ILC3 cells have been found in lesional skin of patients with psoriasis.

Mast Cells

Mast cells (Chapter 44) are components of innate immunity that are also commonly found at the interface between host and environment. They are derived from progenitors in bone marrow and circulate as immature precursors to the periphery. Mast cells take up residence and mature in skin, airways, and the GI tract. They are positioned to be first responders to environmental stimuli, including infectious agents. Stem cell factor (SCF, also known as c-kit ligand) is their main survival and developmental factor.

Mast cells express TLR-1 through TLR-9 and therefore are capable of responding to a wide variety of pathogens. TLR-induced mast-cell activation leads to production of proinflammatory cytokines and chemokines. Murine models of peritonitis, such as cecal ligation and puncture, demonstrate that mast cells enhance resistance to bacterial infection. Mast cells are also well known for mediating allergic reactions through IgE-bound allergens that are anchored by FcεRI on the mast-cell surface. Ligation of FcεRI leads to release of tryptase, histamine, leukotrienes, prostaglandins, and cytokines, which cause type 1 hypersensitivity reactions (Chapter 46).

ACTIVATING INNATE IMMUNITY

The innate immune response is initiated when cells of the innate immune system encounter pathogens and recognize them by means of PRRs binding to microbial molecules (e.g., lipopolysaccharide, DNA, RNA). These interactions activate signaling pathways that lead to the production of secreted factors involved in the inflammatory response, including chemokines (Chapter 14) and cytokines (Chapter 15). Characteristics of cytokines include pleiotropism (e.g., the ability to activate a variety of responses on multiple cell types) and redundancy. Cytokines can function locally and distantly and can affect the production of other cytokines. Exposure to cytokines can induce changes in gene expression that affect cell function (e.g., enhanced microbicidal activity or proliferation). Secretion of cytokines (IL-1β, IL-6, TNF-α) is a transient event, thereby limiting potential destruction of host tissue. However, severe infections (e.g., bacteremia, sepsis) can result in overproduction of TNF-α, IL-1β, IL-6, and IFN-γ, which leads to vascular collapse, disseminated intravascular coagulation, and metabolic disturbances (septic shock) that are often fatal.

Cytokine synthesis is a transient process because the messenger RNA (mRNA) of most cytokines is unstable, thus limiting cytokine production. Production of certain cytokines is also regulated by a posttranslational process. For example, TNF-α is a membrane-bound protein that is proteolytically cleaved by a membrane-associated metalloproteinase. IL-1β is a 33-kDa protein that is proteolytically processed by the IL-1β-converting enzyme caspase-1 to generate the biologically active 17-kDa mature IL-1β (described below).

TNF-α and IL-1β can recruit PMNs and monocytes to sites of infection and enhance their ability to eliminate microbes. TNF-α and IL-1β induce expression of adhesion molecules (Chapter 16), such as selectins (P-selectin, E-selectin) and the integrin ligands, ICAMs (intercellular adhesion molecules) and VCAMs (vascular cell adhesion molecules), on vascular endothelial cells near the sites of infection. Expression of selectins on vascular endothelium induces leukocyte rolling on endothelium. Chemokines, such as CXCL8, activate PMNs and monocyte integrins and increase their affinity for ligands (ICAMs, VCAMs) on vascular endothelium, allowing migration of PMNs and monocytes through endothelium to sites of infection. TNF-α and IL-1β both induce prostaglandin synthesis in the hypothalamus, which induces fever.

KEY CONCEPTS

Pattern Recognition Receptors

- Toll-like receptors (TLRs) contain an extracellular domain with leucine-rich repeats (LRRs) and a cytoplasmic Toll/IL-1 receptor domain.
  - The extracellular domain binds ligand.
  - The cytoplasmic domain links to adapter proteins and signaling pathways.
- Nucleotide oligomerization domain (NOD)-like receptors are a family of 22 proteins.
  - LRRs are used for ligand binding.
  - The NOD is used for oligomerization.
  - A caspase activation and recruitment domain (CARD), a Pyrin domain, or a baculovirus inhibitor of apoptosis (BIR) domain is used for initiation of signaling.
- Retinoic acid-inducible gene (RIG)-like receptors contain two N-terminal CARDs for signaling and an RNA helicase domain.
- C-type lectin receptors (CLRs) contain a C (Ca^{2+})-type recognition domain.
  - They mediate diverse functions that depend upon the signaling pathways activated.
- Scavenger receptors and receptor for advanced glycated end products (RAGE) are diverse groups of receptors.
  - They recognize a variety of ligands.
  - They mediate the uptake of oxidized lipoproteins or glycated proteins.
  - They may be involved in atherosclerotic plaque formation.

Pattern Recognition Receptors

Our understanding of the mechanisms by which pathogens are detected has increased greatly over the past decade. Pathogen recognition by the innate immune system leads to engulfment and destruction of invading pathogens, but clearance is often incomplete. The subsequent adaptive immune response is usually required to complete clearance.

The innate immune system expresses a wide variety of PRRs that mediate pathogen recognition. These include TLRs, nucleotide oligomerization domain (NOD)-like receptors (NLRs), and the retinoic acid-inducible gene-I (RIG-I)-like receptors (RLRs). These receptors play an essential role in initiating the innate immune response. Unlike T-cell and B-cell antigen receptors, PRRs are germline encoded, do not undergo somatic recombination, and are expressed constitutively by immune and nonimmune cells. PRRs recognize PAMPs, components of pathogens that are invariant and required for pathogen survival (Table 3.3). Although PRRs detect the PAMPs expressed by microbes, they may also recognize self-molecules (e.g., host nucleic acids), which may underlie some autoimmune diseases such as systemic lupus erythematosus (Chapter 52) and rheumatoid arthritis (Chapter 53).

TABLE 3.3 Classes of Pattern Recognition Receptors

Pattern Recognition Receptor Family	Receptor	Ligand
Toll-like receptors (TLRs)	TLR1/2	Triacyl lipopeptides
	TLR2	Zymosan
	TLR3	dsRNA
	TLR4	LPS, RSV glycoprotein, HSPs, pneumolysin
	TLR2/6	Diacyl lipopeptide
	TLR7	ssRNA
	TLR8	ssRNA
	TLR9	dsDNA, hemozoin
	TLR10	HIV gp41
	TLR11	Profilin-like protein
NOD-like receptors (NLRs)	NOD1	DAP, MDP
	NOD2	MDP
	CIITA	?
	NAIP	*Legionella pneumophilia*, flagellin?
	IPAF	PAMPs
	NLRP1	PAMPs, MDP, microbial toxins
	NLRP2	TBK1 (negative regulator)
	NLRP3	PAMPs, toxins, DAMPs
	NLRP4–14	TBK1 (negative regulator)?
RIG-like receptors (RLRs)	RIG-I	dsRNA, ssRNA
	MDA5	dsRNA, ssRNA
C-type lectin receptors (CLRs)	Mannose receptor	Bacterial carbohydrates
	Dectin-1	Fungal wall glucans
Scavenger receptors	SRA, SRB	Oxidized lipoproteins, apoptotic cells
	CD36, CD68,	β-Amyloid
RAGE		LPS, microbial DNA, β-amyloid

CIITA, Class II major histocompatibility complex transactivator; *DAMP*, damage-associated molecular pattern; *DAP*, meso-diaminopimelic acid; *dsRNA*, double-stranded RNA; *IPAF*, IL-1beta converting enzyme protease activating factor; *LPS*, lipopolysaccharide; *MDA-5*, melanoma differentiation-associated gene-5; *MDP*, muramyl dipeptide; *NAIP*, neuronal apoptosis inhibitory protein; *NLRP*, NOD-like receptor related protein; *PAMP*, pathogen-associated molecular pattern; *RAGE*, receptor for advanced glycation end products; *RSV*, respiratory syncytial virus; *SR*, scavenger receptor; *ssRNA*, single-stranded RNA.

CLINICAL RELEVANCE

Innate Immunity: TLRs, CARDs, NLRs, and Inflammasomes

- TLRs play a non-redundant role in host defense.
  - Defective TLR function (e.g., IRAK-4, MyD88 deficiency) creates susceptibility to invasive, pyogenic infections.
  - Infection in newborns and infants is increased with TLR defects, suggesting the role of TLR is particularly important early in life.
- NOD2 missense mutations have been associated with Crohn disease and Blau syndrome.
- NLRP3 mutations are associated with Muckle-Wells syndrome, familial cold autoinflammatory syndrome, and neonatal onset multisystem inflammatory disease (NOMID).
- CARD9 mutations create susceptibility to chronic mucocutaneous candidiasis.

Toll-Like Receptors

Toll was initially identified in *Drosophila melanogaster* as a receptor required for dorsal–ventral patterning. Subsequently, the Toll pathway was found to be an essential component of *Drosophila* host defense against fungal infection, which led to cloning of mammalian homologues, the TLRs. Mammalian TLRs consist of 11 members that can recognize a wide variety of PAMPs. TLRs are type 1 integral membrane glycoproteins characterized by an extracellular domain with varying numbers of leucine-rich repeats (LRRs) and a cytoplasmic signaling domain homologous to the IL-1 receptor (IL-1R), referred to as the Toll/IL-1R (TIR) homology domain. The TIR domain links the receptor to adaptor proteins (e.g., myeloid differentiation factor 88 [MyD88]) and downstream signaling molecules. This leads to transcription of genes that regulate inflammation (Fig. 3.4).

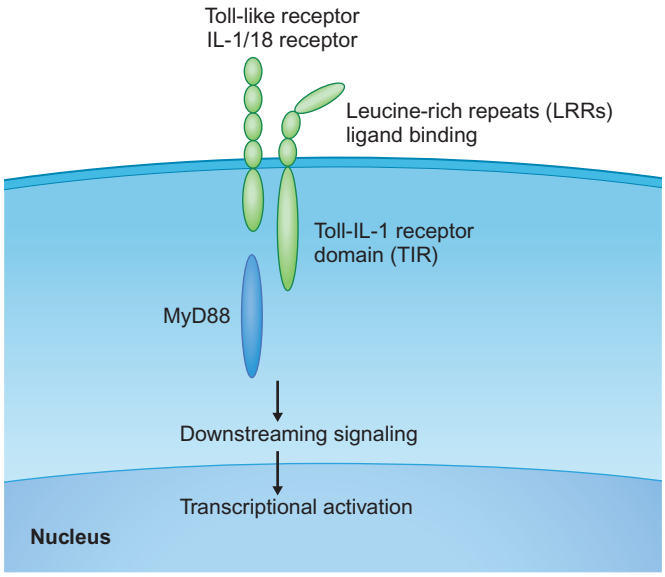


FIG. 3.4 Toll-Like Receptors *(TLRs)* and Interleukin *(IL)*-1/-18 Receptors Share a Common Signaling Pathway. Upon ligand binding, signals are transduced intracellularly by the interaction of the adaptor protein MyD88 with the TIR domain of receptor. MyD88 interacts with IL-1 receptor associated kinase (IRAK)-4 through death domain interactions, activating a signaling cascade, which results in transcriptional activation of genes involved in inflammation.

TLRs are widely expressed on or within cells of the immune system and the epithelia. TLRs detect a wide variety of pathogens (see Table 3.3). They are classified into subfamilies based on their genetic tree. The TLR1, TLR2, and TLR6 subfamily recognizes bacterial lipoproteins, whereas the TLR3, TLR7, TLR8, and TLR9 subfamily recognizes nucleic acids. TLR4, in conjunction with MD-2, recognizes lipopolysaccharide (LPS), TLR5 binds bacterial flagellin, and TLR11, which is functional in mice but probably not in humans. TLR5 recognizes a profilin-like molecule of *Toxoplasma gondii*. However, ligand binding by TLRs can be promiscuous. For example, TLR4 can also bind respiratory syncytial virus F protein and pneumolysin of *S. pneumoniae*; and TLR9 binds malarial hemozoin and hypomethylated CpG-rich DNA.[23] TLRs can also recognize endogenous danger signals: for example, damage-associated molecular patterns (DAMPs) that include heat shock proteins. Recently, TLR10 was found to bind HIV gp41 and high expression of TLR10 was associated with enhanced HIV infection.[24]

The cellular localization of TLRs varies. TLR1, -2, -4, -5, -6, -10, and -11 are found on cell surfaces, whereas TLR3, -7, -8, and -9 are located within endosomes. The cell surface expression of

TLRs, such as TLR4, which recognizes LPS, allows recognition of extracellular molecules released from pathogens. Endosomal expression of TLR3, -7, -8, and -9 allows recognition of microbial nucleic acids following their uptake and degradation in phagolysosomes. Endosomal expression of TLR3, -7, -8, and -9 may prevent activation by host nucleic acids and the development of autoimmunity. The broad cellular expression of TLRs and their diverse and promiscuous agonist recognition allows detection of a wide variety of pathogens despite the existence of a limited number of TLRs.

TLR-mediated cellular responses are essential to host defense. TLR activation stimulates a brief burst of macropinocytosis, which results in antigen uptake at sites of infection and allows antigen presentation to T cells (Chapter 6). TLR activation leads to production of proinflammatory cytokines (e.g., TNF-α, IL-6) and chemokines (e.g., CXCL8). TLR pathway engagement induces transcription and translation of mRNA encoding pro-IL-1β, but production of mature IL-1β requires activation of the inflammasome (described below).

Production of proinflammatory cytokines recruits phagocytes to sites of infection and augments their antimicrobial functions. Production of IL-12p70 by TLR-activated DCs leads to activation of naïve T cells and their subsequent differentiation into effector Th1 cells. Presentation of foreign peptides and increased expression of MHC molecules along with expression of costimulatory molecules (e.g., B7–1, B7–2, IL-12p70) results in subsequent development of adaptive immune responses. IL-12p70 stimulates IFN-γ production by T cells, which further augments the microbicidal activities of phagocytes. Stimulation of TLR3, -7, -8, and -9 elicits the production of proinflammatory cytokines, as well as type 1 IFNs, which play a crucial role in innate antiviral immunity and also influence adaptive immune responses.

Engagement of TLRs activates complex signaling pathways that have been characterized through biochemical analyses and in gene-targeted mice.[25–27] TLRs, IL-1R, and IL-18R share similar signaling pathways (Fig. 3.5). Upon ligand binding, the cytoplasmic adaptor protein MyD88 is recruited to the TIR domain of the receptor for all TLRs, except TLR3.

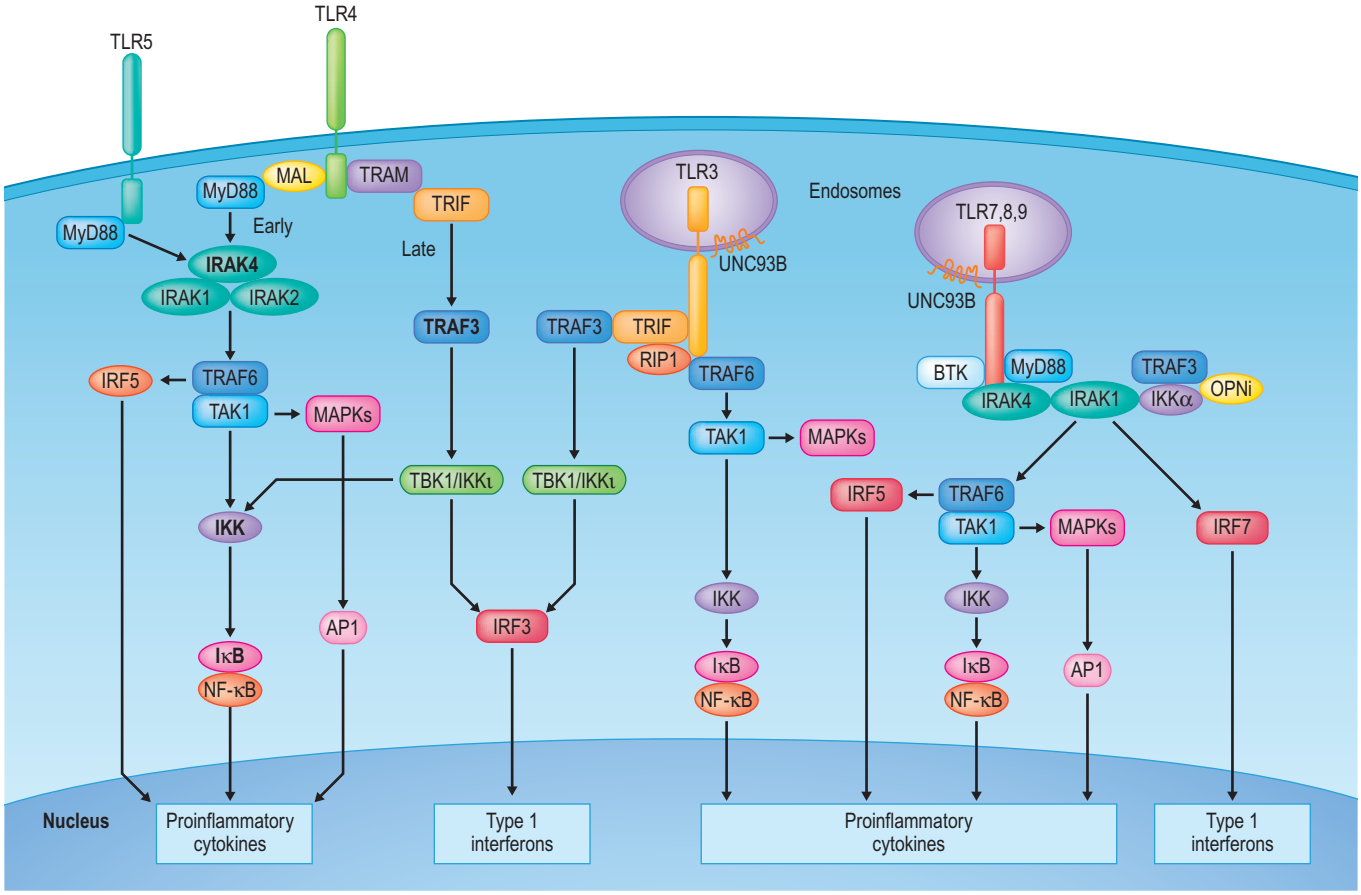


FIG. 3.5 MyD88-Dependent and -Independent Toll-Like Receptor (TLR) Signaling Pathways. Engagement of MyD88-dependent TLR (TLR5) results in activation of IL-1 receptor associated kinase (IRAK)-4 and IRAK-1, IRAK-2, leading to activation of TRAF6 and TAK-1. Subsequently, activation of the IκB kinase (IKK) complex and MAP kinases activates NF-κB and AP1 transcription factors, respectively. The transcription factor IRF5 is also activated downstream of TNF receptor–associated factor 6 (TRAF6). The TLR4 signaling pathway utilizes four adaptor proteins. The adaptors MAL and MyD88 are activated upon ligand interaction at the cell surface, leading to activation of an "early" signaling cascade through the IRAKs. Subsequently, TLR4 is internalized and a "late" signaling cascade, which is dependent on the adaptors TRAM and TRIF, is activated. TLR3 activates a TRIF-dependent pathway that activates RIP1 and TBK1/IKKι, leading to production of proinflammatory cytokines and IFN-β. TLR7, -8, and -9 are MyD88 dependent and activate the transcription factors NF-κB, IRF5, AP1, and IRF7, resulting in production of proinflammatory cytokines and type 1 interferons.

Recruitment of MyD88 leads to recruitment of IL-1 receptor associated kinase-4 (IRAK-4), through death domain interactions. IRAK-4 activation leads to recruitment and activation of IRAK-1 and IRAK-2. In monocytes, IRAK-M is also recruited to this complex and functions as a negative regulator of signaling. Both IRAK-1 and IRAK-2 activation are required for full activation of NF-κB and MAP kinases. IRAK activation leads to interaction with TNF receptor–associated factor 6 (TRAF6), which is an E3 ubiquitin ligase. Along with the E2 conjugating complex of Ubc13 and Uev1a, TRAF6 is K-63 ubiquitinated, recruiting TGF-β–activated protein kinase-1 (TAK-1). TAK-1 then activates the inhibitor of NF-κB (IκB) kinase complex (IKK), which consists of NF-κB essential modifier (NEMO), IKKα, and IKKβ, leading to phosphorylation of IκB (inhibitor of NF-κB) proteins and their subsequent K-48 linked ubiquitination and degradation. NF-κB is subsequently released from inhibition, allowing translocation to the nucleus, where it mediates transcriptional activation of numerous genes involved in inflammation.

The transcription factor interferon regulatory factor-5 (IRF-5) is also activated downstream of TRAF6 and is required for production of proinflammatory cytokines. TAK-1 activation also leads to activation of p38 MAP kinase and c-Jun N terminal kinase (JNK), which then activates the AP1 transcriptional complex (see Fig. 3.5).

TLR signaling can proceed via multiple pathways, impacting both the kinetics and nature of the subsequent innate response. For example, TLR4 also interacts with the adaptors MAL (MyD88-like adaptor protein), TRAM (translocating chain-associating membrane protein), and TRIF (TIR domain–containing adaptor-inducing IFN-β (see Fig. 3.5). TLR4 initially recruits MAL and MyD88 to trigger "early phase" NF-κB and MAP kinase activation. TLR4 is subsequently endocytosed and trafficked to the endosome, where it forms a signaling complex with TRAM and TRIF, which leads to activation of TANK-binding kinase-1 (TBK-1), IKKε, and IRF-3, and "late phase" activation of NF-κB and MAP kinases. Activation of IRF3 induces IFN-β production.

Antiviral TLRs are located in endosomes and interact with an endoplasmic reticulum membrane protein called UNC93B (see Fig. 3.5). Upon activation, TLR3 does not recruit MyD88, but rather TRIF, leading to recruitment of TRAF3 and activation of TBK1, IKKι, IRF-3, and IFN-β production. TRIF also recruits RIP1 and TRAF6, which leads to activation of NF-κB. While the other antiviral TLRs, TLR7, -8, and TLR9, are MyD88 dependent, they activate a pathway utilizing IRAK-1, IKKα, TRAF3, and intracellular osteopontin (iOPN), activating IRF7 and leading to production of IFN-α.[28] TLR7, -8, and -9 also utilize a TRAF6-dependent pathway that leads to NF-κB/MAP kinase activation. Bruton tyrosine kinase (BTK) (Chapter 33) also plays a critical role in TLR8- and TLR9-induced production of TNF-α and IL-6.

Given the ubiquitous expression of TLRs and their essential roles in detection of pathogens, it is not surprising that genetic defects that impair TLR function (e.g., IRAK4 deficiency, MyD88 deficiency) are associated with susceptibility to infections, notably invasive, pyogenic infections (e.g., *S. aureus, S. pneumococcus, P. aeruginosa*) and herpes simplex I (defects in TLR3 function) (Chapter 35).[29] Additionally, numerous single nucleotide polymorphisms (SNPs) in individual TLRs have been identified that have been associated with increased susceptibility to a variety of pathogens or increased susceptibility to autoimmune diseases (e.g., rheumatoid arthritis, systemic lupus erythematosus). However, in many instances the effects of these SNPs on TLR function is unclear and their importance in human health requires further investigation.[30]

The innate immune system in natura encounters intact pathogens that express multiple PAMPs, including bacterial cell wall components as well as microbial DNA and RNA. Thus, DCs and other phagocytes are activated through multiple PRRs. Activation of DCs through combinations of TLRs, such as TLR4 and TLR8, can induce synergistic production of Th1 cell–inducing cytokines, as well as the Th1-inducing ligand, Delta-4, leading to stronger Th1 differentiation of T cells than occurs following activation of a single TLR.[31] Interestingly, the use of combinations of TLR agonists on virus-sized nanoparticles containing antigen induce enhanced and better-sustained antibody responses in mice and nonhuman primates.[32] Thus the use of combinations of TLR agonists as adjuvants in vaccines may result in enhanced efficacy of future vaccines (Chapter 87), as well as in immunotherapy against tumors (Chapter 80).

During an infection, multiple factors can mitigate TLR-induced inflammation. One such factor is adenosine, an endogenous purine metabolite whose levels rise during stress or hypoxia. Adenosine binds receptors expressed on leukocytes, leading to increased intracellular concentrations of cyclic adenosine monophosphate (cAMP), dampening TLR-mediated production of Th1-polarizing cytokines while preserving production of Th2 and anti-inflammatory cytokines. Anti-inflammatory/pro-resolving lipid metabolites, such as resolvins and lipoxins, can differentially regulate TLR4-mediated responses of macrophages, inhibiting TNF response to pure LPS but enhancing uptake, killing, and TNF-α production to whole gram-negative bacteria.[33]

Within the GI tract, the factors that maintain tolerance to commensal host flora while detecting/containing pathogenic bacteria with appropriate inflammatory responses are incompletely understood (Chapter 22). The detection of common PAMPs in pathogenic and nonpathogenic bacteria would be anticipated to activate the same inflammatory response. Nevertheless, the detection of commensal bacteria within the intestines can induce tolerance. TLR signaling can contribute to intestinal homeostasis by regulating intestinal epithelial cell proliferation and epithelial integrity. Expression and localization of TLRs in the intestinal epithelium may directly relate to their role in maintaining homeostasis versus inducing inflammation. For example, within the intestinal epithelium, TLR9 activation through the apical membrane induces tolerance, whereas TLR9 activation via the basolateral membrane induces an inflammatory response through the canonical NF-κB pathway. Differential spatial expression of PRR in epithelia may constitute a critical mechanism of distinguishing nonpathological from pathological bacteria that have breached intestinal epithelium.

NOD-Like Receptors

NOD-like receptors (NLRs) are cytoplasmic PRRs that are unrelated to the transmembrane PRRs. NLRs mediate detection of intracytoplasmic bacterial products. Among the NLRs are five members of the NOD family, 14 members of the NALP family, CIITA, IPAF, and NAIP (see Table 3.3). NLR family proteins possess LRRs for ligand detection; a NOD domain (also referred to as a NACHT domain); a domain for initiation of signaling, such as caspase activation and recruitment domain (CARD); pyrin domains; or baculovirus inhibitor of

apoptosis repeat (BIR) domains. NOD1 and NOD2 were the first NLRs identified. They detect components of bacterial peptidoglycan: NOD1 detects mesodiaminopimelic acid, and NOD2 detects muramyl dipeptide. Direct ligand binding by NLR has not yet been demonstrated, leaving open the possibility that detection of pathogens and other signals by NLRs may be indirect. Following activation, NODs oligomerize and recruit the protein kinase RIP2 and CARD9 via CARD domains, leading to activation of NF-κB and MAP kinases, respectively (Fig. 3.6). NOD2 may also play a role in activation of some inflammasomes (described below).

In humans, missense mutations in NOD2 that impair function have been associated with susceptibility to Crohn disease. Conversely, missense mutations in NOD2 that lead to constitutive activation of NF-κB lead to Blau syndrome, an autosomal dominant disorder characterized by granulomatous arthritis, iritis, and skin granulomas.[34] Although many NLRs function as activators of inflammation, NLRP2 and NLRP4 were found to be negative regulators of innate antiviral inflammation through blockade of TBK1 activation, leading to inhibition of the transcription factor IRF3 and inhibition of type 1 interferon production.[35] Thus, some NLRs play+ homeostatic roles in inflammatory responses.

RIG-I–Like Receptors

RIG-I–like receptors (RLRs) detect the presence of viral nucleic acids generated by intracellular, replicating viruses. The RLRs consist of two receptors: RIG-I and melanoma differentiation-associated gene-5 (MDA-5). Both have two N-terminal CARDs and an RNA helicase domain and mediate virus-induced type 1 IFN expression in fibroblasts and conventional DCs. A third RLR, laboratory of genetics and physiology 2 (LGP2), lacks the N-terminal CARD domains and plays a role in repression of signaling. RIG-I and MDA-5 are activated by double-stranded RNA (dsRNA) generated during viral replication with distinct specificities for viral recognition. RIG-I detects Orthomyxoviridae, Rhabdoviridae, Paramyxoviridae, and Flaviviridae, whereas MDA-5 detects Picornaviridae, Caliciviridae, and Coronaviridae. Polyinosine:cytosine (poly I:C) is a nonspecific dsRNA analogue used experimentally to activate TLR3 and RIG-I/MDA-5. Relatively short poly I:C (<1 kb) is recognized preferentially by RIG-I, whereas long poly I:C (>1 kb) is preferentially recognized by MDA-5.[23]

dsRNA-induced activation of RIG-I and MDA-5 induces their association with a mitochondria-associated adaptor known as interferon-β promoter stimulator-1 (IPS-1) or mitochondrial antiviral signaling protein (MAVS) through CARD domain interactions. Downstream effectors include TBK-1 and IKKι, which activate IRF3 and IRF7, leading to production of type-1 interferons (see Fig. 3.6). It is worth noting that live bacteria are more effective inducers of STAT-1, type I IFN, and inflammasome pathways compared with killed organisms, a property that may reflect the importance of bacterial RNA to innate recognition of live infection.[36] Some viruses have developed strategies to circumvent innate immune responses. For example, the coronavirus that causes severe acute respiratory syndrome (SARS-CoV) expresses a protein encoded by open reading frame-9b (ORF-9b), which targets the MAVS signalosome. ORF-9b causes the degradation of MAVS, TRAF3, and TRAF6, inhibiting host cell interferon production.[37]

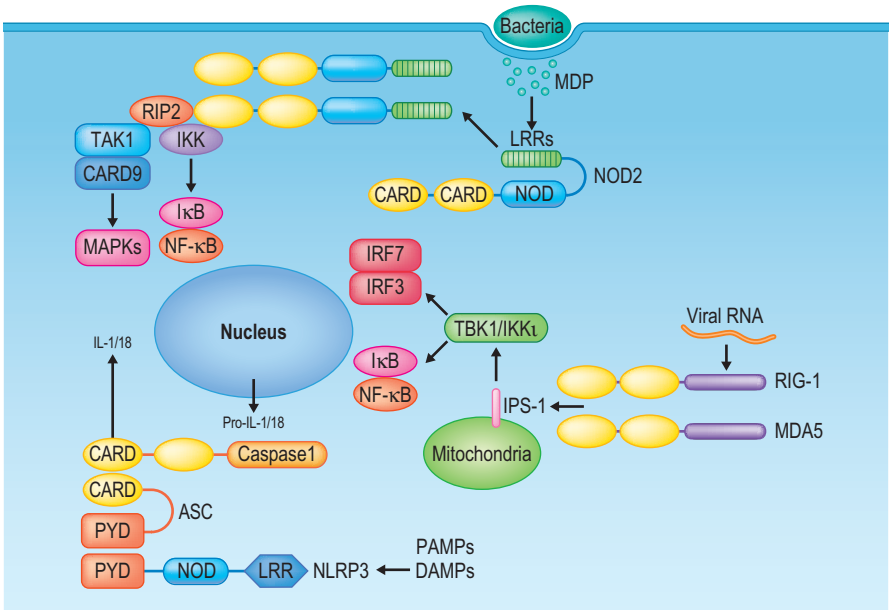


FIG. 3.6 NOD-Like Receptor (NLR), RIG-I–Like Receptor (RLR), and Inflammasome Signaling. NOD2 is activated following exposure to bacterial muramyl dipeptide (MDP) and leads to dimerization and activation of RIP2, TAK1, CARD9, and the IκB kinase (IKK) complex. This results in inflammatory gene transcription. RLRs, like RIG-I and MDA-5, are activated by double-stranded RNA generated by intracellular, replicating viruses, inducing activation of transcription factors, including NF-κB, IRF3, and IRF7, leading to production of proinflammatory cytokines and type 1 interferons. Inflammasomes can be activated by microbial products (PAMPs), as well as endogenous products released by damaged host cells (DAMPs), leading to activation of caspase 1. Active caspase 1 processes pro-IL-1β and pro-IL-18 into mature IL-1β and IL-18. Active caspase 1 also processes GSDMD to form pores in the cell membrane that allows release of IL-1β and IL-18.

C-Type Lectin Receptors

C-type lectin receptors (CLRs) are a diverse group of receptors originally identified as Ca^{2+}-dependent carbohydrate-binding proteins.[38] CLRs are defined as any protein that contains a C-type carbohydrate recognition domain (CRD), regardless of calcium- or carbohydrate-binding ability. CLRs include numerous members with diverse functions, including cell adhesion, regulation of NK-cell function, phagocytosis, endocytosis, platelet activation, complement activation, tissue remodeling, and innate immunity. In myeloid cells, CLRs can mediate internalization of microbes to allow for antigen processing and presentation. Some CLRs function analogously to TLR, resulting in direct cellular activation and generation of inflammatory responses. Other CLRs are capable of binding PAMPs, but function to modulate cell activation. The functions of myeloid CLRs are dictated by the signaling pathways they activate.

Dectin-1 is a CLR expressed on DCs and other myeloid cells that recognizes β-1,3-linked glucans present in the cell wall of fungi, mycobacteria, and plants. Dectin-2 recognizes high mannose structures and α-mannans found in fungi, mycobacteria, and dust mites. Following ligand binding, Dectin-1 and Dectin-2 activate signaling pathways by utilizing the tyrosine kinase Syk, CARD9, and Raf-1, leading to activation of the transcription factors NF-κB, NFAT, and AP1 and production of proinflammatory cytokines (Fig. 3.7). The production of cytokines downstream of Dectin-1 and Dectin-2 (e.g., IL-1, IL-6, TGF-β, IL-23) induces the subsequent differentiation of naïve T cells into Th17 T cells, which play a critical role in antifungal immunity (Chapters 11 and 28). Activation of Syk induces generation of ROIs, which can activate the NLRP3 inflammasome, leading to processing of pro-IL-1β to mature IL-1β. The importance of CLR function in antifungal immunity is demonstrated by inactivating mutations in Dectin-1 and CARD9 that lead to chronic mucocutaneous candidiasis, as well as invasive fungal infections in the case of CARD9 deficiency.[39]

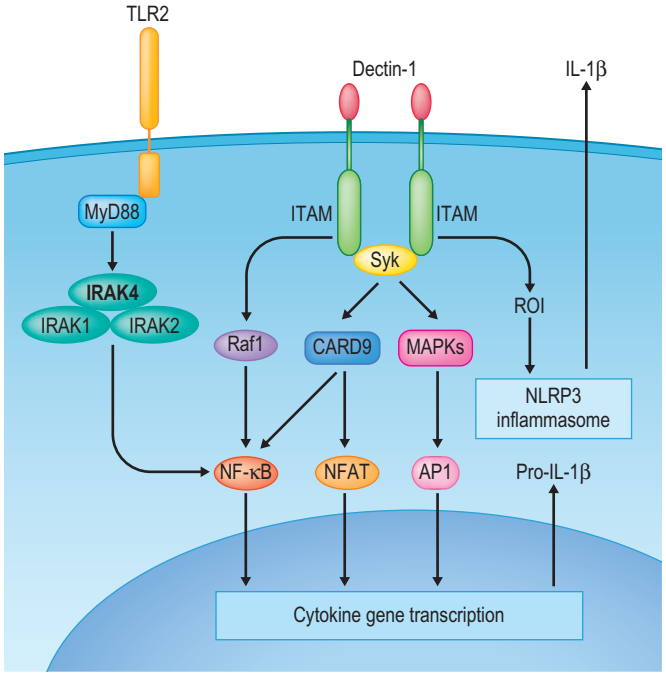


FIG. 3.7 C-Type Lectin Signaling. C-type lectin receptors (CLRs), like Dectin-1, contain immunoreceptor tyrosine-based activation motifs (ITAMs) that interact with the cytosolic tyrosine kinase, Syk, leading to activation of transcription factors including NF-κB, NFAT, and AP1. Dectin-1 activates the serine kinase, Raf1, which contributes to NF-κB activation. Toll-like receptors (TLRs), such as TLR2, can cooperate with Dectin-1 signaling to activate NF-κB, leading to an enhanced inflammatory response.

Mincle (macrophage inducible C-type lectin) recognizes α-mannans and glycolipids and associates with the FcRγ chain. Upon ligand binding, Syk is recruited to the ITAM of FcRγ, leading to cellular activation. Mincle also binds the endogenous nucleoprotein SAP130, which is exposed by dead cells. The Mincle-mediated response to dead cells leads to infiltration of PMNs into damaged tissue and may contribute to tissue repair.

Other CLRs, such as DCIR (DC-inhibitory receptor), possess an inhibitory ITIM motif. DCIR is expressed on myeloid cells, DCs, and B cells. DCIR inhibits TLR8-induced IL-12p70 and TNF-α production by myeloid DCs and TLR9-induced IFN-α/β production by pDCs. Inhibition of TLR responses may reflect inhibition of tyrosine kinases and/or PI3 kinase pathways.

Pathogens express multiple PAMPs that activate a variety of PRRs. Indeed, CLRs and TLRs cooperate in antimicrobial responses. Coordinated activation of Dectin-1 and TLR2, for example, via stimulation with zymosan, enhances production of proinflammatory cytokines via activation of both Dectin-1-Syk and TLR2-MyD88 signaling pathways (see Fig. 3.7). DC-SIGN, which recognizes mycobacteria and viruses, can enhance TLR-induced NF-κB activation through a Raf-1–dependent signaling pathway.

Scavenger Receptors

Scavenger receptors are a diverse group of receptors that include CD36, CD68, SR class A, and SR class B.[40] The receptors mediate the uptake of oxidized lipoproteins into cells. Scavenger receptors also mediate the uptake of microbes and contribute to the response of macrophages to mycobacteria. SR class A can also mediate an inflammatory response to β-amyloid fibrils that may contribute to Alzheimer disease (see Table 3.3). Scavenger receptors play a pathological role in the generation of cholesterol-laden foam cells that comprise atherosclerotic plaques in blood vessels.

Receptor for Advanced Glycation End Products

RAGE is an inflammatory transmembrane receptor that belongs to the immunoglobulin superfamily of receptors (see Table 3.3). RAGE is capable of binding a variety of DAMP molecules, including advanced glycation end products, HMGB1, S100 proteins, and DNA. Additionally, RAGE can bind various PAMPs, including LPS, microbial DNA and respiratory viruses. RAGE preassembles into dimers and multimers at the cell surface. Upon ligand binding, intracellular signaling pathways are activated through the adaptor protein diaphanous-1/mDia-1, which include Rho GTPases, phosphoinositol-3 kinase, mitogen-activated kinases, AKT, and transcription factors (e.g., NF-κB, STAT3), inducing production of proinflammatory cytokines (e.g., TNFβ, IL-6). Chronic activation of RAGE has been implicated in numerous pathological conditions, including cardiovascular disease, diabetic inflammatory complications, Alzheimer disease, and some cancers.[41]

PERSPECTIVES

- While short-lived, polymorphonuclear leukocytes (neutrophils) are the most abundant and earliest cells of the innate immune system to respond to infection.
- Several days later in an infection, monocytes and macrophages predominate.
- Activated neutrophils, monocytes, and macrophages kill phagocytosed bacteria with reactive oxygen intermediates and antimicrobial peptides and proteins (APPs).
- Dendritic cells take up and present foreign antigens and thus link the innate to the adaptive immune system.
- Natural killer (NK) cells can kill infected or malignant cells without prior activation.
- Mast cells are found at the interface between the host and the environment, are first responders to microbes, and recruit other inflammatory cells.

Inflammasomes

A variety of stimuli, including PAMPs, bacterial toxins, the common vaccine adjuvant aluminum hydroxide (alum), and ultraviolet (UV) light, as well as endogenous "danger signals" released by stressed or damaged host cells, referred to as DAMPs (e.g., adenosine triphosphate [ATP], uric acid, hyaluronan), induce the processing of pro-IL-1β to mature IL-1β. The cytosolic cellular multimolecular machinery responsible for IL-1β processing is termed the inflammasome. Prototypical inflammasomes include the NOD-like receptor-related protein-1 (NLRP1) inflammasome, the NLRP3 inflammasome, and the IL-1β-converting enzyme protease-activating factor (IPAF) inflammasome. NLRP recruit ASC (apoptosis-associated speck-like protein containing a CARD) via homotypic PYRIN domain interactions. ASC is an adaptor protein that contains a PYRIN domain and a CARD domain. Caspase-1 is subsequently recruited via CARD domains, leading to proteolytic processing of IL-1β and IL-18 (see Fig. 3.6). Although NLRP1, NLRP3, and IPAF form prototypical inflammasomes, other NLR, including NOD2 and neuronal apoptosis inhibitory protein (NAIP), can also form inflammasomes or modulate their activities.[42]

The mechanism through which mature IL-1β and IL-18 are released has been termed pyroptosis, a form of inflammatory death of the infected cell (Chapter 17). During pyroptosis, inflammatory caspases (caspase 1) cleave gasdermin D (GSDMD), a cytosolic protein in phagocytes and epithelial cells. Cleavage of GSDMD leads to its localization at the cell membrane where it oligomerizes to form membrane pores. Pore formation disrupts membrane integrity, leading to cell death and release of small molecules, such as IL-1β and IL-18 and GSDMD itself (see Fig. 3.6). GSDMD can bind cardiolipin in bacterial cell membranes, perhaps leading to permeabilization of the bacterial membrane and death.[43]

Aberrant inflammasome activation can result in autoinflammatory diseases, characterized by recurrent inflammatory episodes. Mutations in the *NLRP3* gene are associated with a spectrum of autosomal dominant inherited autoinflammatory diseases, including Muckle-Wells syndrome (sensorineural deafness, urticaria, fevers, chills, arthritis), familial cold autoinflammatory syndrome (rash, conjunctivitis, fever, chills, arthralgias elicited by cold exposure), and neonatal-onset multisystem inflammatory disease (NOMID) (rashes, arthritis, chronic meningitis). These disorders are the result of constitutive production of IL-1β. Treatment of these disorders involves anti-inflammatory therapy, including IL-1 antagonists.[44]

INNATE IMMUNITY IN CLINICAL PRACTICE

Our growing understanding of the molecular and cellular basis for innate immunity is paralleled by increasing appreciation of its importance to clinical medicine. Examples include (1) recognition of novel primary immunodeficiencies in the innate immune pathways presenting with heightened susceptibility to bacterial (e.g., MyD88 deficiency, IRAK4 deficiency) and fungal infection (e.g., CARD9 deficiency); (2) appreciation of immune ontogeny (i.e., the change in immune function in a given individual across their life span, which parallels age-specific susceptibility to infection); (3) clinical studies of the potential use of APP congeners to prevent or treat infection; (4) leveraging PRR agonists, such as TLR agonists, as vaccine adjuvants (e.g., the TLR4 agonist monophosphoryl lipid A [MPLA], which is a detoxified form of lipopolysaccharide or endotoxin used in a licensed human papillomavirus [HPV] vaccine [Cervarix]); and (5) employing certain live attenuated vaccines, such as bacille Calmette-Guérin (BCG) or measles, mumps, rubella (MMR) vaccine that activate multiple PRRs to reduce infections to antigenically distinct pathogens—so-called beneficial heterologous or "nonspecific" effects—that may be mediated via epigenetic reprograming of monocytes and enhancement of innate immune responses.[45]

Our understanding of the complexity of innate immunity has advanced considerably over the past decade. The importance of the innate immune system to human health is underscored by single gene mutations, such as IRAK4 deficiency, that result in immune deficiencies and infection, particularly in early life.[29] Continued elucidation of the cell types comprising the innate immune system has expanded our understanding of the delicate balance between tolerance to nonpathogenic commensal bacteria and inflammatory responses to pathogenic, invasive microbes. As innate immunity is expressed in an age-specific manner, a better understanding of the ontogeny of the innate immune system, as well as the mechanisms by which innate and adaptive immunity interact, will guide development of adjuvants, resulting in more effective vaccines and tumor immunotherapy.

REFERENCES

1. Netea MG, Joosten LA, Latz E, et al. Trained immunity: a program of innate immune memory in health and disease. *Science*. 2016;352(6284):aaf1098.
2. Brandner JM, Zorn-Kruppa M, Yoshida T, Moll I, Beck LA, De Benedetto A. Epidermal tight junctions in health and disease. *Tissue Barriers*. 2015;3(1–2):e974451.
3. Hazlett L, Wu M. Defensins in innate immunity. *Cell Tissue Res*. 2011;343(1):175–188.
4. Martin L, van Meegern A, Doemming S, Schuerholz T. Antimicrobial peptides in human sepsis. *Front Immunol*. 2015;6:404.
5. Schutte BC, McCray Jr. PB. [beta]-defensins in lung host defense. *Annu Rev Physiol*. 2002;64:709–748.
6. Hiemstra PS, Amatngalim GD, van der Does AM, Taube C. Antimicrobial peptides and innate lung defenses: role in infectious and noninfectious lung diseases and therapeutic applications. *Chest*. 2016;149(2):545–551.
7. Palmer CD, Guinan EC, Levy O. Deficient expression of bactericidal/permeability-increasing protein in immunocompromised hosts: translational potential of replacement therapy. *Biochem Soc Trans*. 2011;39(4):994–999.
8. Sharon N. Carbohydrates as recognition determinants in phagocytosis and in lectin-mediated killing of target cells. *Biol Cell*. 1984;51(2):239–245.
9. Ricklin D, Hajishengallis G, Yang K, Lambris JD. Complement: a key system for immune surveillance and homeostasis. *Nat Immunol*. 2010;11(9):785–797.

10. Rickert RC. Regulation of B lymphocyte activation by complement C3 and the B cell coreceptor complex. *Curr Opin Immunol*. 2005;17(3):237–243.
11. Kaplan AP. Enzymatic pathways in the pathogenesis of hereditary angioedema: the role of C1 inhibitor therapy. *J Allergy Clin Immunol*. 2010;126(5):918–925.
12. Daha MR. Role of complement in innate immunity and infections. *Crit Rev Immunol*. 2010;30(1):47–52.
13. Cros J, Cagnard N, Woollard K, et al. Human CD14dim monocytes patrol and sense nucleic acids and viruses via TLR7 and TLR8 receptors. *Immunity*. 2010;33(3):375–386.
14. Shapouri-Moghaddam A, Mohammadian S, Vazini H, et al. Macrophage plasticity, polarization, and function in health and disease. *J Cell Physiol*. 2018;233(9):6425–6440.
15. Liu L, Wei Q, Lin Q, et al. Anti-spike IgG causes severe acute lung injury by skewing macrophage responses during acute SARS-CoV infection. *JCI Insight*. 2019;4(4).
16. Reeves EP, Lu H, Jacobs HL, et al. Killing activity of neutrophils is mediated through activation of proteases by K^+ flux. *Nature*. 2002;416(6878):291–297.
17. Papayannopoulos V. Neutrophil extracellular traps in immunity and disease. *Nat Rev Immunol*. 2018;18(2):134–147.
18. Kobayashi Y. The regulatory role of nitric oxide in proinflammatory cytokine expression during the induction and resolution of inflammation. *J Leukoc Biol*. 2010;88(6):1157–1162.
19. Berzins SP, Smyth MJ, Baxter AG. Presumed guilty: natural killer T cell defects and human disease. *Nat Rev Immunol*. 2011;11(2):131–142.
20. Umetsu DT, Dekruyff RH. Natural killer T cells are important in the pathogenesis of asthma: the many pathways to asthma. *J Allergy Clin Immunol*. 2010;125(5):975–979.
21. Villablanca EJ, Mora JR. A two-step model for Langerhans cell migration to skin-draining LN. *Eur J Immunol*. 2008;38(11):2975–2980.
22. Qiu Y, Wang W, Xiao W, Yang H. Role of the intestinal cytokine microenvironment in shaping the intraepithelial lymphocyte repertoire. *J Leukoc Biol*. 2015;97(5):849–857.
23. Kawai T, Akira S. The role of pattern-recognition receptors in innate immunity: update on Toll-like receptors. *Nat Immunol*. 2010;11(5):373–384.
24. Henrick BM, Yao XD, Zahoor MA, Abimiku A, Osawe S, Rosenthal KL. TLR10 senses HIV-1 proteins and significantly enhances HIV-1 infection. *Front Immunol*. 2019;10:482.
25. Beutler B, Jiang Z, Georgel P, et al. Genetic analysis of host resistance: Toll-like receptor signaling and immunity at large. *Annu Rev Immunol*. 2006;24:353–389.
26. Kawai T, Akira S. Toll-like receptors and their crosstalk with other innate receptors in infection and immunity. *Immunity*. 2011;34(5):637–650.
27. O'Neill LA. The interleukin-1 receptor/Toll-like receptor superfamily: 10 years of progress. *Immunol Rev*. 2008;226:10–18.
28. Shinohara ML, Lu L, Bu J, et al. Osteopontin expression is essential for interferon-alpha production by plasmacytoid dendritic cells. *Nat Immunol*. 2006;7(5):498–506.
29. Picard C, von Bernuth H, Ghandil P, et al. Clinical features and outcome of patients with IRAK-4 and MyD88 deficiency. *Medicine (Baltimore)*. 2010;89(6):403–425.
30. Mukherjee S, Huda S, Sinha Babu SP. Toll-like receptor polymorphism in host immune response to infectious diseases: a review. *Scand J Immunol*. 2019;90(1):e12771.
31. Napolitani G, Rinaldi A, Bertoni F, Sallusto F, Lanzavecchia A. Selected Toll-like receptor agonist combinations synergistically trigger a T helper type 1-polarizing program in dendritic cells. *Nat Immunol*. 2005;6(8):769–776.
32. Kasturi SP, Skountzou I, Albrecht RA, et al. Programming the magnitude and persistence of antibody responses with innate immunity. *Nature*. 2011;470(7335):543–547.
33. Palmer CD, Mancuso CJ, Weiss JP, Serhan CN, Guinan EC, Levy O. 17(R)-Resolvin D1 differentially regulates TLR4-mediated responses of primary human macrophages to purified LPS and live *E. coli*. *J Leukoc Biol*. 2011;90(3):459–470.
34. Philpott DJ, Sorbara MT, Robertson SJ, Croitoru K, Girardin SE. NOD proteins: regulators of inflammation in health and disease. *Nat Rev Immunol*. 2014;14(1):9–23.
35. Yang Y, Lang X, Sun S, et al. NLRP2 negatively regulates antiviral immunity by interacting with TBK1. *Eur J Immunol*. 2018;48(11):1817–1825.
36. Sander LE, Davis MJ, Boekschoten MV, et al. Detection of prokaryotic mRNA signifies microbial viability and promotes immunity. *Nature*. 2011;474(7351):385–389.
37. Shi CS, Qi HY, Boularan C, et al. SARS-coronavirus open reading frame-9b suppresses innate immunity by targeting mitochondria and the MAVS/TRAF3/TRAF6 signalosome. *J Immunol*. 2014;193(6):3080–3089.
38. Osorio F, Reis e Sousa C. Myeloid C-type lectin receptors in pathogen recognition and host defense. *Immunity*. 2011;34(5):651–664.
39. Gross O, Gewies A, Finger K, et al. Card9 controls a non-TLR signalling pathway for innate anti-fungal immunity. *Nature*. 2006;442(7103):651–656.
40. Greaves DR, Gordon S. The macrophage scavenger receptor at 30 years of age: current knowledge and future challenges. *J Lipid Res*. 2009;50(suppl):S282–S286.
41. Hudson BI, Lippman ME. Targeting RAGE signaling in inflammatory disease. *Annu Rev Med*. 2018;69:349–364.
42. Schroder K, Tschopp J. The inflammasomes. *Cell*. 2010;140(6):821–832.
43. Liu X, Lieberman J. A mechanistic understanding of pyroptosis: the fiery death triggered by invasive infection. *Adv Immunol*. 2017;135:81–117.
44. Geyer M, Muller-Ladner U. Actual status of antiinterleukin-1 therapies in rheumatic diseases. *Curr Opin Rheumatol*. 2010;22(3):246–251.
45. Levy O, Netea MG. Innate immune memory: implications for development of pediatric immunomodulatory agents and adjuvanted vaccines. *Pediatr Res*. 2014;75(1–2):184–188.

4

Antigen Receptor Genes, Gene Products, and Co-Receptors

Harry W. Schroeder Jr., Laurent Gapin, and Raul Martin Torres

In 1890, von Behring and Kitasato reported the existence of an agent in the blood that could neutralize diphtheria toxin. The following year, glancing references were made to "*Antikörper*" in studies describing the ability of the agent to discriminate between two immune substances, or bodies. The term *antigen* (Chapter 6) is a shortened form of *Antisomatogen* + *Immunkörperbildner*, the substance that induces the production of an antibody. Thus, the definition of *antibody* and *antigen* represent a classic tautology.

In 1939, Tiselius and Kabat used electrophoresis to separate immunized serum into albumin, α-globulin, β-globulin, and γ-globulin fractions. Absorption of the serum against the antigen depleted the γ-globulin fraction, yielding the terms *gammaglobulin*, *immunoglobulin* (Ig), and *IgG*. Subsequently, "sizing" columns were used to separate Igs into those that were "heavy" (pentameric IgM), "regular" (IgA, IgE, IgD, IgG, or monomeric IgM), and "light" (light-chain dimers), culminating with the discovery of the last major class of Ig, IgE, in 1966.

In 1949, Porter used papain to digest IgG molecules into two types of fragments, termed Fab (fragment antigen binding) and Fc (fragment crystallizable) (Chapter 8). The constancy of the Fc fragment permitted its crystallization, and thus the elucidation of its sequence and structure. The variability of the Fab fragment precluded analysis until Bence Jones myeloma proteins were identified as clonal, isolated light chains.

In 1976, Hozumi and Tonegawa demonstrated that the variable portion of κ chains was the product of the rearrangement of a variable (V) and joining (J) gene segment. In 1982, Alt and Baltimore reported that terminal deoxynucleotidyl transferase (TdT) could be used to introduce non-germline encoded sequence between rearranging V, D for diversity, and J gene segments, potentially freeing the preimmune heavy-chain repertoire from germline constraints. In 1984, Weigert and colleagues determined that during affinity maturation variable domains could undergo mutation at a rate of 10^{-3} per base pair, per generation. These discoveries clarified how lymphocytes could generate an astronomically diverse antigen receptor repertoire from a handful of gene elements.

In 1982 Allison and colleagues raised antisera against a cell-surface molecule that could uniquely identify individual T-cell clones. A year later, Kappler and a consortium of colleagues demonstrated that these surface molecules were heterodimers composed of variable and constant region domains, just like Ig. Subsequently, Davis and Mak independently cloned the β chain of the T-cell antigen receptor (TCR). Initial confusion regarding the identity of the companion α chain led to the realization that there were two mutually exclusive forms of TCR, αβ and γδ.

PARATOPES AND EPITOPES

Igs and TCRs both belong to the eponymous immunoglobulin superfamily (IgSF).[1] The study of antibodies precedes that of TCR by decades; hence much of what we know is based on knowledge first gleaned from the study of Igs.

Ig–antigen interactions typically take place between the *paratope*, the site on the Ig at which the antigen binds, and the *epitope*, which is the site on the antigen that is bound. Thus, lymphocyte antigen receptors do not recognize antigens; they recognize epitopes borne on those antigens. This makes it possible for the cell to discriminate between two closely related antigens, each of which can be viewed as a collection of epitopes. It also permits the same receptor to bind divergent antigens that share equivalent or similar epitopes, a phenomenon referred to as *cross-reactivity*.

Although both Igs and TCRs can recognize the same antigen, they do so in markedly different ways. Igs tend to recognize intact antigens in soluble form, preferentially identifying surface epitopes that are often composed of conformational structures that are noncontiguous in the antigen's primary sequence. In contrast, most TCRs recognize fragments of antigens, both surface and internal, that have been processed by a separate antigen-presenting cell and then bound to a major histocompatibility complex (MHC) class I or class II molecule (Chapters 5 and 6).

THE B-CELL AND T-CELL RECEPTOR ANTIGEN RECOGNITION COMPLEX

Both B-cell antigen receptor (BCR) and TCR cytoplasmic domains are exceptionally short. In order for surface binding of antigen to elicit a response from the cell, the BCR and TCR each associate noncovalently with signal transduction complexes: heterodimeric Igα:Igβ (also known as CD79a:CD79b, respectively) for B cells, and multimeric CD3 for T cells. Loss-of-function mutations in either of these complexes leads to cell death, which becomes clinically manifest as hypogammaglobulinemia in the case of B cells (Chapter 33), or severe combined immune deficiency (SCID) in the case of T cells (Chapter 34).

IMMUNOGLOBULINS AND T-CELL RECEPTOR STRUCTURES

The Immunoglobulin Domain, the Basic Immunoglobulin Superfamily Building Block

Igs consist of two heavy (H) and two light (L) chains (Fig. 4.1). The L chain can be either a κ or a λ chain. TCRs consist of either

an αβ or a γδ heterodimer. Each component chain contains two or more IgSF domains, each of which consists of two sandwiched β pleated sheets "pinned" together by a disulfide bridge between two conserved cysteine residues.[1] Considerable variability is allowed to the amino acids that populate the external surface of the IgSF domain and the loops that link the β strands. These solvent-exposed surfaces offer multiple targets for docking with other molecules.

Two types of IgSF domains, "constant" (C) and "variable" (V), are used in Igs and TCRs (see Fig. 4.1). C-type domains, which are the most compact, have seven antiparallel strands distributed as three strands in the first sheet and four strands in the second. Side chains positioned to lie between the two strands tend to be nonpolar in nature, creating a hydrophobic core of considerable stability. V-type domains add two additional antiparallel strands to the first sheet, creating a five-strand–four-strand distribution. The two additional strands, which encode framework region 2 (FR2), are used to steady the interaction between heterodimeric V domains, allowing them to create a stable antigen-binding site.[2]

While each Ig and TCR chain contains only one amino-terminal V Ig domain, the number of carboxy-terminal C domains varies. Ig H chains contain between three and four C domains, whereas both Ig L chains and all four TCR chains contain only one C domain each. IgH chains with three C domains tend to include a spacer hinge region between the first (C_H1) and second (C_H2) domains. Each V or C domain consists of approximately 110 to 130 amino acids, averaging 12,000 to 13,000 kDa. A typical L or TCR chain will thus mass approximately 25 kDa, and a three C domain Cγ H chain with its hinge will mass approximately 55 kDa.

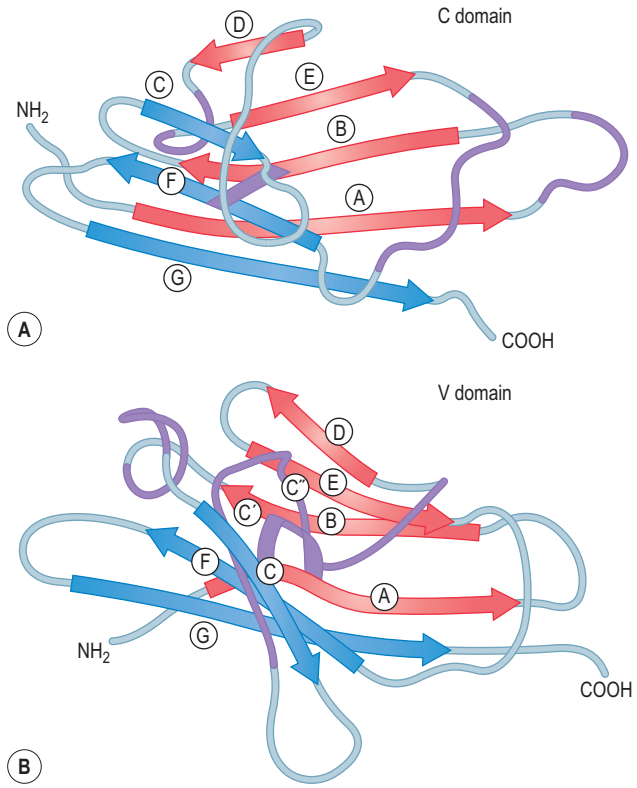


FIG. 4.1 Immunoglobulin Superfamily Domain Structures. (A) A typical compact C domain structure. The β strands are labeled A through G. The sequence at the core is conserved and nonpolar. The external surface and the β loops are available for docking and often vary in sequence. (B) A typical V domain structure. Two additional strands, C′ and C″, have been added. Note the projection of the C–C′ strands and loop away from the core.

Idiotypes and Isotypes

Immunization of heterologous species with monoclonal antibodies (mAbs; or a restricted set of Igs) has shown that Igs and TCRs contain both common and individual antigenic determinants. Individual determinant(s), termed *idiotype(s)*, are contained within V domains. Common determinants, termed *isotypes*, are specific for the constant portion of the antibody and allow grouping of Igs and TCRs into recognized classes. Each class defines an individual type of C domain. Determinants common to subsets of individuals within a species, yet differing between other members of that species, are termed *allotypes* and define inherited polymorphisms that result from allelic forms of the genes.[3]

The V Domain

Three hypervariable intervals, termed complementarity-determining regions (CDRs), that are situated between four framework regions of stable sequence (Fig. 4.2) can be distinguished by comparisons of the primary sequences of V domains.[4] The international ImMunoGeneTics information system, or IMGT, maintains an extremely useful website, http://www.imgt.org that contains a large database of Ig and TCR sequences as well as a multiplicity of software tools for their analysis.

Antigen Recognition and the Fragment Antigen Binding

Initial studies of Ig structure were facilitated by the use of papain and pepsin to fragment IgG molecules. Papain digests IgG into two antigen-binding fragments (Fab) and a single crystallizable (or constant) fragment (Fc). Pepsin splits IgG into an Fc fragment and a single dimeric $F(ab')_2$ that can cross-link as well as bind antigens. The Fab contains one complete L chain in its entirety and the V and C_H1 portion of one H chain (see Fig. 4.2). The Fab can be further divided into a variable fragment (Fv) composed of the V_H and V_L domains, and a constant fragment (Fb) composed of the C_L and C_H1 domains. Single Fv fragments can be genetically engineered to recapitulate the monovalent antigen-binding characteristics of the original, parent antibody.[5] The extracellular domains of TCRαβ and TCRγδ correspond to Ig Fab.

Effector Function and the Fragment Crystallizable

The Fc portion (see Fig. 4.2) encodes the effector functions of the Ig. These functions are generally inflammatory reactions that include fixation and activation of complement, and binding of antibody to Fc receptors on the surface of other cells (Chapter 8). Each Ig class and subclass exhibits its own set of effector functions.[6] For example, the IgG C_H2 domain plays a key role in complement fixation and in binding to class-specific Fc receptors on the surface of effector cells. Both these interactions are important in initiating the process of phagocytosis, in allowing certain subclasses to traverse the placenta, and in influencing the biologic functions of lymphocytes, platelets, and other cells.

Gm Allotype System

A series of serologically defined C-domain allotypes have been identified. In the case of the H chain, they are termed Gm for

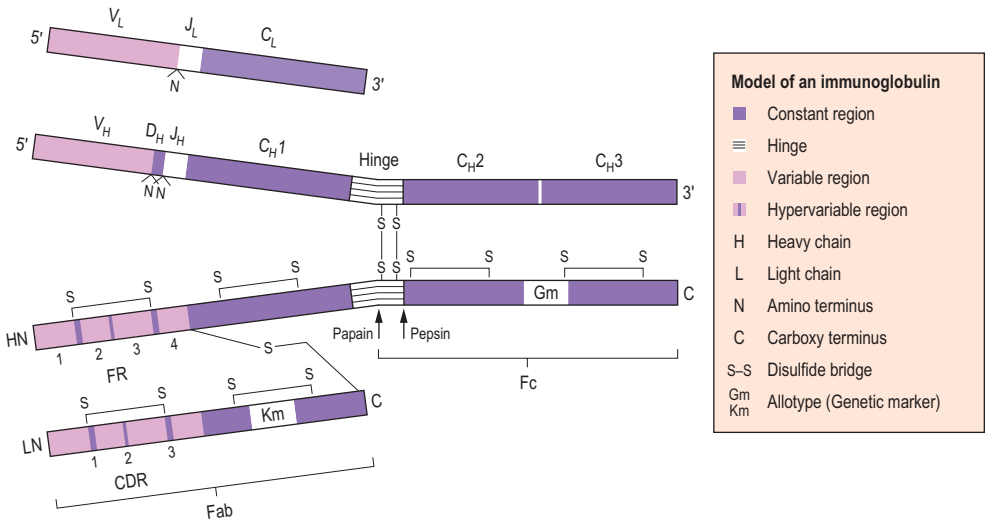


FIG. 4.2 A Two-Dimensional Model of an Immunoglobulin G Molecule. The top H and L chains illustrate the composition of these molecules at a nucleotide level. The bottom chains illustrate the nature of the protein sequence. See text for further details. *CDR*, Complementarity-determining region; *Fab*, fragment antigen binding; *Fc*, fragment crystallizable; *FR*, framework region.

TABLE 4.1 Selected Properties of Immunoglobulin (Ig) Classes

	IgG	IgA	IgM	IgD	IgE
Molecular weight	160,000	170,000 or polymer	900,000	160,000	180,000
Approximate concentration in serum (mg/dL)	700–1500	75–400	50–250	0.3–30	0.0015–0.2
Valence	2	2 (monomer)	10 (small antigen) 5 (large antigen)	2	2
Molecular formula	γ_2L_2	$(\alpha_2L_2)n$	$(\mu_2L_2)_5$	δ_2L_2	ε_2L_2
Half-life (days)	23	6	5	3	2.5
Special property	Placental passage	Secretory Ig	Primary response lymphocyte surface	Lymphocyte surface	Immediate hypersensitivity reactions

KEY CONCEPTS

Immunoglobulin and T-Cell Receptor Structure

- Both Igs and TCRs are heterodimeric proteins.
  - Igs consist of two identical H and two L chains.
  - αβ TCRs consist of one α and one β chain.
  - γδ TCRs consist of one γ and one δ chain.
- Ig and TCR contain two or more IGSF domains, which are identified by their characteristic beta barrel structure.
  - Each Ig and TCR chain contains a V-type IgSF domain that will form one-half of the antigen-binding site.
  - Each V domain contains three hypervariable intervals known as CDRs.
- The CDRs of paired heterodimers chains are juxtaposed to form the antigen-binding site.
- The C domains of Ig H chains define the Ig class or subclass.
- The two distal C IgH domains determine the effector function of the antibody.

the gammaglobulin fraction of the serum in which they were first identified.[3] Allotypes have been identified for γ1, γ2, γ3, γ4, α2, and ε H chains and for the κ L chain. Associations between certain Gm allotypes and predisposition to develop certain diseases of immune function have been reported.

IMMUNOGLOBULIN CLASSES AND SUBCLASSES

The constant domains of the H chain define the class and subclass of the antibody. Table 4.1 lists the five major classes of Igs in human and describes some of their physical and chemical features. Two of the five major H-chain classes, α and γ, have undergone duplication. IgG1, IgG2, IgG3, and IgG4 all have the same basic structural design and differ only in the primary sequence of their constant regions and in the location of their interchain disulfide bonds. The H chain in each of these subclasses is referred to as γ1, γ2, etc. IgA consists of the two subclasses, α1 and α2. Table 4.2 compares the four subclasses of IgG, the two of IgA, and the classes of IgM, IgD, and IgE from the standpoint of their biologic functions. In humans, the two L-chain classes, κ and λ, are expressed at roughly equal frequencies. No specific effector function has been identified for either L-chain class.

Immunoglobulin M

IgM exists in monomeric, pentameric, and hexameric forms. The 8 S monomeric 180 kDa IgM has the molecular formula μ_2L_2. It is a minor fraction in serum, but in its transmembrane form IgM plays a key role in B-cell development and function as the antigen recognition portion of the B-cell antigen receptor. The major form in serum is the 19 S, 900 kDa pentameric IgM, which contains five

TABLE 4.2 Selected Biologic Properties of Classes and Subclasses of Immunoglobulins (Igs)

	IgG				IgA				
	1	2	3	4	1	2	IgM	IgD	IgE
Percentage of total (%)	65	20	10	5	90	10			
Complement fixation	++	+	++	–	–	–	++	–	–
Complement fixation (alternative)			+	+	+/–	+/–			
Placental passage	+	+	+	+	–	–	–	–	–
Fixing to mast cells or basophils	–	–	–	–	–	–	–	–	+
Binding to:									
Lymphocytes	+	+	+	+	–	–	+	–	–
Macrophages	+	+/–	+	+/–	–	–	–	–	–
Neutrophils	+	+	+	+	+	+	–	–	–
Platelets	+	+	+	+	–	–	–	–	–
Reaction with stapylococcal protein A	+	+	–	+	–	–	–	–	–
Half-life (days)	23	23	8–9	23	6	4.5	5	3	2.5
Synthesis (mg/kg/day)	25	?	3.5	?	24	?	7	0.4	0.02

subunits [$(\mu_2L_2)_5$] linked together by disulfide bridges, and by one molecule of an additional polypeptide chain, the J chain, which joins two of the subunits by a disulfide bridge.[7]

IgM is the first Ig to be secreted (as pentameric IgM) during development and in the absence of antigenic challenge. It is also the predominant Ig produced during the primary immune response. Occasionally, particularly in the case of carbohydrate antigens such as isohemagglutinins, it will remain the major or sole antibody class. IgM differs from most other Igs in having an extra C_H domain in place of a hinge.

IgM avidly fixes complement. This property is focused in CH3, the homologue of IgG CH2.[8] Although the valence of each μ_2L_2 subunit is 2, when binding to large protein antigens 5 of the 10 antigen-binding sites in pentameric IgM appear blocked due to steric hindrance. As a consequence, the valence for large antigens is five.

Immunoglobulin G

IgG, the major Ig class, generally exists in a monomeric form (γ_2L_2) in the serum, although it can form hexameric assemblies at cell surfaces.[9] IgG accounts for the bulk of serum antibody activity in response to most protein antigens.

The four IgG subclasses are numbered in relation to their serum levels relative to each other, with IgG1 predominant and IgG4 the least common.[10] IgG1 and IgG3 fix complement and bind phagocyte Fcγ receptors well, whereas IgG2 fixes complement but binds Fcγ receptors more poorly. Recently it has been shown that IgG1 can use non-covalent Fc-Fc interactions to form hexameric assemblies, which facilitate complement activation. IgG4 does not fix complement effectively in the native state but has been reported to do so after proteolytic cleavage. IgG1 and IgG3 are most frequently elicited by viral antigens, IgG2 by carbohydrates, and IgG4 by helminthic parasites.

IgG4 can attenuate allergic responses by inhibiting the activity of IgE.[11] IgG4 can function as a blocking antibody, preventing cross-linking of receptor bound IgE. It can co-stimulate the inhibitory IgG receptor FcγRIIb, which can negatively regulate FcεRI signaling and thus inhibit effector cell activation. Finally, the disulfide bonds of the IgG4 hinge are easily reduced, which allows the H chains to separate and randomly re-associate to produce a mixed population of IgG4 molecules with randomized heavy-chain and light-chain pairs. This impairs the ability of IgG4 to form immune complexes and thus has an anti-inflammatory effect, facilitating immunotherapy for allergic diseases (allergy shots).

Overproduction of IgG4 is seen in a disparate group of inflammatory diseases. Fibro-inflammatory masses can develop in virtually all organs except the brain, with an unexplained preference for salivary glands, lymph nodes, and pancreas. Together, these are referred to as IgG4-related disease (IgG4-RD).[12]

Immunoglobulin A

IgA generally exists in a monomeric form (α_2L_2) in the serum. However, it can also interact with J chain to form a polymer $(\alpha_2L_2)_{2,3}$-J. Second in concentration to IgG in serum, IgA functions as the predominant form of Ig in mucosal secretions.[13]

Secretory IgA (SIgA) is largely synthesized by plasma cells associated with mucosal tissues. In the secretions, the molecule typically exists in polymeric form with two subunits in association with the 70 kDa secretory component $(\alpha_2L_2)_2$-SC. SC is synthesized by the epithelial cells that line the lumen of the gut. It appears to render the SIgA complex more resistant to proteolytic digestion and it enhances the immune functions of SIgA.

Immunoglobulin E

IgE is largely found in extravascular spaces. Its plasma turnover is rapid, with a half-life of about 2 days. IgE antibodies help protect the host from parasitic infections (Chapter 30). In Westernized, affluent societies, IgE is primarily associated with allergy. Through their interaction with Fcε receptors on mast cells and basophils, IgE antibodies, in the presence of antigens, induce the release of histamine and various other vasoactive substances, which are responsible for clinical manifestations of various allergic states.[14]

Immunoglobulin D

Although the H chain of IgD can undergo alternative splicing to a secretory form, IgD serum antibodies in human are uncommon and are absent in the serum of mice and primates. Instead, IgD typically is co-expressed with IgM on the surface of mature lymphocytes. The appearance of IgD is associated with the transition of a B lymphocyte from a cell that can be tolerized to antigen to a cell that will respond to antigen with the production of antibody (Chapter 7).

T-CELL RECEPTOR αβ AND γδ

TCR α, β γ, and δ chains are members of the IgSF and thus share a number of structural similarities with Ig. Each chain contains

a leader peptide, and extracellular, transmembrane, and intracytoplasmic components. The extracellular component can be divided into three domains: a polymorphic V domain encoded by VJ (α and γ chains) or VDJ (β and δ chains) gene segments, a C domain, and a hinge region.[15] The hinge region typically contains an extra cysteine (none in γ chains encoded by Cγ2) that forms a disulfide bond with the other partner of the heterodimer. The transmembrane domains all include a lysine plus or minus an arginine residue that facilitate the association of the TCR heterodimer with components of the CD3 signal transduction complex, each of which has a matching negatively charged residue in their own transmembrane portions (see below). The intracytoplasmic components are tiny and play a relatively limited role in signal transduction.

T-Cell Receptor αβ

The TCR α and β chains are glycoproteins with molecular weights that vary from 42 to 45 kDa, depending upon the primary amino acid sequence and the degree of glycosylation. Deglycosylated forms have a molecular mass of 30 to 32 kDa. These chains share a number of invariant residues in common with Ig heavy and light chains, in particular residues that are thought to be important for interactions between heavy and light chains. The structures of a number of partial or full-length TCRs have been solved by X-ray crystallography (Fig. 4.3).[16] In general, the structure of the TCR αβ heterodimer is similar, but not identical, to that of an Ig Fab fragment.

T-Cell Receptor γδ

The TCR γ and δ chains are glycoproteins with a more complex molecular size pattern than α and β chains. TCRs that use the Cγ1 gene segment, which contains a cysteine-encoding exon are disulfide-linked (MW 36 to 42 kDa). TCRs that use Cγ2 exist in two non-disulfide-linked forms, one of 40 to 44 kDa and one of 55 kDa.[17] The differences in molecular size are due to variability of both N-linked glycosylation and primary amino acid sequence. The 55-kDa form uses a Cγ2 allele that contains three (rather than two) exons encoding the connecting piece, as well as more N-linked carbohydrate. The TCR δ chain is more straightforward, being 40 to 43 kDa in size and containing two sites of N-linked glycosylation. The overall architecture of the γδ TCR closely resembles that of αβ TCRs and antibodies, although the angle between the V and C domains, known as the elbow angle, appears more acute.

Ligand Recognitions

TCR αβ T cells primarily recognize peptide-MHC complexes (pMHC) (see Fig. 4.3; Chapters 5 and 6); however, other types of ligands exist. Some αβ TCRs can bind nonpeptidic antigens (atypical antigens) that are bound to "non-classic" MHC class Ib molecules. These specificities tend to define unique populations of T cells. For example, some natural killer T cells (NKT) express a distinctive combination of TCR αβ chains that recognize lipid-based antigens presented by CD1d molecules, while mucosal-associated invariant T cells (MAIT), using other combinations of TCR αβ chains, recognize vitamin metabolites presented by the MHC class-I related gene protein (MR1).[18] Many γδ T cells recognize atypical antigens that may or may not be associated with an antigen-presenting molecule, although some can bind peptides. Finally, many αβ TCRs bind superantigens (SAgs) in a predominantly Vβ-dependent fashion (Chapter 6).

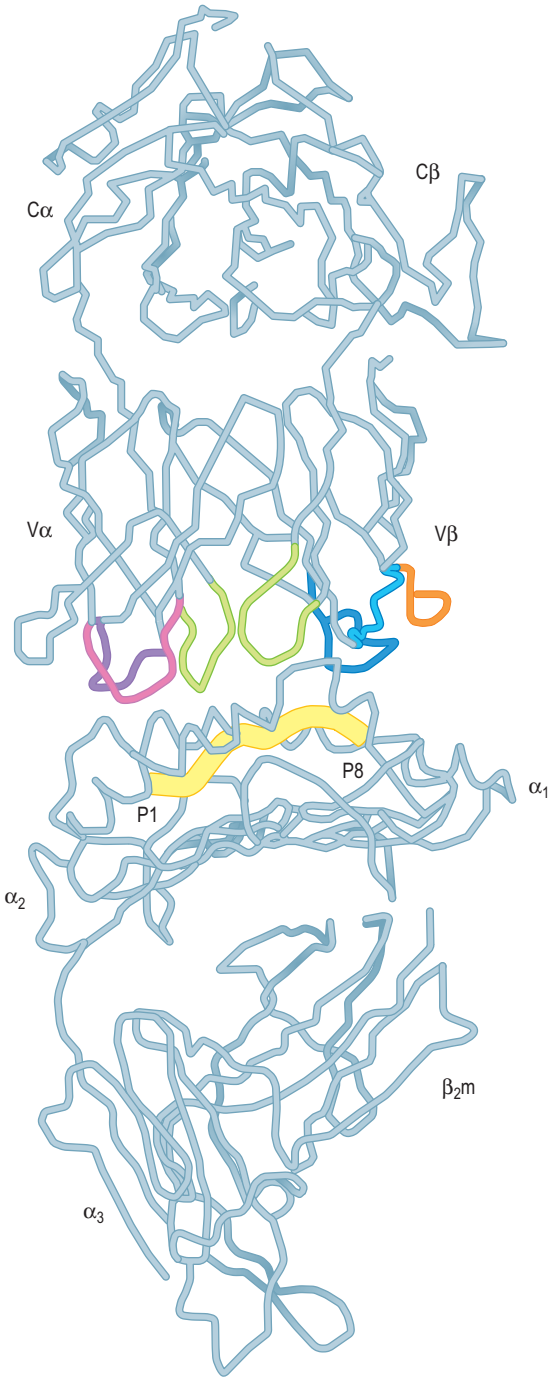


FIG. 4.3 Backbone Representation of Murine αβ T-Cell Receptor Bound to Murine Major Histocompatibility Complex Class I and an Octamer Peptide. The T-cell receptor is above. The Vα CDR1 and CDR2 are *magenta*, Vβ CDR1 and CDR2 are *blue*, both CDR3s are *green*, and the Vβ HV4 is *orange*. β_2m refers to β_2-microglobulin. The peptide is in *yellow*, and the NH_2-terminal and COOH-terminal residues are designated *P1* and *P8*. (Reproduced with permission from Garcia KC, Degano M, Stanfield RL, et al. An alphabeta T cell receptor structure at 2.5 A and its orientation in the TCR-MHC complex. *Science*. 1996;274(5285):209–219.)

Binding to Peptide-Major Histocompatibility Complex

TCRs recognize peptide antigens bound to the binding groove of MHC-encoded glycoproteins (see Fig. 4.3). TCR recognition of pMHC requires a trimolecular complex in which all the components (antigen, MHC, and TCR) contact one another.[16] Thus, recognition is highly influenced by polymorphisms in the MHC molecule (Chapter 5). As in the case of Ig, TCR CDR1 and CDR2 are encoded in the germline V regions, whereas CDR3 is formed at the junction of the V gene with a J gene segment (TCR α and γ) or D and J gene segments (TCR β and δ chains). Vβ also has a fourth region of variability within Framework 3 that is juxtaposed to the other CDRs in the tertiary structure. This region, variously termed hypervariable region 4 (HV4) or CDR4, can participate in SAg binding.

The co-crystallization of different combinations of soluble TCR αβ interacting with MHC class I bound to antigen peptide (pMHC) has made it possible to directly address the manner in which antigen recognition occurs (see Fig. 4.3). The TCR αβ combining site is relatively flat, allowing it to interact with a rather broad surface at the point of contact with pMHC.

The TCR footprint on the pMHC complex tends to occur in a diagonal across the MHC antigen-binding groove, with TCR Vα positioned over the MHC α2 helix and TCR Vβ overlying the MHC α1 helix. This geometry would permit consistent access of the CD8 co-receptors to the MHC class I molecule. The CDR1 and CDR2 loops, which are entirely encoded by germline sequence, tend to interact more with the MHC molecule; whereas the CDR3 loops, which are composed of both germline and somatic (N region) sequence, appear to dominate the interaction with MHC-bound peptide.

The binding of TCR to pMHC appears to be driven by enthalpy: that is, binding increases the stability of the CDR loops, especially CDR3. These results have led to the suggestion that initial binding focuses on the interaction between CDRs 1 and 2 and the MHC. After this initial recognition, the CDR3s change their shape to maximize the area of contact. Conformational flexibility, or "induced fit," would allow TCRs to rapidly sample many similar pMHC complexes, stopping only when their CDR3s are able to stabilize the interaction.

T-Cell Receptor Binding Affinity

The affinity with which the TCR ultimately binds its ligand is a critical determinant of T-cell activation. It is, however, only one factor in determining the overall avidity of the interaction, since other cell-surface molecules of the T cell (e.g., CD4, CD8, CD2, and various integrins) bind to cell-surface molecules on the antigen-bearing cell to stabilize cell–cell TCR–ligand interactions. Furthermore, since both the TCR and the pMHC ligand are surface membrane proteins, each T cell can provide multiple TCRs in the same plane that can bind multiple pMHC molecules on the surface of the antigen-presenting cell. This makes binding of TCR to pMHC functionally multivalent, enhancing the apparent affinity of the interaction.

Atypical Antigens

Some αβ T cells can recognize lipid antigens when they are complexed with members of the CD1 family. There are four members of the CD1 family that are expressed on the cell surface. TCR that recognize CD1a, b, and c likely follow the same recognition scheme as T cells. But crystal structures of NKT TCR that recognize CD1d show binding that is parallel, rather than diagonal. Allelic polymorphism in CD1 is limited, which theoretically would restrict the range of lipid antigens that can be bound. However, a surprisingly large variety of different lipid-based antigens can be accommodated.[19]

Rather than binding to a single groove on the MHC, lipids attach themselves to one of several hydrophobic pockets that can be found on the surface of CD1. Pocket volume can range from 1300 to 2200 $Å^{trs3}$. The number and length of the pockets differ between the various CD1 isoforms. For example, CD1b has three pockets that share a common portal of entry, as well as a fourth pocket that connects two of the three pockets to each other. This connecting pocket permits the insertion of lipids with a long alkyl chain, such as mycobacterial mycolic acid.

MAIT cells are generally low in frequency in laboratory mice. However, they are abundant in humans, on average representing approximately 5% of total blood T cells, 10% of CD8 T cells, and up to 45% of liver T cells. Natural MR1 ligands include derivatives of vitamin B_9 (folate) and also unstable pyrimidine intermediates derived from the vitamin B_2 (riboflavin) synthesis pathway. A range of small organic molecules, drugs, drug metabolites and drug-like molecules, including salicylates and diclofenac, have also been described as MR1-binding ligands.[20] The MR1 Ag-binding cleft is ideally predisposed to bind to these small metabolites, forming an aromatic cradle that closely sequesters the ligands, with some of the ligands forming a covalent bond (Schiff base) with MR1. The MAIT TCR docks in a conserved manner with MR1, analogous to typical TCR–MHC-I–peptide docking, in which the α and β chains of the MAIT TCR are positioned over the α2 and α1 helices of MR1, respectively.

γδ T cells are activated by conserved stress-induced ligands, enabling them to rapidly produce cytokines that regulate pathogen clearance, inflammation, and tissue homeostasis in response to tissue stress.[21] Antigen recognition by γδ TCRs resembles recognition of intact antigens by antibodies more closely than recognition of pMHC by αβ TCR. γδ TCRs can recognize protein antigens, such as nonclassical MHC molecules and viral glycoproteins, as well as small phosphate- or amine-containing compounds, such as pyrophosphomonoesters from mycobacteria and alkylamines.

Binding to nonpeptide antigens plays an important role in the biology of γδ T cells. About 5% of peripheral blood T cells bear γδ TCRs, and most of these are encoded by Vγ9 JγP and Vδ2 gene segments. (In an alternative nomenclature, Vγ9 is known as Vγ2 and JγP as Jγ1.2. See the IMGT database at: http://www.imgt.org.) These Vγ9 JγPVδ2 TCRs recognize nonpeptide pyrophosphate- or amine-containing antigens, such as pyrophosphomonoesters from mycobacteria or isobutylamine from various sources. Other common naturally occurring small phosphorylated metabolites that stimulate γδ T cells include 2,3-diphosphoglyceric acid, glycerol-3-phosphoric acid, xylose-1-phosphate, and ribose-1-phosphate. In addition to mycobacteria, Vγ9 JγPVδ2 T-cell populations are seen to expand in response to listeriosis, ehrlichiosis, leishmaniasis, brucellosis, salmonellosis, mumps meningitis, malaria, and toxoplasmosis.

Superantigens

SAgs are a special class of TCR ligands that have the ability to activate large fractions (5% to 20%) of the T-cell population (Chapter 6). Activation requires simultaneous interaction between the SAg, the TCR Vβ domain, and an MHC class II molecule on the surface of an antigen-presenting cell. Unlike conventional antigens, SAgs do not require processing to allow them to bind class II molecules

KEY CONCEPTS

Features Common to Immunoglobulin and T-Cell Receptor Genes

- Ig and TCR variable domains are created by site-specific V(D)J recombination.
- Starting with a small set of individual gene segments, combinatorial gene segment rearrangement, combinatorial association of H and L chains, or TCR β and α, and mechanisms of junctional diversity generate a broad repertoire of antigen-binding structures.
- Each receptor is assembled in a stepwise fashion
  - Igs: $D_H \rightarrow J_H$; $V_H \rightarrow D_H J_H$; cytoplasmic μ chain expression; Vκ→Jκ and, if needed, $V_\lambda \rightarrow J_\lambda$; surface IgM expression.
  - TCRαβ: Dβ→Jβ; Vβ→DβJβ; cytoplasmic β-chain expression; Vα→Jα; surface αβ TCR expression.
- CDRs 1 and 2 begin with exclusively germline sequence.
- CDR3 is created by the (V[D]J) joining reaction and often includes nongermline N nucleotides between the V and the D, and between the D and the J.
  - Thus, CDR-H3, CDR-B3, and CDR-D3 are the most variable components of IgM, TCRαβ, and TCRγδ; respectively.
- The antigen-binding site is a product of a nested gradient of diversity. Conserved framework regions surround CDR1 and CDR2, which in turn surround the paired CDR3 intervals that form the center of the antigen-binding site.
- The variability of the Ig and TCR repertoires is restricted during pregnancy, limiting the immune response of the fetus and newborn infant.

or activate T cells. Instead of binding to the peptide antigen-binding groove, SAgs interact with polymorphic residues on the periphery of the class II molecule. And, instead of binding to TCR β CDR3 residues, SAg interact with polymorphic residues in CDR1, CDR2, and HV4. Soluble TCR β chains can also bind the appropriate SAg in the absence of a TCR α chain. As a consequence, although SAg link the TCR to the MHC, the T-cell responses are not "MHC-restricted" in the conventional sense,

KEY CONCEPTS

Features Specific to Immunoglobulin Genes

- Variable domain somatic hypermutation (SHM) permits affinity maturation, which further diversifies the B-cell repertoire.
- Class switch recombination (CSR) allows replacement of an upstream C domain by a downstream one, altering effector function while maintaining antigen specificity.

since a T cell with the appropriate Vβ will respond to a SAg bound to a variety of polymorphic class II molecules.

IMMUNOGLOBULIN GENE ORGANIZATION

The component chains of Igs and TCRs are each encoded by a separate multigene family.[22,23] The paradox of variability in the V region in conjunction with a nearly invariable constant region was resolved when it was shown that immunoglobulin V and C domains are encoded by independent elements, or gene segments, within each gene family. As a result, several gene elements are used to encode a single polypeptide chain. For example, while κ constant domains are encoded by a single Cκ exon in the κ locus on chromosome 2, κ variable domains represent the joined product of Vκ and Jκ gene segments (Fig. 4.4).

V_L gene segments typically contain their own promoter, a leader exon, an intervening intron of approximately 100 nucleotides, an exon that encodes the first three framework regions (FR 1, 2, and 3), the first two CDRs in their entirety, the amino terminal portion of CDR3, and a recombination signal sequence (RSS). A J_L (J for joining) gene segment begins with its own recombination signal, the remaining portion of CDR 3, and the complete FR 4 (see Fig. 4.2).

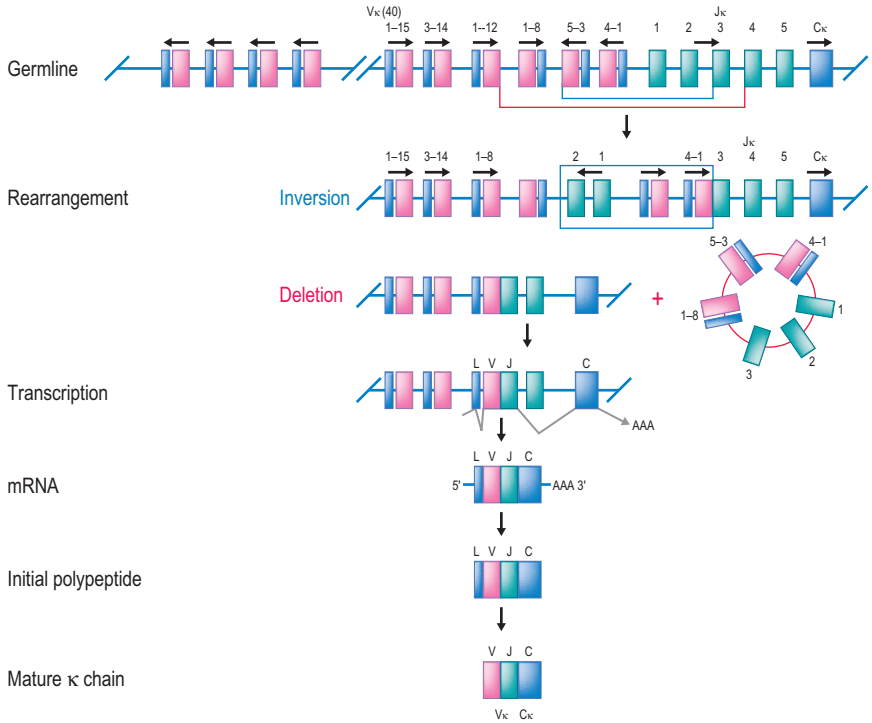


FIG. 4.4 Rearrangement Events in the Human κ Locus. *C*, Constant region of the κ light chain; *J*, joining region; *V*, variable region. See text for further description.

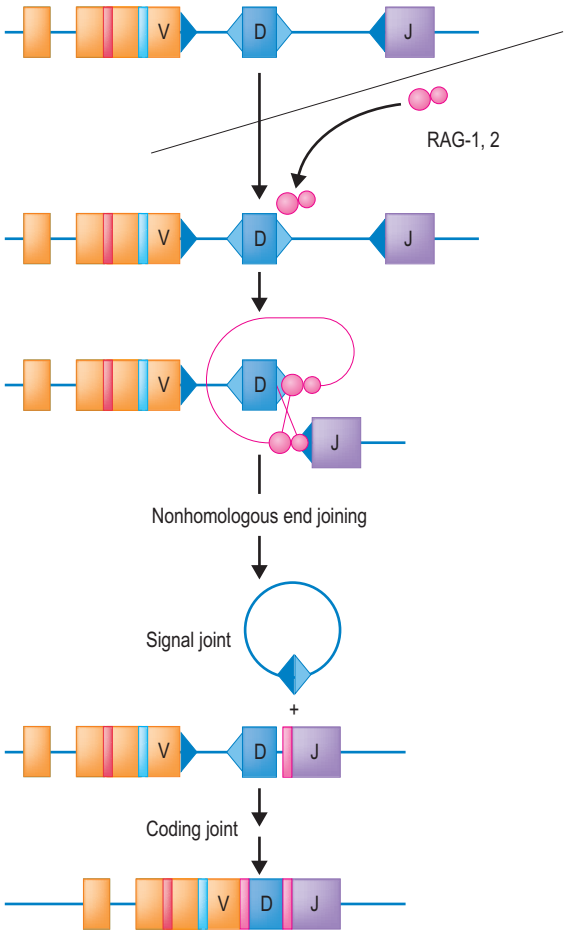


FIG. 4.5 VDJ Recombination. Lymphoid-specific RAG-1 and RAG-2 bind to the recombination signal sequences (RSS) flanking V, D, or J gene segments, juxtapose the sequences, and introduce precise cuts adjacent to the RSS. Components of the nonhomologous end-joining repair pathway subsequently unite the cut RSSs to form a signal joint, and the coding sequences of the rearranging gene segments to form a coding joint.

(Use of the same abbreviation—V—for both the complete variable domain of an Ig peptide chain and for the gene segment that encodes only a portion of that same variable domain is the result of historic precedent. It is unfortunate that one must depend on the context of the surrounding text in order to determine which V region of the antibody is being discussed. The same holds true for the use of J to represent both the J gene segment and the J joining protein.)

The creation of a V domain is directed by the RSSs that flank the rearranging gene segments.[24] Each RSS contains a strongly conserved seven base-pair, or heptamer, sequence (e.g., CACAGTG) that is separated from a less well-conserved nine base-pair, or nonamer, sequence (e.g., ACAAAACCC) by either a 12- or 23-base-pair (bp) spacer. For example, Vκ gene segments have a 12-bp spacer and Jκ elements have a 23-bp spacer. These spacers place the heptamer and nonamer sequences on the same side of the DNA molecule, separated by either one or two turns of the DNA helix. A one-turn RSS (12-bp spacer) will preferentially recognize a two-turn signal sequence (23-bp spacer). This helps prevent wasteful V-V or J-J rearrangements.

Initiation of the V(D)J recombination reaction requires recombination activating genes 1 and 2 (*RAG1* and *RAG2*). These genes are expressed only in developing lymphocytes.[25] The gene products, RAG-1 and RAG-2, act by precisely introducing a DNA double-strand break (DSB) between the terminus of the rearranging gene segment and its adjacent RSS (Fig. 4.5). These breaks are then repaired by ubiquitously expressed components of a DNA repair process that is known as nonhomologous end joining (NHEJ). Lack-of-function mutations in NHEJ proteins yields susceptibility to DNA damage in all cells of the body and can lead to a SCID phenotype (Chapter 34).

The NHEJ process creates precise joins between the RSS ends, and imprecise joins of the coding ends. TdT, which is expressed only in lymphocytes, adds non-germline encoded nucleotides (N-nucleotides) to the coding ends of the recombination product.

Lymphoid-specific expression of RAG-1 and RAG-2 limits V(D)J recombination to B and T lymphocytes. To ensure that TCR genes are rearranged to completion only in T cells and Ig genes are rearranged to completion only in B cells, V(D)J recombination is further regulated by limiting the accessibility of the appropriate gene segments to the specific lineage as well as to the specific stage of development. For example, H-chain genes are typically assembled before L-chain genes.

The RAG-1 and RAG-2 recombinases cooperatively associate with 12- and 23-bp RSSs and their flanking coding gene segments to form a synaptic complex. Typically, the initial event will be recognition of the nonamer sequence of a 12-bp spacer RSS by RAG-1, which appears to function as the catalytic component of the recombinase. RAG-1 binding to the heptamer provides specificity. RAG-2 does not bind DNA independently but does make contact with the heptamer when in a synaptic complex with RAG-1. Binding of a second RAG-1 and RAG-2 complex to the 23-bp, two-turn RSS permits the interaction of the two synaptic complexes to form what is known as a paired complex. Creation of this paired complex is facilitated by the actions of the DNA-bending proteins HMGB1 and HMGB2 and by the presence of a divalent metal ion.

After paired complex assembly, the RAG proteins single-strand cut the DNA at the heptamer sequence. The 3′ OH of the coding sequence ligates to 5′ phosphate and creates a hairpin loop. The clean-cut ends of the signal sequences enable formation of precise signal joints. However, the hairpin junction created at the coding ends must be resolved by re-nicking the DNA, usually within four to five nucleotides from the end of the hairpin. This forms a 3′ overhang that is amenable to further diversification. It can be filled in via DNA polymerases, nibbled back, or may serve as a substrate for TdT-catalyzed N addition. DNA polymerase μ, which shares homology with TdT, appears to play a role in maintaining the integrity of the terminus of the coding sequence.

The cut ends of the coding sequence are then repaired by the nonhomologous end-joining proteins. NHEJ proteins involved in V(D)J recombination include Ku70, Ku80, DNA-PKcs, Artemis, XRCC4, XLF (Cernunnos), and ligase 4.

Ku70 and Ku80 form a heterodimer (Ku) that directly associates with DNA DSBs to protect the DNA ends from degradation, permit juxtaposition of the ends to facilitate coding end ligation, and help recruit other members of the repair complex. The DNA protein kinase catalytic subunit (DNA-PKcs) phosphorylates Artemis, inducing an endonuclease activity that plays a role in the opening of the coding joint hairpin. Thus absence

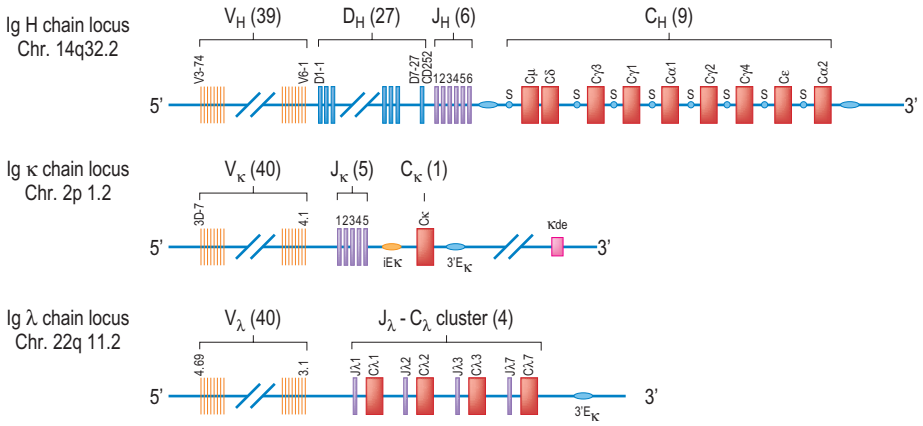


FIG. 4.6 Chromosomal Organization of the Ig H, κ, and λ Gene Clusters. The typical numbers of functional gene segments are shown. The κ gene cluster includes a κ deleting element that can rearrange to sequences upstream of Cκ in cells that express λ chains, reducing the likelihood of dual κ and λ light-chain expression. These maps are not drawn to scale.

of DNA-PKcs or Artemis inhibits proper coding joint formation. Signal joint formation is normal in Artemis deficiency, but it is impaired in the absence of DNA-PKcs. Finally, XRCC4, XLF, and ligase 4 help rejoin the ends of the broken DNA.

Depending on the transcriptional orientation of the rearranging gene segments, recombination can result in either inversion or deletion of the intervening DNA (see Fig. 4.3). The products of inversion remain in the DNA of the cell, whereas deletion leads to the loss of the intervening DNA. The increased proximity of the V promoter to the C domain enhancers promotes the subsequent transcription of the Ig gene product.

There is a price to the use of V(D)J recombination to create antigen receptor diversity. Aberrant recombination in nonreceptor genes can create deleterious genomic rearrangements that promote B-cell and T-cell neoplasias.[24] For example, deletional recombination at the SIL/SCL locus and in *Notch1, Izkf1, PTEN*, and other critical genes appear to be major drivers of lymphoid neoplasms in humans and in mice.

The κ Locus

The κ locus is located on chromosome 2p11.2. It contains 5 Jκ and 75 Vκ gene segments upstream of Cκ (Fig. 4.6). The Vκ gene segments can be grouped into six different families of varying size.[26] Each family is composed of gene segments that share extensive sequence and structural similarity.[27]

One-third of the Vκ gene segments contain frameshift mutations or stop codons that preclude them from forming functional protein, and of the remaining sequences less than 30 of the Vκ gene segments have actually been found in functional Igs. Each of these active Vκ gene segments has the potential to rearrange to any one of the 5 Jκ elements, generating a potential repertoire of more than 140 distinct VJ combinations. Even more diversity is created at the site of gene segment joining. The terminus of each rearranging gene segment can undergo a loss of 1 to 5 nucleotides during the recombination process. In humans, but not mice, N addition can either replace some or all of the lost nucleotides or can be inserted in addition to the original germline sequence. Each codon created by N addition increases the potential diversity of the repertoire 20-fold. Thus, the focus for the diversity of the κ repertoire lies in the VJ junction that defines CDR-L3.

The λ Locus

The λ locus, on chromosome 22q11.2, contains four functional Cλ exons, each of which is associated with its own Jλ (see Fig. 4.6). The Vλ genes are arranged in three distinct clusters, each containing members of different Vλ families.[28] Depending on the individual haplotype, there are approximately 30 to 36 potentially functional Vλ gene segments and an equal number of pseudogenes.

In addition to normal κ and λ peptides, H chains can also form a complex with unconventional λ light chains, known as surrogate or pseudo light chains (SLC). The genes encoding the SLC proteins, VpreB and λ5 (λ14.1), are located within the λ light-chain locus on chromosome 22 and are restricted in expression to discreet B-cell developmental stages (Chapter 7). Together, these two genes create a product with considerable homology to conventional λ light chains. A critical difference between these unconventional SLC genes and other L chains is that VpreB-λ5 gene rearrangement is not required for SLC expression.

The H-Chain Locus

The H-chain locus, on chromosome 14q32.33, is considerably more complex than the κ and λ loci. There are approximately 80 V_H gene segments near the telomere of the long arm of chromosome 14.[29] Of these, approximately 39 are functional and can be grouped into 7 different families of related gene segments. Adjacent to the most centromeric V_H, V6-1, are 27 D_H (D for diversity) gene segments (see Fig. 4.6) and 6 J_H gene segments. Each V_H and J_H gene segment is associated with a two-turn RSS, which prevents direct V→J joining. A pair of one-turn RSSs flanks each D_H. Recombination begins with the joining of a D_H to a J_H gene segment, followed by the joining of a V_H element to the amino terminal end of the DJ intermediate. The V_H gene segment contains FR1, -2, and -3, CDR1 and -2, and the amino terminal portion of CDR3; the D_H gene segment forms the middle of CDR3; and the J_H element contains the carboxy terminus of CDR3 and FR4 in its entirety (see Fig. 4.1). Random assortment of one of approximately 50 active V_H and one of 27 D_H with one of the 6 J_H gene segments can generate up to 104 different VDJ combinations (Fig. 4.7).

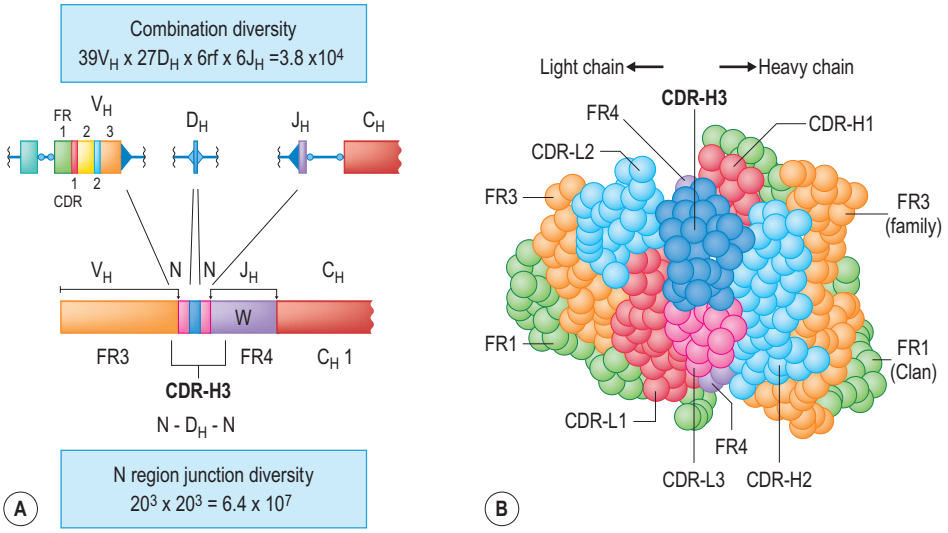


FIG. 4.7 The Antigen-Binding Site Is the Product of a Nested Gradient of Diversity. (A) VDJ rearrangement yields 38,000 different combinations. The CDR-H3 sequence contains both germline V, D, and J sequence and non-germline encoded N nucleotides. The addition of nine N nucleotides on either side of the D gene segment yields 64 million different combinations. (B) The antigen-binding site is created by the juxtaposition of the three complementarity-determining regions (CDRs) of the H chain and the three CDRs of the light chain. The view is looking into the binding site as an antigen would see the CDRs. The V_H domain is on the right side. The central location of CDR-H3, which due to N addition is the focus for repertoire diversity, is readily apparent. *FR*, Framework region.

Although combinatorial joining of individual V, D, and J gene segments maximizes germline-encoded diversity, the major source of variation in the preimmune repertoire is focused on the CDR-H3 interval which is created by VDJ joining (see Fig. 4.7). First, D_H gene segments can rearrange by either inversion or deletion and each D_H can be spliced and translated in each of the three potential reading frames. Thus, each D_H gene segment has the potential to encode six different peptide fragments. Second, the terminus of each rearranging gene segment can undergo a loss of one or more nucleotides during the recombination process. Third, the rearrangement process creates a hairpin ligation between the 5′ and 3′ termini of the rearranging gene segment. Nicking to resolve the hairpin structure leaves a 3′ overhang that creates a palindromic extension, termed a P junction. Fourth, non-germline-encoded nucleotides (N regions) can be used to replace or add to the original germline sequence. Every codon added by N-region addition increases the potential diversity of the repertoire 20-fold. N regions can be inserted both between the V and the D, and between the D and the J. Together, the imprecision of the joining process and variation in the extent of N addition permits generation of CDR-H3's of varying length and structure. As a result, more than 10^{10} different H-chain VDJ junctions, or CDR-H3's, can be generated at the time of gene segment rearrangement. Together, somatic variation in CDR3, combinatorial rearrangement of individual gene segments, and combinatorial association between different L and H chains yields a potential preimmune antibody repertoire of greater than 10^{16} different Igs.

Class Switch Recombination

Located downstream of the VDJ loci are nine functional C_H gene segments (see Fig. 4.7). Each C_H contains a series of exons, each encoding a separate domain, hinge, or terminus. All C_H genes can undergo alternative splicing to generate two different types of carboxy termini: either a membrane terminus that anchors Ig on the B-lymphocyte surface, or a secreted terminus that occurs in the soluble form of the Ig. With the exception of $C_H1\delta$, each C_H1 constant region is preceded by both an exon that cannot be translated (an I exon) and a region of repetitive DNA, termed the switch (S). Through recombination between the Cμ switch region and one of the switch regions of the seven other H-chain constant regions (a process termed *class switching* or *class switch recombination*), the same VDJ heavy-chain variable domain can be juxtaposed to any of the H-chain classes (Fig. 4.8).[30] Thus, the system can tailor both the receptor and the effector ends of the antibody molecule to meet a specific need.

Somatic Hypermutation

A final mechanism of Ig diversity is engaged only after exposure to antigen.[30] With T-cell help, the variable domain genes of germinal center lymphocytes undergo *somatic hypermutation* at a rate of up to 10^{-3} changes per base pair per cell cycle. SHM is correlated with transcription of the locus and current studies suggest that at least two separate mechanisms are involved. The first mechanism targets mutation hot spots with the RGYW (purine/G/pyrimidine/A) motif and the second mechanism incorporates an error-prone DNA synthesis that can lead to a nucleotide mismatch between the original template and the mutated DNA strand. SHM allows affinity maturation of the antibody repertoire in response to repeated immunization or exposure to antigen as B cells bearing receptors that have mutated to higher affinity for the cognate antigenic epitope are preferentially stimulated to proliferate, especially under conditions of limiting antigen concentration.

Activation-Induced Cytidine Deaminase

Activation-induced cytidine deaminase (AID) plays a key role in both CSR and SHM.[31] AID is a single-strand DNA (ssDNA) cytidine deaminase that can be expressed in activated germinal center B cells. Both SHM and CSR require transcription.

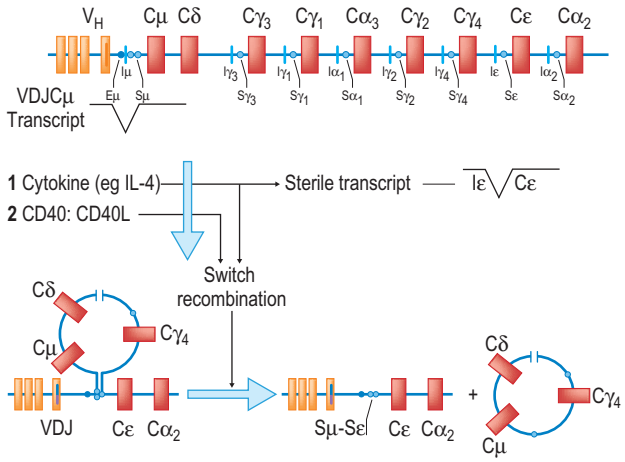


FIG. 4.8 Immunoglobulin H-Chain Class Switching. The molecular events involved in switching from expression of one class of immunoglobulin to another are depicted. At the top is the gene organization during μ-chain synthesis. At the bottom a class switch recombination event has resulted in the deletion of the intervening DNA. Exposure to the appropriate cytokine or T-cell:B-cell interaction through the CD40:CD40L pathway results in activation of the I exon that yields a sterile ε transcript (Iε-Cε) (see Chapter 7). The CD40:CD40L interaction is necessary for the subsequent replacement of Cμ by another constant gene (in this case, Cε). The S loci indicate switch-specific recombination signals.

Transcription helps target AID to the requisite chromosomal location, and also contributes to formation of requisite ssDNA substrates. For example, transcription of an Ig V domain or of the switch region upstream of the C_H1 domain opens the DNA helix to generate ssDNA that can then be deaminated by AID to form mismatched dU/dG DNA base pairs. Both CSR and SHM then co-opt the activities of normal cellular base excision repair (BER) and mismatch repair (MMR) to convert AID cytidine deamination lesions to mutations and/or double strand breaks. The BER protein uracil DNA glycosylase (UNG) removes the mismatched dU base, creating an abasic site. Cleavage of the DNA backbone at this abasic site by an apurinic/apyridimic (AP) endonuclease, generates a ssDNA nick adjacent to the abasic site. The nick is then processed to a single-nucleotide gap. The gap is filled in by DNA polymerase β; and then sealed by DNA ligase 1 or DNA ligase 3. The MMR proteins MSH2 and MSH6 can also bind and process the dU:dG mismatch. Deficiencies of AID, UNG underlie some forms of the hyper-IgM syndrome (Chapter 33). UNG and MMR double deficiency ablates CSR. It also eliminates both C/G transversion mutations and spreading of mutations, leaving only C/G transition mutations.

The benefits of diversity created by AID are balanced by the tendency of AID to target non-Ig genes. AID can create clusters of mutations in a number of genes, including BCL6, CD95, CD79A, CD79B, PIM1, MYC, RHOH, and paired box 5 (PAX5).[32] The process is termed *kataegis*. These mutations clusters can contribute to the development of lymphoproliferative disorders.

Diversity and Constraints on Immunoglobulin Sequence

In theory, combinatorial rearrangement of V(D)J gene segments, combinatorial association of H and L chains, flexibility in the site of gene segment joining, N-region addition, P junctions, hypermutation, and class switching can create an antibody repertoire, the diversity of which is limited only by the total number of B cells in circulation at any one time. In practice, constraints and biases on both the structure and sequence of the antibody repertoire are apparent.

The representation of individual V gene elements is nonrandom. Among Vκ and V_H elements, half of the potentially functional V gene elements contribute minimally to the expressed repertoire. Among Vλ elements these restrictions are even greater, with three gene segments contributing to half of the expressed repertoire.

Particular patterns of amino acid composition in the sequences of the V domains create predictable canonical structures for several of the hypervariable regions. In κ chains, CDR2 is found in a single canonical structure, whereas four structures are possible for CDR1.[33] In the H chain, most germline CDR1 and CDR2 elements encode one of three or one of five distinct canonical structures, respectively.[34] Preservation of these key amino acids during affinity maturation tends to maintain the canonical structure of CDR1 and CDR2 even while they are undergoing SHM.

The enhanced sequence diversity of the CDR3 region is mirrored by its structural diversity. Few canonical structures have been defined for the H-chain CDR3, and even in κ chains 30% of the L-chain CDR3 can be quite variable. However, at the sequence level there is a preference for tyrosine and glycine residues and a bias against the use of highly charged or hydrophobic amino acids in the H-chain CDR3, which reflects preferential use of only one of the six potential D_H reading frames, natural selection of reading frame content, and selection during development.[35]

The T-Cell Receptor αδ-Chain Locus

The α and δ loci are located on chromosome 14q11-12. This region is unusual in that the gene segments encoding the two different TCR chains are actually intermixed (Fig. 4.9). There are 38 to 40 Vα, 5 Vα/Vδ, no Dα, and 50 Jα functional gene segments, as well as one Cα gene.[36]

The δ locus lies between the Vα and Jα gene segments. There are three committed Vδ, five Vα/Vδ, three Dδ, and three Jδ gene segments, as well as one Cδ gene. Vδ3 lies 3′ of Cδ, and thus must rearrange by inversion. Although V-region use by α and δ chains is largely independent of one another, this unusual gene organization is accompanied by sharing of five V gene segments. For example, Vδ1 can rearrange either to Dδ/Jδ or to Jα elements, and thus can serve as the V region for both γδ and αβ TCRs.

In the large majority of αβ T cells analyzed, the α chain on both chromosomes has rearranged. This occurs by the rearrangement of the 5′ RSS δRec to a pseudo-J segment, ΨJα, at the beginning of the Jα cluster (see Fig. 4.9) as well as by the subsequent rearrangement of Vα to Jα on both chromosomes. Both types of rearrangement delete all of the Dδ, Jδ, and Cδ genes, thus preventing co-expression of αβ and γδ TCRs.

The T-Cell Receptor β-Chain Locus

The β locus is positioned at chromosome 7q35.44. It contains 40 to 48 functional Vβ genes, 2 Dβ, 2 Jβ clusters, each containing 6 or 7 gene segments, and 2 Cβ genes (see Fig. 4.9). There is one Vβ immediately downstream of Cβ2, which rearranges by inversion. Each Cβ is preceded by its own Dβ–Jβ cluster. There is no apparent preference for Vβ gene rearrangement to either Dβ–Jβ cluster. Dβ1 can rearrange to the Jβ1 cluster or the Dβ2–Jβ2 cluster. Dβ2 can only rearrange to Jβ2 gene segments.

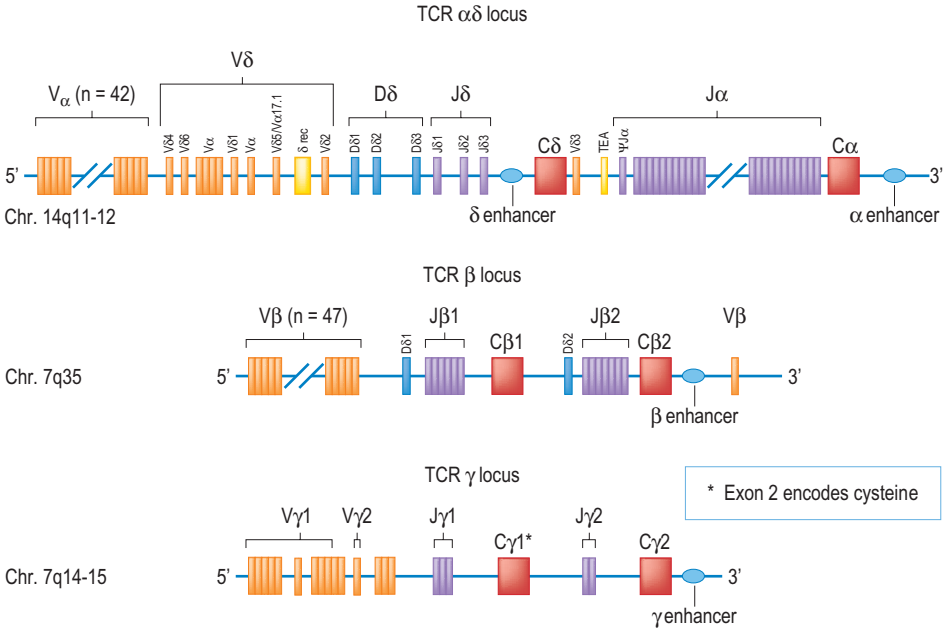


FIG. 4.9 Chromosomal Organization of the T-Cell Receptor αδ, β, and γ Gene Clusters. Typical numbers of functional gene segments are shown. These maps are not drawn to scale. *TCR*, T-cell receptor.

The two Cβ segments differ by only six amino acids and are functionally indistinguishable from each other.

The T-Cell Receptor γ-Chain Locus

The γ locus is located at chromosome 7p14-15. There are four to six functional Vγ region segments intermixed with pseudogenes, no Dγ, and two J clusters with a total of five J segments. Each J cluster is 5′ to its C region (see Fig. 4.9). The Vγ segments have been divided into six families, although only Vγ1 (nine members, five of them functional) and Vγ2 (one member) encode functional proteins. The number of Cγ gene exons varies: Cγ1 has three, while there are two alleles of Cγ2 that have four and five, respectively. The first Cγ exon encodes most of the extracellular portion of this region. The last Cγ exon encodes the intracytoplasmic portion of the molecule. The middle exon(s) (one for Cγ1, two or three for Cγ2) encode the connecting piece, which does (Cγ1), or does not (Cγ2), include a cysteine. Since this cysteine can form a disulfide bond with another cysteine in the δ chain, TCRs using Cγ1 contain a covalently linked γ–δ pair, while TCRs using Cγ2 do not. The nomenclature of the human γ locus differs between laboratories and reports and is extensively cross-referenced on the IMGT website (http://www.imgt.org).

Allelic Exclusion

Because of the inherently imprecise nature of coding joints, only one in three V(D)J Ig or TCR rearrangements will be in-frame and capable of creating a functional protein. Theoretically, one in nine cells might be expected to express two different Ig or TCR chains. However, almost all B cells express the functional products of only one IgH allele and one IgL allele, and mature αβ T cells express only one functional TCRβ gene. The process of limiting the number of receptors expressed by an individual cell is known as *allelic exclusion*.

The mechanisms that ensure monoallelic expression are regulated at the level of gene rearrangement. Mechanisms that have been shown to contribute to allelic exclusion include asynchronous replication of the two alleles, with rearrangement occurring at the allele that replicates early; localization of the active allele to a more central, euchromatic region of the nucleus; and DNA demethylation of the active allele. Once a functional V domain has been generated, rearrangement terminates with the expression of a membrane-bound Ig (mIg) or TCR product capable of transducing a signal. In pre-B cells, a functional μ H chain associates with the surrogate light chain to form the pre-BCR. Similarly, in developing T-cell progenitors a productive TCR β chain associates with pre-Tα to form the pre-TCR. These preliminary antigen receptors signal to shut down RAG expression, promote cell division, and limit the accessibility of the IgH and TCRβ loci to further rearrangement while promoting the accessibility of the IgL and TCRα loci, respectively.

In pre-B cells, the κ locus is the first to rearrange, with λ rearrangement occurring in cells that have failed to produce a proper κ chain. Surface expression of an acceptable membrane-bound IgM BCR invokes the mechanism of allelic exclusion among the L-chain loci, termed *isotypic exclusion*, and promotes further maturation of the B cell.

Productive TCRα rearrangement in $CD4^+CD8^+$ T-cell progenitors allows the expression of a functional TCR αβ heterodimer (Chapter 9). Unlike IgH and TCRβ; TCRα does not undergo allelic exclusion at the level of gene rearrangement. Instead, in cells that express two functional TCRα alleles, one of the two alleles tends to preferentially pair with the one functional TCRβ chain. This is termed *phenotypic allelic exclusion*.

Allelic exclusion can be overcome by selection pressures. Cells that express self-reactive antigen receptors can downregulate IgH or TCR expression and reactivate gene rearrangement to replace one of the two offending chains. This process, termed *receptor editing*, occurs most often in the IgL or TCRα loci, whose gene structures lend themselves to repeated rearrangement. The V_H in the H chain can also be replaced by rearrangement to a cryptic RSS at the terminus of the V_H gene segment.

KEY CONCEPTS

B-Cell Receptor and Co-Receptors

- The BCR–antigen complex consists of a mIg that is responsible for antigen recognition and an Igα/β heterodimer that is responsible for transducing the recognition signal into the cell.
- BCR engagement leads to the phosphorylation of tyrosines in the Igα/β *I*mmunoreceptor *T*yrosine-based *A*ctivation Motifs (ITAMs). This signal is then transmitted to one or more other intracellular signaling pathways.
- Recognition of antigen by B lymphocytes can also involve binding of antigen complexed with C3d and IgG to additional B-cell co-receptors.
- Binding of complexed antigen by individual co-receptors can lead to either positive or negative signals, each of which can influence the ultimate outcome of an antigen-B lymphocyte interaction.
- Deficiency of the components of the BCR–antigen complex impairs B-cell development and can lead to agammaglobulinemia.

B-CELL RECEPTOR COMPLEX: STRUCTURE AND FUNCTION

Although the ability of surface Ig to recognize antigen was appreciated very early, the mechanism by which mIg transmitted an antigen recognition event to the cell took longer to understand. The predominant isotypes expressed on the surface of mature B cells, mIgM and mIgD, contain only three amino acid residues exposed to the cytoplasm. It was thought unlikely that these Ig heavy chains could function as signal transduction molecules by themselves. Subsequently, it was shown that all membrane Ig isotypes associated noncovalently with a heterodimeric complex consisting of two transmembrane proteins, Igα (CD79a) and Igβ (CD79b), each of which is capable of transducing signals into the cell (Table 4.3).

TABLE 4.3 The B-Cell Receptor and Its Co-Receptor Molecules

Molecule	M_r	Chromosome	Function
BCR			
mIgM (μ_2L_2)	180,000	14 (IgH; 14q.32) 2 (Igκ; 2p12) 22 (Igλ; 22q11.2)	Antigen recognition
Igα (CD79a)	47,000	19 (19q13.2)	Signal transducer
Igβ (CD79b)	37,000	17 (17q23)	Signal transducer
Co-Receptors			
CD21	140,000	1 (1q32)	Activating co-receptor Ligand for C3d, EBV, CD23
CD19	95,000	16 (16p11.2)	Activating co-receptor Signal transducer
FcγRIIB (CD32)	40,000	1 (1q23-24)	Inhibitory co-receptor Low-affinity receptor for IgG
CD22	140,000	19 (19q13.1)	Inhibitory co-receptor Adhesion molecule Signal transducer

BCR, B-cell receptor; *EBV*, Epstein–Barr virus; *IgG*, immunoglobulin G; *mIgM*, membrane-bound immunoglobulin M.

Membrane-Bound Immunoglobulin

Igs mediate their effector functions as secreted products of plasma cells. However, Igs also serve as the membrane-bound antigen recognition component of the BCR complex. Although all Ig classes can be expressed at the B-cell surface, the vast majority of circulating mature B cells co-express membrane-bound IgM and IgD. Appropriate activation of a naïve IgM and IgD expressing B cell leads to plasma cell differentiation and antibody secretion. Membrane-bound IgM and IgD are the product of alternative splicing of the Ig transcript at the 3′, or carboxy, terminus of the heavy chain (Fig. 4.10). The two membrane exons encode the transmembrane hydrophobic stretch of amino acids and an evolutionarily conserved cytoplasmic tail encoding lysine, valine, and lysine.

Signal Transduction and the Immunoglobulin-α/β (CD79a/CD79b) Heterodimer

The heterodimeric signal transduction component of the BCR complex that associates with membrane Ig has been designated CD79. It is composed of an Igα (CD79a) and Igβ (CD79b) heterodimer. CD79 is responsible for transporting mIg to the cell surface and for transducing BCR signals into the cell.[37]

CD79a/Igα is encoded by *CD79a/MB-1* (chromosome 19q13.2) as a 226-amino acid glycoprotein of approximately 47 kDa. The exact molecular weight depends on the extent of glycosylation. *CD79b/B29* (chromosome 17q23) encodes CD79b/Igβ, which is a 229-amino acid glycoprotein of approximately 37 kDa. *CD79a* and *CD79b* share an exon–intron structure, which is similar to that of the genes that encode the CD3 TCR co-receptor molecules. These similarities suggest that both BCR and TCR co-receptors are the progeny of a common ancestral gene. Igα and Igβ both contain a single IgSF Ig domain (111 residue C-type for Igα and 129 residue V-type for Igβ). Each also contains a highly conserved transmembrane domain and a 61- (Igα) or 48- (Igβ) amino acid cytoplasmic tail that also exhibits striking amino acid evolutionary conservation.

Igα and Igβ are expressed by the earliest committed B-cell progenitors prior to Igμ H-chain rearrangement. The CD79 heterodimer has been observed on the surface of early B-cell progenitors in the absence of Igμ, although neither protein is required for progenitors to commit to the B-cell lineage.[38] Later in development, Igα and Igβ are co-expressed together with Ig

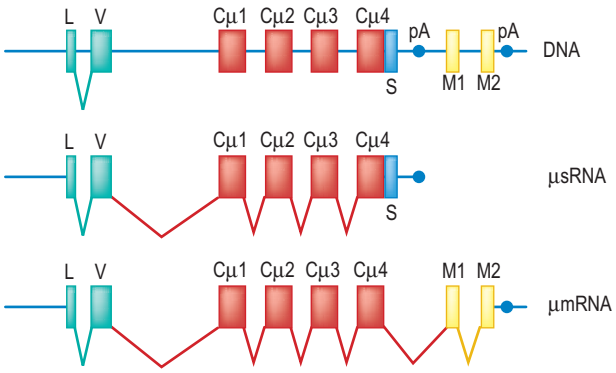


FIG. 4.10 Membrane and Secretory Immunoglobulin M Are Created by Alternative Splicing. Alternative splicing of the Cm carboxy-terminal exons results in mRNA transcripts encoding either secreted IgM (μsRNA) or membrane-bound IgM (μmRNA).

of all isotypes on the surface of B cells as a mature BCR complex.[39] The CD79 proteins are specific to the B lineage and are expressed throughout B lymphopoiesis. This has led to their use as markers for the identification of B-cell neoplasms.[40]

The signaling capacity of both Igα and Igβ resides within an *ITAM* that has the consensus sequence of D/IxxYxxL(x)7YxxL, where x is any amino acid. Similar ITAMs are also found within the cytoplasmic domain of the molecules that associate with, and signal for, the T-cell antigen receptor (CD3) and certain Fc receptors (Chapter 18). The phosphorylation of both tyrosines in both Igα/β ITAMs is considered an obligate initial step in the propagation of antigen receptor engagement to the cell nucleus.[37,41]

Tyrosine-phosphorylated ITAMs serve as efficient binding sites for Src homology 2 (SH2) domains, which are present within a large number of cytosolic signaling molecules. Whether Igα and Igβ make qualitatively different contributions towards BCR signaling or are functionally redundant remains unclear, as evidence exists to support both views. Moreover, the high degree of evolutionary conservation within the non-ITAM portion of the cytoplasmic domains suggests additional, as yet uncharacterized, signaling roles for the cytoplasmic tails of these molecules over and above positive signaling via the ITAMs.

Igα and Igβ are covalently associated by a disulfide bridge via cysteine residues that exist within the IgSF extracellular domains of both molecules. The association of the Igα/β heterodimer with mIg occurs though interaction within the transmembrane domains of these proteins.[39] The core BCR complex consists of a single Ig molecule associated with a single Igα/β heterodimer (H_2L_2/Igα/Igβ) (Fig. 4.11). A current model for the initiation of signals originating from the BCR (see Fig. 4.11) proposes that antigens induce the clustering of BCR complexes, increasing their local density. The increase in density leads to the transfer of phosphate groups to the tyrosine residues of the Ig-α/β ITAM motifs.[37,41]

Src-family tyrosine kinases, of which LYN, FYN, and BLK are most often implicated, are believed to be responsible for ITAM phosphorylation upon aggregation of Igα/β. They have been shown to physically associate with the heterodimer. It has been suggested that only a fraction of Src-family tyrosine kinases are associated with the Igα/β heterodimer and, upon aggregation, transphosphorylate juxtaposed heterodimers. However, the exact mechanism by which Igα/β undergoes initial tyrosine phosphorylation after antigen engagement remains uncertain. Phosphorylated ITAMs subsequently serve as high-affinity docking sites for cytosolic effector molecules that harbor SH2 domains. The recruitment of the SYK tyrosine kinase, via its tandem SH2 domains, to doubly phosphorylated Igα/β ITAMs then appears to propagate a BCR-mediated signal. Association of SYK with the BCR leads to its subsequent tyrosine phosphorylation by either Src-family or other Syk tyrosine kinases, further increasing kinase activity. Together, the concerted actions of the Syk and Src-family protein tyrosine kinases activate a variety of intracellular signaling pathways that can lead to the proliferation, differentiation, or death of the cell.

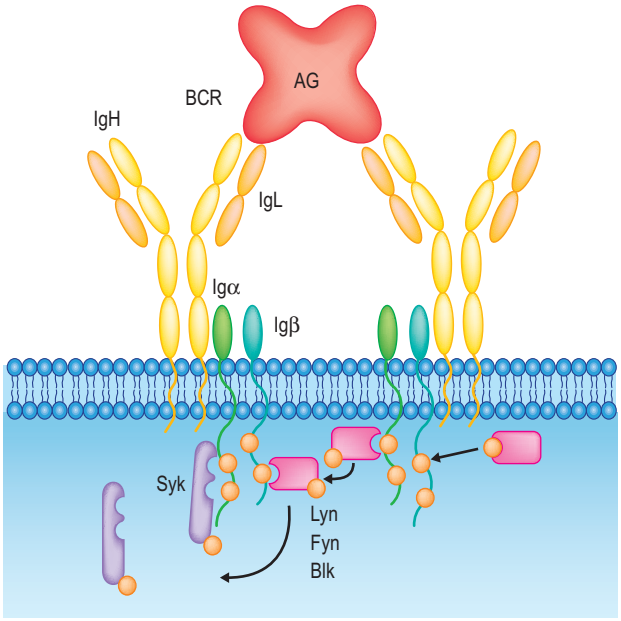


FIG. 4.11 The B-Cell Receptor Core Complex. The B-cell receptor (*BCR*) core complex can be divided into an antigen recognition unit fulfilled by mIgM and a noncovalently associated signal transduction unit composed of the Igα/β (CD79ab) heterodimer. Antigen engagement of membrane-bound immunoglobulin M oligomerizes the BCR, allowing preassociated Src-family protein tyrosine kinases to phosphorylate neighboring Igαβ *I*mmunoreceptor *T*yrosine-based *A*ctivation Motif (ITAM) tyrosines. This promotes association of the SYK tyrosine kinase with the tyrosine phosphorylated ITAMs, allowing SYK to become a substrate for other Syk or Src-family tyrosine kinases, leading to its activation.

Clinical Consequences of Disruptions in B-Cell Receptor Signaling

Both the development of B lymphocytes and the maintenance of the mature antigen-responsive B-cell pool demand the presence of an intact BCR and its downstream signaling pathway(s). Disruption of these pathways can present clinically with hypogammaglobulinemia and an absence of B cells.

The most common such genetic lesions is BTK deficiency, which is an X-linked trait (Chapter 33). BTK plays an important role in BCR signaling both during development and in response to antigen. Loss-of-function mutations in BTK results in the arrest of human B-cell development at the pre-B-cell stage.

BTK is intact in approximately 10% to 15% of patients with hypogammaglobulinemia and absence of B cells. Mouse models where BCR components or signaling pathways have been disrupted by targeted mutagenesis have provided insight into the basis of these atypical hypogammaglobulinemia disorders.[37] These studies have shown that an inability to express either a functional μ IgH chain, Igα, Igβ, or the signaling adaptor molecule, BLNK, lead to an early, severe arrest in B lymphopoiesis, with subsequent agammaglobulinemia.[42] Together, these experimental findings highlight the central role of the BCR in the generation and function of B lymphocytes. Thus, mutations in any component of the antigen receptor complex or immediate downstream effectors have the potential to disrupt B-cell development and create an agammaglobulinemic state.

Besides its important role in the maturation, differentiation, and survival of B lymphocytes, the B-cell antigen receptor is responsible for initiating the humoral response to foreign antigen. Some of the variables that can influence the

ultimate outcome of BCR–antigen interaction include the nature of the foreign antigen, the mode of activation, the developmental stage of the B cell, and the microenvironment in which antigen encounter occurs. Exactly how these variables ultimately result in the differential activation of diverse intracellular signaling pathways with fundamentally divergent outcomes is under study. Emerging from these studies is an appreciation of the role of BCR co-receptors, which have been shown to be capable of modulating antigen receptor signaling in response to antigen.

B-Cell Receptor Co-Receptors

The initiation of a humoral immune response results from antigen interaction with the antigen receptors on mature peripheral lymphocytes. However, the manner in which mature B and T lymphocytes recognize antigen is fundamentally different (Chapter 6). Surface Ig, as a component of the BCR on B lymphocytes, typically recognizes an antigenic epitope in its native three-dimensional configuration, which, upon engagement with mIg, is capable of transmitting a signal to the cell interior. In contrast, the antigen receptor expressed by T lymphocytes typically recognizes an antigen-derived peptide associated with an appropriate MHC structure (Chapter 5). Further, in order for this T-cell recognition event to be productive, a CD4 or CD8 co-receptor must also bind to the MHC structure presenting the foreign antigen.

Antigen recognition by the BCR on B lymphocytes is also influenced by co-receptors present on mature B cells (see Table 4.3). In this case the co-receptors may also recognize antigen, but only in a form that has been modified by other components of the immune system, as described below. In general, these co-receptors and co-receptor complexes can be divided into those that regulate BCR signaling in a positive manner and those that regulate in a negative manner. Thus, the ultimate outcome of signaling via the BCR depends not only on the signals transduced via the Igα/β heterodimer, but also how these signals are perceived by the cell in association with the signals propagated by the various co-receptors that are concomitantly engaged.

Co-Receptors That Positively Regulate B-Cell Receptor Signaling

CD21

Mature B lymphocytes express two receptors for complement C3 components, CD35 (CR1) and CD21 (CR2) (Chapter 40). Of these, CD21 fulfills the requirements of a BCR co-receptor (vide infra). The expression of CD21 is restricted to mature B cells and follicular dendritic cells, whereas CD35 is also found on erythrocytes, monocytes, and granulocytes. CD21 is a 140-kDa surface glycoprotein encoded by the CR2 locus on chromosome 1q32 (see Table 4.3). Expression of CD21 begins at approximately the same time as IgD during B lymphopoiesis (Chapter 7). CD21 is subsequently expressed on all mature B cells until terminal differentiation. Within the mature population, marginal zone B cells express higher levels relative to follicular B cells. The extracellular domain of CD21 is composed of 15 to 16 short consensus regions (SCRs), each composed of 60 to 70 amino acids, and a relatively short 34-amino acid cytoplasmic tail. The two-amino terminal SCRs constitute the region that interacts with one of the third complement component (C3) cleavage products, iC3b, C3d, g, and C3d (Chapter 40).[43]

CD21 is a receptor for Epstein–Barr virus (EBV), which similarly binds the two N-terminal SCRs via its major envelope glycoprotein gp350/220. CD21, through its oligosaccharide chains, also binds CD23, the low-affinity IgE receptor (FcεRII). Whereas EBV utilization of CD21 for cell entry has clear physiologic consequences in terms of infection, B-cell immortalization, and the potential for oncogenesis, the in vivo relevance of any CD21 to CD23 interaction remains unclear.

CD19

CD19 is an IgSF surface glycoprotein of 95 kDa that is expressed from the earliest stages of B-cell development until plasma cell terminal differentiation, when its expression is lost.[44] Follicular dendritic cells also express CD19. *CD19* maps to chromosome 16p11.2, where it encodes a 540-amino acid protein with two extracellular C-type IgSF domains as well as a large, approximately 240-residue, cytoplasmic tail that exhibits extensive conservation between mouse and human. This relatively large cytoplasmic domain includes nine conserved tyrosine residues, which, upon phosphorylation, serve as docking sites for other SH2-containing effector molecules. The signaling capacity of CD19 has been shown to result from tyrosine phosphorylation, which occurs upon engagement of the BCR, CD19 or, optimally, by co-ligation of CD19 and IgM. Known signaling effector molecules that have been identified in association with tyrosine-phosphorylated CD19 include the LYN and FYN protein tyrosine kinases, the Rho-family guanine nucleotide exchange factor, VAV, and phosphatidyl inositol 3-kinase.[44] Although specific ligands for CD19 have been proposed, the physiological relevance of CD19 engagement by putative ligands has not been demonstrated.

In vitro studies using mAbs directed against CD21 or CD19 provided initial evidence that these B-cell surface antigens could influence mIg-mediated signaling.[45,46] Genetic deficiencies of CD21 (CVID7) or CD19 (CVID3) promote the development of common variable immune deficiency, which is characterized by hypogammaglobulinemia (Chapter 33). In mice, CD21 and CD19 deficiency demonstrate impaired antibody response to T-dependent antigens. The paucity of $CD5^+$ B cells in CD19-deficient mice suggests a role for this molecule in the generation and maintenance of the B1 lineage of B cells (Chapter 7). CD19 is expressed from the earliest stages of B-cell ontogeny in both mice and humans and, accordingly, a signaling function for CD19 in B lymphopoiesis has been demonstrated.[47]

CD21–CD19 Co-Receptor Complex

A mechanism by which these molecules could augment BCR-mediated signaling was provided by the identification of a CD21–CD19 co-receptor complex on mature B cells that also includes CD81 (Fig. 4.12). CD81, also known as TAPA-1, is a 26-kDa tetraspan molecule widely expressed on a number of cell types, including lymphocytes. The CD21–CD19 co-receptor model predicted that, as a result of complement activation, C3d would be deposited on an antigen, thereby providing a bridge by which a CD21–CD19 receptor complex could associate with mIgM and the BCR complex.[44,46] Clustering of CD19 close to the BCR by the C3d–antigen complex would effectively recruit the signal transduction effector molecules associated with CD19 to the Igα/β heterodimer. As a consequence, the CD19-associated LYN and FYN tyrosine kinases, VAV, and PI3-kinase signaling effector molecules would be in a position to exert their activities on the Igα/β heterodimer-mediated signaling events initiated by antigen engagement of mIgM.

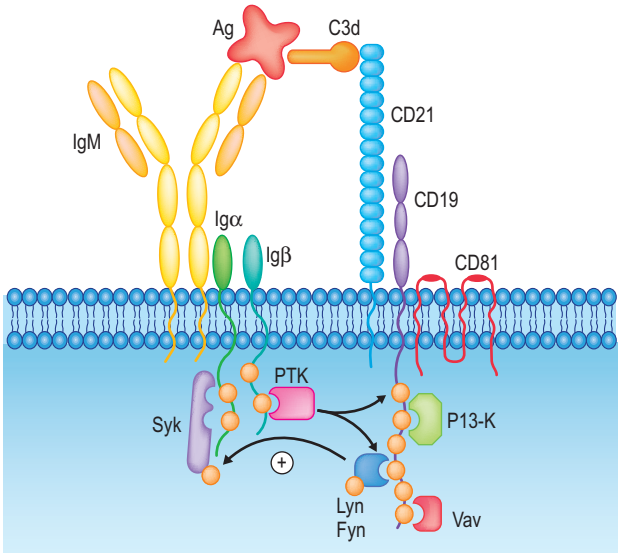


FIG. 4.12 Proposed Mechanisms for the Augmentation of B-Cell Receptor Signaling by the CD21–CD19 Co-Receptor. Co-ligation of the B-cell receptor (BCR) and CD21–CD19 complex by C3d–antigen complex allows a CD79-associated Src-family tyrosine kinase to phosphorylate tyrosine residues within the CD19 cytoplasmic domain. Subsequently, tyrosine-phosphorylated CD19 effectively recruits key SH2-containing signaling molecules to the BCR complex, allowing the initial BCR-mediated signal to quickly disseminate along different intracellular signaling pathways. *Ag,* antigen; *PI3-K,* phosphatidyl inositol 3-kinase; *PTK,* protein tyrosine kinase.

Strong support for CD21–CD19 co-receptor physiological function in BCR signaling was subsequently provided by experiments using a murine model of immune response. Immunization with an antigen covalently attached to C3d dramatically reduced the signaling threshold necessary for antigen to elicit an immune response.[48] Antigen bearing either two or three copies of C3d was, respectively, 1000 and 10,000 times more immunogenic than antigen alone. Thus, the CD21–CD19 co-receptor complex provides a link between the innate and adaptive immune responses. In vivo, CD19-deficient mice appear to have more severely affected T-dependent immune responses than do CD21-deficient animals, suggesting alternative roles for CD19 in regulating BCR signals beyond the CD21–CD19 co-receptor complex.

Co-Receptors That Negatively Regulate B-Cell Receptor Signaling

FcγRIIB

Among the several receptors for the Fc portion of Ig expressed by B cells, the Fc receptor for IgG, FcγRIIB (a member of the CD32 cluster), has an important role in negatively regulating BCR-mediated signal transduction.[49] FcγRIIB is a 40-kDa single-chain molecule that is encoded by single gene located on chromosome 1q23-24. Alternative splicing of different cytoplasmic exons permits expression of three isoforms. The extracellular domain of FcγRIIB is composed of two C-type IgSF domains that can bind with low affinity to IgG. All three FcγRIIB isoforms share a common cytoplasmic region that is important for negatively regulating activation signals delivered by associated surface receptors. The region within the cytoplasmic domain of FCγRIIB responsible for the inhibitory activity of this Fc receptor towards the BCR has been identified as a sequence that contains a tyrosine residue critical for its activity. In analogy to the ITAM, which provides an activation signal, this inhibitory sequence has been referred to as an *I*mmunoreceptor *T*yrosine-based *I*nhibitory Motif, or *ITIM*. The ITIM is carried by the canonical sequence of I/L/VxYxxI/V/L (where x is any amino acid). ITIMs are found in a number of other transmembrane structures, all of which share the ability to negatively regulate signaling by activating receptors.

The ability of passively administered soluble antibody to inhibit humoral responses has long been appreciated and was initially thought to occur by soluble antibody effectively masking all available antigen epitopes. The molecular mechanism accounting for this suppression is now known to be mediated by the binding of IgG to FcRγIIB and the subsequent recruitment of cytosolic phosphatases to the FcRγIIB ITIM upon tyrosine phosphorylation. Thus, the inhibitory effect of IgG on BCR-mediated B-cell activation is explained by the interaction of the FcγRIIB ITIM, and specifically associated phosphatases, with the BCR (Fig. 4.13). Co-ligation of the BCR and FcRγIIB by antigen–IgG complexes results in the tyrosine phosphorylation of the FcRγIIB ITIM, presumably by the BCR-associated tyrosine kinases. Phosphorylated FcRγIIB ITIMs then recruit two different SH2-containing phosphatases, SHIP and SHP-1, which function to remove phosphate groups from inositol lipids or tyrosines, respectively. Although both phosphatases can negatively regulate BCR-mediated signaling events, SHIP appears to be the most relevant phosphatase in FcRγIIB inhibition of BCR signaling (see Fig. 4.13). Thus, once the majority of antigen exists in immune complexes together with antigen-specific IgG, attenuation of an ongoing immune response occurs by the juxtaposition of FcRγIIB with the BCR.

CD22

CD22 is a 135- to 140-kDa transmembrane glycoprotein that is restricted in its expression to the B lineage.[50] CD22 expression is limited to the cytoplasm of progenitor and pre-B cells in early B-cell development. Expression on the surface of the B cell occurs concomitant with the appearance of surface, or membrane, IgD. Upon B-cell activation, CD22 expression is initially transiently upregulated and subsequently down-modulated upon terminal differentiation to Ig-secreting plasma cells. Although the onset of CD22 expression follows a similar pattern during murine B lymphopoiesis, it is not restricted to the cytoplasm in early B lymphopoiesis, but rather is expressed on the surface from the progenitor stage onward. The basis or function of CD22 intracellular retention in human B-cell development is not understood.

CD22 maps to chromosome 19q13.1 and encodes alternatively spliced forms of CD22, CD22α, and CD22β, of which the latter is the predominant species expressed by B cells. The CD22β isoform contains seven extracellular IgSF domains, of which all but one are of the C type. The single exception is the N-terminal domain, which is of the V type. CD22α lacks the IgSF third and fourth domains, although the significance of this minority alternatively spliced product remains unclear. The CD22 murine homolog has only been found as a full-length CD22β isoform. The extracellular domain of CD22 is homologous to the carcinoembryonic antigen subfamily of adhesion molecules, which includes the myelin-associated glycoprotein (MAG) and CD33. CD22 also functions

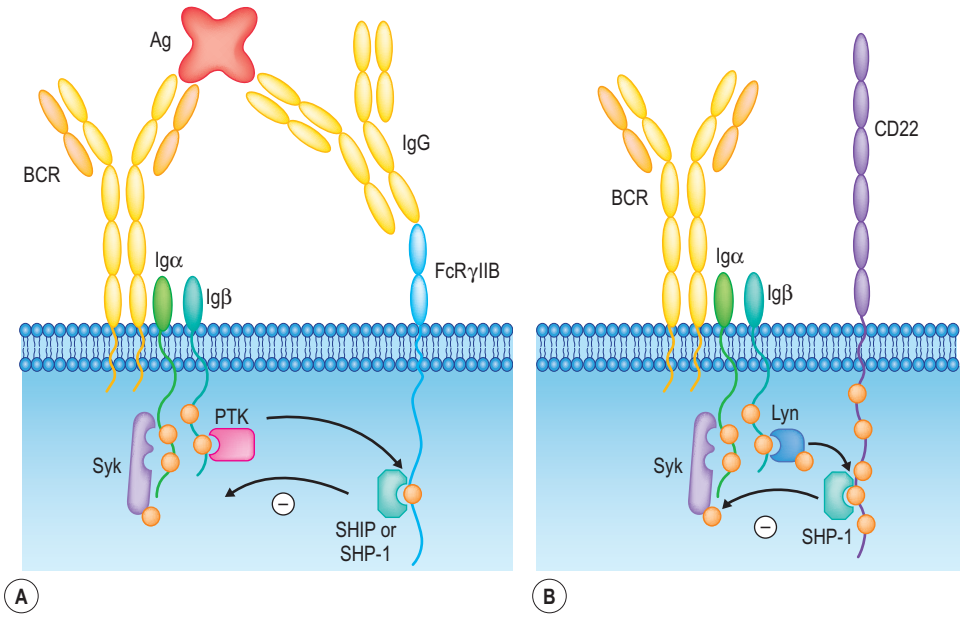


FIG. 4.13 Negative Regulation of B-Cell Receptor Signaling by FcγRIIB and CD22. (A) Soluble IgG–antigen immune complexes juxtapose the B-cell receptor (*BCR*) with FcγRIIB. The BCR-associated LYN tyrosine kinase subsequently tyrosine phosphorylates the FcγRIIB *I*mmunoreceptor *T*yrosine-based *I*nhibitory Motif (ITIM). In turn, this leads to the recruitment of the SH2-containing inositol phosphatase SHIP and tyrosine phosphatase SHP-1 to the phosphorylated FcRγIIB ITIM. Both of these phosphatases have demonstrable inhibitory activity on BCR-mediated signaling. Although SHIP is believed to be the major effector in the FcRγIIB-mediated inhibition of BCR signaling, the exact mechanism of its action in this context has not yet been elucidated. (B) CD22 associated with the BCR is tyrosine-phosphorylated upon antigen–BCR engagement. SH2-containing signaling molecules dock on tyrosine-phosphorylated residues, including the SHP-1 tyrosine phosphatase that can subsequently dephosphorylate signaling molecules previously activated by a membrane-bound immunoglobulin M-mediated signal.

as an adhesion molecule belonging to the Siglec subfamily of the Ig superfamily, whose members function as mammalian sialic acid-binding Ig-like lectins.[51] The two N-terminal IgSF domains have been shown to mediate adhesion to both B and T lymphocytes via the binding of structures carrying α2,6 sialic acids.

In addition to acting as an adhesion molecule, CD22 is also capable of modulating BCR signaling (see Fig. 4.13). A fraction of CD22 associates with the BCR, and CD22 is rapidly tyrosine-phosphorylated upon mIgM engagement. Tyrosine-phosphorylated CD22 associates with several SH2-containing signaling molecules, including the LYN and SYK tyrosine kinases, PI3-kinase, phospholipase C-γ and SHP-1. The 140-amino acid cytoplasmic domain of CD22 includes six conserved tyrosine residues. Three of these tyrosines are located within conserved consensus ITIM sequences and possess a demonstrable capacity to bind the SH2 domain of the SHP-1 phosphatase. The presence of the multiple ITIMs and association with SHP-1 indicated that CD22 might impinge on BCR signaling in a negative manner. Physiological evidence that CD22 could act as a co-receptor to negatively regulate mIgM signaling was provided by the generation of CD22-deficient mice by targeted mutagenesis.[51] CD22-deficient B cells exhibited hyperactive B-cell responses upon BCR triggering, and an increased incidence of serum autoantibodies. This suggests that B-cell tolerance is altered and B cells are more readily activated in the absence of this negative regulator of BCR signaling.

KEY CONCEPTS

T-Cell Receptor/CD3 Complex

- Cell-surface expression of the TCR heterodimers requires association with a complex of invariant proteins designated CD3.
- Each TCR/CD3 complex contains three CD3 dimers.
- Assembly of the TCR/CD3 complex involves interactions between TCR transmembrane basic residues and transmembrane acidic residues in each of the CD3 subunits.
- Signal transduction by the TCR involves the phosphorylation of ITAMs in the cytoplasmic domains of CD3 proteins.
- Phosphorylated CD3 ITAMs recruit and activate the ZAP-70 protein tyrosine kinase.
- Deficiency of CD3 proteins impairs T-cell development and can produce SCID.

THE T-CELL RECEPTOR/CD3 COMPLEX

The αβ and γδ TCR heterodimers, which are responsible for the recognition of specific antigen by T lymphocytes, associate with a complex of invariant proteins designated CD3. This association is necessary for TCR cell-surface expression and enables the TCR heterodimers, which have only short cytoplasmic domains, to couple to the intracellular signaling events that lead to the activation of T-cell effector function.[52] There are four CD3 proteins: γ, δ, ε, and ζ (Fig. 4.14).

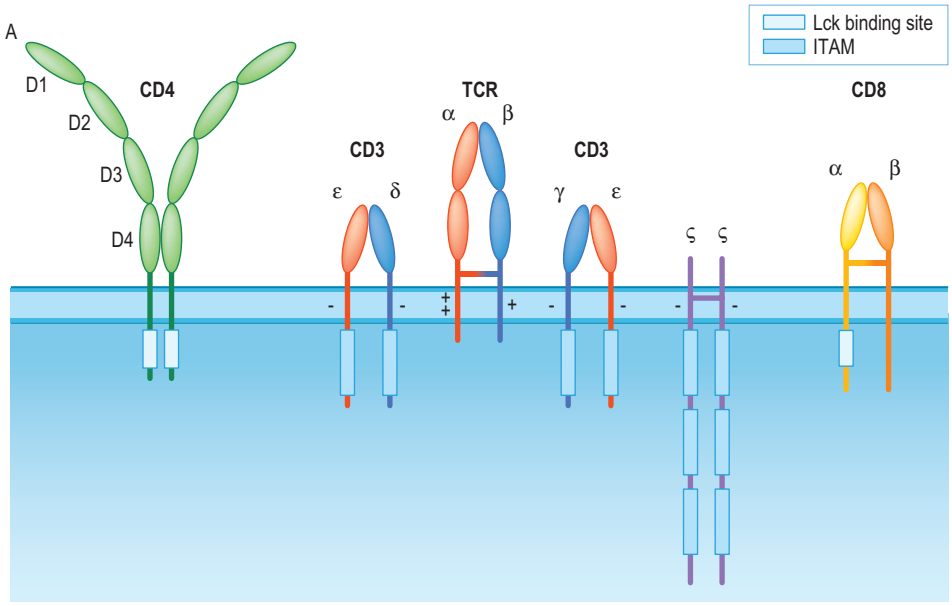


FIG. 4.14 Schematic Representation of the Human T-Cell Receptor and CD4 and CD8 Co-Receptors. IgSF domains are represented by ovals. The four extracellular domains of CD4 are labeled D1–D4. Basic (+) and acidic (–) transmembrane charged residues are indicated, as are known and predicted sites of disulfide bonds. For schematic simplicity the cytoplasmic domains of the CD3 chains are shown as extending into the cytoplasm. The cytoplasmic domains of CD3ε and CD3ζ are positively charged and likely are associated with the inner leaflet of the plasma membrane.

CD3 Proteins

CD3γ, CD3δ, and CD3ε are structurally similar, and the genes encoding them map to a locus in chromosome 11q23. The polypeptides range in size from 20 to 25 kDa. Each has an extracellular C-type IgSF domain, a transmembrane region that contains an acidic residue (aspartic acid in CD3δ and CD3ε, glutamic acid in CD3γ), and a cytoplasmic domain with a single ITAM. The cytoplasmic domain of CD3ε (but not of CD3δ or CD3γ) has a net positive charge and can bind to the negatively charged inner leaflet of the plasma membrane with its ITAM inserted into the lipid bilayer. The CD3 chains are present in the TCR/CD3 complex in the form of noncovalently linked CD3γε and CD3δε heterodimers; interactions between the extracellular IgSF domains lead to the formation of these CD3 heterodimers.

The 16-kDa CD3ζ differs substantially from the other CD3 proteins and is structurally homologous to the γ chain of the high-affinity IgE receptor (FcRγ chain). The extracellular domain of CD3ζ has only nine amino acids and is of unknown structure. As is the case with the other CD3 chains, the transmembrane region of CD3ζ contains an acidic residue (aspartic acid). The large cytoplasmic domain of CD3ζ has 3 ITAMs in tandem, which, like the ITAM of CD3, also associate with the inner leaflet of the plasma membrane.[16] CD3ζ is usually present in the TCR/CD3 complex in the form of disulfide-linked CD3ζζ homodimers that form through interactions within the transmembrane domain.

Stoichiometry of the T-Cell Receptor/CD3 Complex

The αβTCR/CD3 complex is univalent and consists of a single αβ TCR heterodimer together with three CD3 dimers: CD3γε, CD3δε, and CD3ζζ (see Fig. 4.14). The γδTCR/CD3 complex, in contrast, lacks CD3δ. On naïve T cells, this receptor complex contains two CD3γε heterodimers and one CD3ζζ homodimer. Following activation of γδT cells, the TCR/CD3 complex incorporates the FcRγ chain, either as a homodimer or as a heterodimer with CD3ζ.[16,52,53]

Assembly and Cell-Surface Expression of the T-Cell Receptor/CD3 Complex

Assembly begins with formation of the individual TCRαβ, CD3δε, and CD3γε heterodimers, processes that are driven by interactions between the extracellular domains of the pairing polypeptides. The subsequent higher-order assembly of the TCRαβ with the CD3 dimers depends upon interactions between the potentially charged residues within their transmembrane regions. As noted above, each of the CD3 subunits has a transmembrane acidic residue while the transmembrane domains of the αβ and γδ TCRs contain basic residues. Mutation of any of these transmembrane acidic or basic residues to neutral alanine impairs formation of the TCR/CD3 complex. TCRαβ appears to associate first with CD3δε and then with CD3γε. TCRα binds CD3δε, and TCRβ likely interacts with CD3γε. The incorporation of a CD3ζζ homodimer into the complex requires the prior formation of a TTCRαβ–CD3γε–CD3δε hexamer and involves interactions between the arginine residue in the transmembrane domain of TCRα and the two co-localized aspartic acids in the transmembrane domains of the CD3ζζ homodimer.[16,54]

Formation of the TCR/CD3 complex is tightly regulated. For example, when there are deficiencies of CD3γ, CD3δ, or CD3ε, TCRα and β are retained in the endoplasmic reticulum and are rapidly degraded. In the absence of CD3ζ, the TCRαβ–CD3γε CD3δε hexamer is exported to the Golgi but then is targeted to a lysosomal degradation pathway rather than the cell surface.[16,52–54]

A cryoelectron microscopy structure of a human TCRαβ in complex with the CD3 hexamer was obtained at 3.7 Å resolution,

revealing that the octameric TCR/CD3 complex is assembled with 1:1:1:1 stoichiometry of TCRαβ:CD3γε:CD3δε:CD3ζζ.[55] Assembly of the extracellular domains of TCR/CD3 is mediated by the constant domains and connecting peptides of TCRαβ that pack against CD3γε-CD3δε, forming a trimer-like structure proximal to the plasma membrane. The transmembrane segment of the CD3 complex adopts a barrel-like structure formed by interaction of the two transmembrane helices of CD3ζζ with those of CD3γε and CD3δε. Insertion of the transmembrane helices of TCRαβ into the barrel-like structure via both hydrophobic and ionic interactions results in transmembrane assembly of the TCR/CD3 complex.

Clinical Consequences of Altered or Missing Functions of the T-Cell Receptor/CD3 Complex

Homozygous mutations leading to complete deficiencies of either CD3δ, CD3ε or CD3ζ protein produce a form of severe combined immunodeficiency (SCID) (Chapter 34) characterized by severe T-cell lymphopenia, but the presence of phenotypically normal B cells and NK cells ($T^-B^+NK^+$SCID).[56,57]

Mutations in *CD3G* leading to deficiency of CD3γ produce considerable clinical heterogeneity ranging from severe immunodeficiency in infants to mild forms of autoimmunity in adulthood. Homozygous deficiency in CD3γ impairs, but does not abrogate, T-cell development, leading to mild T lymphopenia, reduction in cell-surface expression of TCR/CD3 complex on peripheral T cells by 75% to 80%, and impaired in vitro proliferative T-cell responses to lectins and to anti-CD3 mAbs. In peripheral blood, there are differential effects on phenotypically defined T-cell subsets, with very few CD8 T cells, a 10-fold reduction in $CD45RA^+$ CD4 T cells ("naïve helper" subset), and normal numbers of $CD45RO^+$ CD4 T cells ("memory" cells).[58]

Early Events in T-Cell Receptor/CD3 Signaling

Stimulation of the TCR/CD3 complex by pMHC leads to the phosphorylation of tyrosine residues in the CD3 ITAMs by the SRC-like protein tyrosine kinase LCK.[59] The phosphorylated CD3 ITAMs in turn create high-affinity binding sites for the SH2 domains of the ZAP-70 protein tyrosine kinase, leading to its recruitment to the TCR/CD3 complex and to its activation (Chapter 10).[59,60] The consequences of ZAP-70 deficiency (selective T-cell immunodeficiency in humans) underscore the centrality of its role in T-cell activation (Chapter 34).

The TCR appears to act as a mechanosensor in order to trigger the cascade of complex biochemical events leading to the activation of T-cell effector function. As the T cell migrates over the cell surface of an antigen-presenting cell or target cell, the binding of the pMHC complex to the TCR causes the TCR to act as a lever, converting horizontal force into a vertical force that acts upon the CD3 chains, exposing their ITAMs for phosphorylation. Following the initiation of signaling, sustained signaling appears to involve multimerization of TCR/CD3 complexes and engagement of co-receptors.[16,54]

T-CELL CO-RECEPTORS: CD4 AND CD8

Expression of CD4 and CD8 divides mature T cells into two broad distinct subsets: CD4 T cells (Chapter 9), which recognize peptides in the context of class II MHC molecules, and CD8 T cells (Chapter 9), which recognize antigens presented by class I MHC molecules. Indeed, CD4 binds directly to class II MHC molecules, and CD8 interacts directly with class I MHC molecules (Fig. 4.15) (Chapter 6). The cytoplasmic domains of CD4 and CD8 associate with LCK, and serve to bring LCK into contact with the CD3 chains of the pMHC-engaged TCR/CD3 complexes, leading to the phosphorylation of CD3 ITAMs and initiation of TCR signaling (Chapter 10).

The expression of the CD4 and CD8 co-receptors is highly regulated during T-cell development in the thymus (Chapter 9). Thymocytes initially express neither co-receptor ("double negative"). $CD4^-CD8^-$ thymocytes destined to become TCRαβ T cells progress through a $CD4^+CD8^+$ ("double-positive") stage to become mature CD4 or CD8 T cells. Positive and negative selection of thymocytes on the basis of their TCR specificities and commitment to the CD4 or CD8 lineages occur during the double-positive stage.

CD4: Structure and Binding to Major Histocompatibility Complex Class II Molecules

A member of the IgSF, CD4 is a 55-kDa glycoprotein whose relatively rigid extracellular region contains four IgSF domains (designated D1-4). Its cytoplasmic domain contains two cysteine residues that mediate a noncovalent interaction with LCK through a "zinc clasp"–like structure formed with a dicysteine motif in the N-terminal region of LCK.[59,61–63]

The N-terminal domain (D1) of CD4 binds between the membrane-proximal α2 and β2 domains of MHC class II. Thus CD4 interacts with pMHC class II at a distance from the α-helices and peptide contacted by the TCR, enabling the TCR and CD4 to bind the same MHC class II molecule simultaneously.

Although MHC molecules are highly polymorphic, the CD4 contact sites are highly conserved. In humans, CD4 targets nonpolymorphic residues shared by all three MHC class II molecules (human leukocyte antigen [HLA]-DR, DP, and DQ). The crystal structure of the TCRαβ–pMHC–CD4 ternary complex assumes a V-shape with pMHC at the apex and with TCRαβ and CD4 forming the arms of the V. There is no direct interaction between the co-receptor and the TCR heterodimer, indicating that pMHC brings the TCR and CD4 together. The approximately 70 Å of separation between the membrane proximal domains of TCRαβ and CD4 would allow the CD3 chains to lie within the open angle between TCRαβ and CD4, promoting interactions between CD3 chains and CD4-associated LCK.[59,61,63]

Experiments using soluble forms of CD4 and pMHC reveal that monomeric CD4 binds pMHC with very low affinity (Kd approximately 200 μM). The binding of CD4 to pMHC is of lower affinity than that TCRαβ to pMHC (Kd 1 to 10 μM) and displays a far more rapid off time. Because of the low affinity and the rapid off time, it is unlikely that interactions of CD4 with MHC class II molecules initiate the interaction between a T cell and an antigen-presenting cell (Chapter 6). Rather, these binding characteristics are more compatible with a model in which the initial event is the interaction between the TCR and pMHC, followed by the recruitment of CD4, which acts primarily to promote signaling events through the delivery of LCK.[59,61,63]

CD8: Structure and Binding to Major Histocompatibility Complex Class I Molecules

There are two CD8 polypeptides, α and β, and these are expressed on the cell surface either as a disulfide-linked CD8αα homodimer or as a disulfide-linked CD8αβ heterodimer.

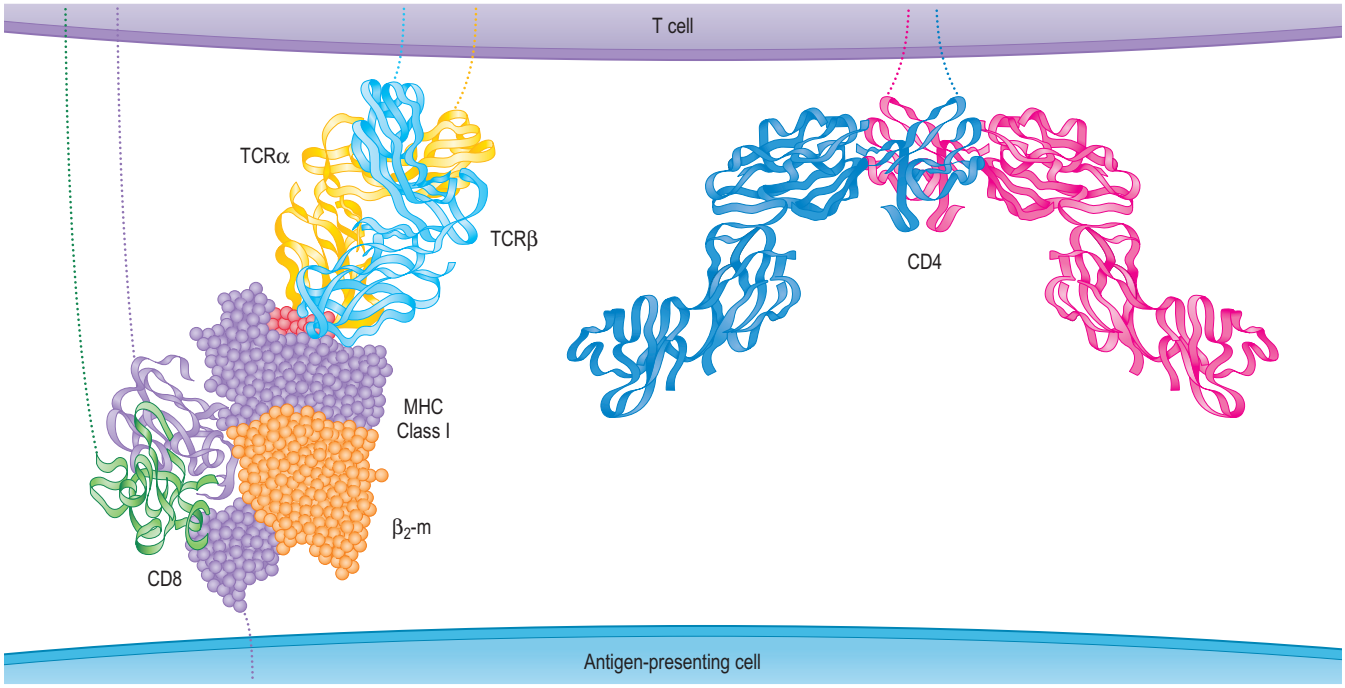


FIG. 4.15 Illustration of the Interactions Between the T-Cell Receptor, Peptide-Major Histocompatibility Complex, and CD8. A composite illustration of the human leukocyte antigen (HLA)-A*0201 structure in complex with a Tax peptide and its cognate T-cell receptor α and β chains (protein data bank [pdb] designation 1BD2) with the human CD8αα/HLA-A*0201 structure (pdb designation 1AKJ) was generated by superposition of the HLA moiety of the two structures. The HLA heavy chain is indicated as major histocompatibility complex *(MHC)*, its light chain (β_2-microglobulin) as β_2-m, the CD8αα homodimer as *CD8*, the T-cell receptor α and β chains as *TCRα* and *TCRβ*. In addition, the *CD4* homodimer (pdb file 1WIO) is shown to scale. Connecting peptides, transmembrane, and cytoplasmic domains are drawn by hand and indicated by *dotted lines*. (Figure courtesy of David H. Margulies, National Institute of Allergy and Infectious Diseases, National Institutes of Health.)

On most αβ T cells, CD8αβ is the predominant form of CD8 while natural killer (NK) cells (Chapter 12), intestinal intraepithelial T cells, MAIT cells, and γδ T cells mostly express CD8αα.[59,61–63]

CD8α, a 34- to 37-kDa protein, and CD8β, a 32-kDa protein, share about 20% amino acid sequence homology. Both are glycoproteins and IgSF members. Although CD8 subserves a co-receptor function similar to that of CD4, in structure it differs substantially from CD4. The CD8 extracellular regions have single N-terminal IgSF V domains at the end of extended mucin-like stalk regions of 48 amino acids (CD8α) or 35 to 38 amino acids (CD8β). A striking difference between the two forms of CD8 lies within the cytoplasmic domain. CD8α, like CD4, contains a cysteine-based motif that enables it to interact with LCK through a "zinc clasp"-like structure. In contrast, CD8β lacks this motif and does not associate with LCK. Interestingly, CD8αβ appears to be a more effective activator of TCR signaling than CD8αα. This may reflect the palmitoylation of the cytoplasmic domain of CD8β, which allows CD8αβ to associate with lipid rafts during T-cell activation.[59,61,63,64]

The structure of CD8αα pMHC class I complexes demonstrates that CD8αα binds to conserved residues in the α3 domain of MHC class I (i.e., a nonpolymorphic, membrane-proximal region of the molecule distinct from the peptide-binding groove engaged by the TCR) (Chapter 5). Compared to the interaction of CD4 and MHC class II, binding is more antibody-like, with a loop of the MHC α3 domain locked between the CDR-like loops of the two CD8α IgSF V domains. Models of the structure of the TCRαβ–pMHC–CD8 ternary complex propose a "V" shape similar to that of the crystal structure of TCRαβ–pMHC–CD4, with pMHC at the apex of the "V" and the TCR and CD8 forming the arms of the "V." CD8 binds to pMHC with lower affinity and with faster kinetics than the TCR. Thus, the binding properties of the CD8 co-receptor, like those of CD4, are consistent with a model in which the TCR initiates pMHC binding, followed by engagement of CD8 to the same pMHC.[59,61–64]

CO-STIMULATORY AND INHIBITORY T-CELL MOLECULES: THE CD28 FAMILY

Although the T-cell response to antigen requires the binding of the TCR and its co-receptors to pMHC, additional receptor–ligand interactions affect the outcome by delivering signals that promote activation (co-simulation) or that inhibit it (Table 4.4). Prominent among these are the interactions of members of the CD28 family with their cell-surface ligands on antigen-presenting cells.[65] This family includes CD28, inducible co-stimulator (ICOS), cytotoxic T-lymphocyte-associated antigen-4 (CTLA-4), B- and T-lymphocyte attenuator (BTLA), and program death-1 (PD-1). CD28 and ICOS are co-stimulatory receptors; the major functions of CTLA-4, PD-1, and BTLA are inhibitory. CD28 and CTLA-4 are T-cell specific, whereas BTLA and PD-1 are also expressed by B cells and ICOS by NK cells. CD28, CTLA-4, and PD-1 are the targets of therapeutic interventions in current clinical practice.

TABLE 4.4 CD28 Superfamily

Receptor	Expression	Ligand	Function on T cells
CD28	Most CD4 T cells	B7-1 (CD80)	Co-stimulation of IL-2 production and proliferation
	50% CD8 T cells	B7-2 (CD86)	Promotes T-cell survival
ICOS	Activated and memory T cells NK cells Not expressed by naïve T cells	ICOS ligand	Promotes T-cell differentiation and effector T-cell function
CTLA-4	Upregulated after T-cell activation	B7-1 (CD80)	Inhibits IL-2 production and proliferation
		B7-2 (CD86)	Promotes peripheral T-cell tolerance
PD-1	Upregulated after activation of T and B cells, myeloid cells	PD-L1 (B7-H1)	Inhibits proliferations and cytokine production
		PD-L2 (B7-DC)	Promotes peripheral T-cell tolerance
BTLA	T and B cells, myeloid cells, dendritic cells	HVEM (herpesvirus-entry mediator)	Inhibits T-cell proliferation

BTLA, B- and T-lymphocyte attenuator; *CTLA-4*, cytotoxic T-lymphocyte-associated antigen-4; *ICOS*, inducible co-stimulator; *IL-2*, interleukin-2; *NK*, natural killer; *PD-1*, program death-1.

All members of the CD28 family have a single extracellular IgSF V domain and have, as their ligands, members of the B7 family of cell-surface molecules. CD28, CTLA-4, and ICOS are disulfide-linked homodimers whose cytoplasmic domains contain the SH2-binding motif YXXM. In contrast, PD-1 and BTLA are monomers whose cytoplasmic domains each contain an ITIM and an *I*mmunoreceptor *T*yrosine-based *S*witch *M*otif (ITSM).

CD28 and Cytotoxic T-Lymphocyte-Associated Antigen-4

Half of CD8 T cells and virtually all human CD4 T cells constitutively express CD28. CD28 binds to B71 (CD80) and B7.2 (CD86) through a MYPPPYY motif in its extracellular domain. Interactions with these ligands leads to the phosphorylation of the YMNM sequence in the CD28 cytoplasmic domain and to the recruitment of phosphatidylinositol 3-kinase and Grb2. CD28 stimulation usually does not elicit a cellular response in the absence of TCR signaling. Rather, CD28 signals act in concert with those of the TCR to promote cytokine production, T-cell expansion, and T-cell survival. TCR signaling in the absence of CD28 co-stimulation can induce T-cell anergy (Chapter 10).

CTLA-4 inhibits the response to TCR and CD28 signals and acts to terminate peripheral T-cell responses. Its importance in human immunology is underscored by observations that *CTLA4* haploinsufficiency produces a syndrome of immune dysregulation characterized by decreased numbers of T regulatory cells (Treg), hyperactive effector T cells, hypogammaglobulinemia, and clinical autoimmunity (Chapter 33).

The majority of CTLA-4 resides in intracellular compartments. T-cell activation promotes the cell-surface expression of CTLA-4 by regulating both its transport to the surface and its subsequent internalization. CTLA-4 also binds B7.1 and B7.2 but does so with substantially greater affinity than does CD28. Moreover, the binding of CTLA-4 to these ligands is divalent, whereas that of CD28 is monovalent. Thus the inhibitory complexes formed by CTLA-4 are more stable than the co-stimulatory interactions involving CD28. CTLA-4 can inhibit T-cell activation by outcompeting CD28 for B7 ligands and, through transendocytosis, by removing B7 molecules from the antigen-presenting cell. In addition, CTLA-4 can induce "reverse signaling" through B7.1 and B7.2 to the antigen-presenting cell, upregulating the enzyme indoleamine 2,3-dioxygenase (IDO), which in turn breaks down tryptophan, a requirement for T-cell proliferation.

The importance of CD28 co-stimulation has made it an attractive target for therapeutic intervention.[66,67] Indeed, two soluble fusion proteins composed of the extracellular domain of human CTLA-4 and the constant regions of human IgG1, abatacept and belatacept, are effective therapies for the treatment of rheumatoid arthritis (Chapter 53) and the prevention of renal allograft rejection (Chapter 89). These fusion proteins are thought to inhibit CD28 co-stimulation through blockade of its B7 ligands, but some of their immunosuppressive effects may be indirect through the induction of IDO and consequent local depletion of tryptophan. Conversely, inhibition of CTLA-4 by mAbs can promote durable immune responses against certain malignancies.

Program Death-1

PD-1 is a key inhibitory receptor that attenuates TCR signaling, promotes T-cell tolerance, and is associated with T-cell exhaustion. PD-1 is not found on resting T cells, and its expression during T-cell activation requires transcriptional activation. PD-1 binds to two ligands: programmed death ligand-1 (PDL-1), which is widely expressed, and PDL-2, which is found primarily on professional antigen-presenting cells. Engagement of ligand induces tyrosine phosphorylation of the ITIM and ITSH motifs in the cytoplasmic domain of PD-1, leading to the recruitment of the tyrosine phosphatase SHP-2. Continued stimulation of T cells by antigen leads to sustained expression of PD-1 and differentiation into a state of hyporesponsiveness termed *T-cell exhaustion*. Blockade of PD-1 has shown considerable promise in the treatment of diverse human malignancies.[68]

REFERENCES

1. Williams AF, Barclay AN. The immunoglobulin superfamily—domains for cell surface recognition. *Annu Rev Immunol*. 1988;6:381–405.
2. Padlan EA. Anatomy of the antibody molecule. *Mol Immunol*. 1994;31(3):169–217.
3. Sanchez-Mazas A, Fernandez-Vina M, Middleton D, et al. Immunogenetics as a tool in anthropological studies. *Immunology*. 2011;133(2):143–164.
4. Ehrenmann F, Kaas Q, Lefranc MP. IMGT/3Dstructure-DB and IMGT/DomainGapAlign: a database and a tool for immunoglobulins or antibodies, T cell receptors, MHC, IgSF and MhcSF. *Nucleic Acids Res*. 2010;38(Database issue):D301–D307.
5. Nelson AL. Antibody fragments: hope and hype. *MAbs*. 2010;2(1):77–83.
6. Schroeder Jr HW, Cavacini L. Structure and function of immunoglobulins. *J Allergy Clin Immunol*. 2010;125(2 Suppl 2):S41–S52.
7. Johansen FE, Braathen R, Brandtzaeg P. Role of J chain in secretory immunoglobulin formation. *Scand J Immunol*. 2000;52(3):240–248.
8. Chen FH, Arya SK, Rinfret A, et al. Domain-switched mouse IgM/IgG2b hybrids indicate individual roles for C mu 2, C mu 3, and C mu 4 domains in the regulation of the interaction of IgM with complement C1q. *J Immunol*. 1997;159(7):3354–3363.

9. Diebolder CA, Beurskens FJ, de Jong RN, et al. Complement is activated by IgG hexamers assembled at the cell surface. *Science*. 2014;343(6176):1260–1263.
10. Vidarsson G, Dekkers G, Rispens T. IgG subclasses and allotypes: from structure to effector functions. *Front Immunol*. 2014;5:520.
11. James LK, Till SJ. Potential mechanisms for IgG4 inhibition of immediate hypersensitivity reactions. *Curr Allergy Asthma Rep*. 2016;16(3):23.
12. Chen LYC, Mattman A, Seidman MA, Carruthers MN. IgG4-related disease: what a hematologist needs to know. *Haematologica*. 2019;104(3):444–455.
13. de Sousa-Pereira P, Woof JM. IgA: structure, function, and developability. *Antibodies (Basel)*. 2019;8(4):57.
14. Sutton BJ, Davies AM, Bax HJ, Karagiannis SN. IgE antibodies: from structure to function and clinical translation. *Antibodies (Basel)*. 2019;8(1):19.
15. Davis MM, Bjorkman PJ. T-cell antigen receptor genes and T-cell recognition. *Nature*. 1988;334:395–402.
16. Wucherpfennig KW, Gagnon E, Call MJ, et al. Structural biology of the T-cell receptor: insights into receptor assembly, ligand recognition, and initiation of signaling. *Cold Spring Harb Perspect Biol*. 2010;2(4):a005140.
17. O'Brien RL, Roark CL, Jin N, et al. Gammadelta T-cell receptors: functional correlations. *Immunol Rev*. 2007;215:77–88.
18. Godfrey DI, Uldrich AP, McCluskey J, et al. The burgeoning family of unconventional T cells. *Nat Immunol*. 2015;16(11):1114–1123.
19. Van Rhijn I, Godfrey DI, Rossjohn J, Moody DB. Lipid and small-molecule display by CD1 and MR1. *Nat Rev Immunol*. 2015;15(10):643–654.
20. Keller AN, Eckle SB, Xu W, et al. Drugs and drug-like molecules can modulate the function of mucosal-associated invariant T cells. *Nat Immunol*. 2017;18(4):402–411.
21. Nielsen MM, Witherden DA, Havran WL. Gammadelta T cells in homeostasis and host defence of epithelial barrier tissues. *Nat Rev Immunol*. 2017;17(12):733–745.
22. Tonegawa S. Somatic generation of antibody diversity. *Nature*. 1983;302:575–581.
23. Krangel MS. Mechanics of T cell receptor gene rearrangement. *Curr Opin Immunol*. 2009;21(2):133–139.
24. Roth DB. V(D)J recombination: mechanism, errors, and fidelity. *Microbiol Spectr*. 2014;2(6). https://doi.org/10.1128/microbiolspec.MDNA3-0041-2014.
25. Carmona LM, Schatz DG. New insights into the evolutionary origins of the recombination-activating gene proteins and V(D)J recombination. *FEBS J*. 2017;284(11):1590–1605.
26. Zachau HG. The immunoglobulin kappa gene families of human and mouse: a cottage industry approach. *Biol Chem*. 2000;381(9–10):951–954.
27. Kirkham PM, Schroeder Jr. HW. Antibody structure and the evolution of immunoglobulin V gene segments. *Semin Immunol*. 1994;6(6):347–360.
28. Kawasaki K, Minoshima S, Nakato E, et al. One-megabase sequence analysis of the human immunoglobulin lambda gene locus. *PCR Meth App*. 1997;7(3):250–261.
29. Matsuda F, Ishii K, Bourvagnet P, et al. The complete nucleotide sequence of the human immunoglobulin heavy chain variable region locus. *J Exp Med*. 1998;188(11):2151–2162.
30. Methot SP, Di Noia JM. Molecular mechanisms of somatic hypermutation and class switch recombination. *Adv Immunol*. 2017;133:37–87.
31. Hwang JK, Alt FW, Yeap LS. Related mechanisms of antibody somatic hypermutation and class switch recombination. *Microbiol Spectr*. *3(1):MDNA3*. 2015:0037–2014.
32. Casellas R, Basu U, Yewdell WT, et al. Mutations, kataegis and translocations in B cells: understanding AID promiscuous activity. *Nat Rev Immunol*. 2016;16(3):164–176.
33. Tomlinson IM, Cox JP, Gherardi E, et al. The structural repertoire of the human V kappa domain. *EMBO J*. 1995;14(18):4628–4638.
34. Chothia C, Lesk AM, Gherardi E, et al. Structural repertoire of the human VH segments. *J Mol Biol*. 1992;227:799–817.
35. Khass M, Blackburn T, Burrows PD, et al. A new role for VpreB: an invariant surrogate antigen that selects Ig antigen binding sites. *Sci Immunol*. 2016;1(1):aaf6628.
36. Arden B, Clark SP, Kabelitz D, et al. Human T-cell receptor variable gene segment families. *Immunogenet*. 1995;42(6):455–500.
37. Wang LD, Clark MR. B-cell antigen-receptor signalling in lymphocyte development. *Immunology*. 2003;110(4):411–420.
38. Pelanda R, Braun U, Hobeika E, et al. B cell progenitors are arrested in maturation but have intact VDJ recombination in the absence of Ig-alpha and Ig-beta. *J Immunol*. 2002;169(2):865–872.
39. Neuberger MS, Patel KJ, Dariavach P, et al. The mouse B-cell antigen receptor: definition and assembly of the core receptor of the five immunoglobulin isotypes. *Immunol Rev*. 1993;132:147–161.
40. Tsuganezawa K, Kiyokawa N, Matsuo Y, et al. Flow cytometric diagnosis of the cell lineage and developmental stage of acute lymphoblastic leukemia by novel monoclonal antibodies specific to human pre-B-cell receptor. *Blood*. 1998;92(11):4317–4324.
41. Gauld SB, Cambier JC. Src-family kinases in B-cell development and signaling. *Oncogene*. 2004;23(48):8001–8006.
42. Niiro H, Clark EA. Regulation of B-cell fate by antigen-receptor signals. *Nat Rev Immunol*. 2002;2(12):945–956.
43. Roozendaal R, Carroll MC. Complement receptors CD21 and CD35 in humoral immunity. *Immunol Rev*. 2007;219:157–166.
44. Fujimoto M, Poe JC, Inaoki M, Tedder TF. CD19 regulates B lymphocyte responses to transmembrane signals. *Semin Immunol*. 1998;10(4):267–277.
45. Carroll MC. The role of complement and complement receptors in induction and regulation of immunity. *Annu Rev Immunol*. 1998;16:545–568.
46. Fearon DT, Carroll MC. Regulation of B lymphocyte responses to foreign and self-antigens by the CD19/CD21 complex. *Annu Rev Immunol*. 2000;18:393–422.
47. Otero DC, Anzelon AN, Rickert RC. CD19 function in early and late B cell development: I. Maintenance of follicular and marginal zone B cells requires CD19-dependent survival signals. *J Immunol*. 2003;170(1):73–83.
48. Dempsey PW, Allison ME, Akkaraju S, et al. C3d of complement as a molecular adjuvant: bridging innate and acquired immunity. *Science*. 1996;271(5247):348–350.
49. Anania JC, Chenoweth AM, Wines BD, Hogarth PM. The human FcgammaRII (CD32) family of leukocyte FcR in health and disease. *Front Immunol*. 2019;10:464.
50. Clark EA, Giltiay NV. CD22: a regulator of innate and adaptive B cell responses and autoimmunity. *Front Immunol*. 2018;9:2235.
51. Nitschke L. CD22 and Siglec-G: B-cell inhibitory receptors with distinct functions. *Immunol Rev*. 2009;230(1):128–143.
52. Brazin KN, Mallis RJ, Das DK, et al. Structural features of the alphabetaT-CR mechanotransduction apparatus that promote pMHC discrimination. *Front Immunol*. 2015;6:441.
53. Rudolph MG, Stanfield RL, Wilson IA. How TCRs bind MHCs, peptides, and co-receptors. *Annu Rev Immunol*. 2006;24:419–466.
54. Kuhns MS, Davis MM, Garcia KC. Deconstructing the form and function of the TCR/CD3 complex. *Immunity*. 2006;24(2) 133–129.
55. Dong D, Zheng L, Lin J, et al. Structural basis of assembly of the human T cell receptor-CD3 complex. *Nature*. 2019;573(7775):546–552.
56. de Saint-Basile G, Geissmann F, Flori E, et al. Severe combined immunodeficiency caused by deficiency in either the delta or the epsilon subunit of. *J Clin Invest*. 2004;114(10):1512–1517.
57. Roberts JL, Lauritsen JP, Cooney M, et al. T-B+NK+ severe combined immunodeficiency caused by complete deficiency of the CD3zeta subunit of the T-cell antigen receptor complex. *Blood*. 2007;109(8):3198–3206.
58. Recio MJ, Moreno-Pelayo MA, Kilic SS, et al. Differential biological role of CD3 chains revealed by human immunodeficiencies. *J Immunol*. 2007;178(4):2556–2564.
59. Gao GF, Rao Z, Bell JI. Molecular coordination of alphabeta T-cell receptors and coreceptors CD8 and CD4 in their recognition of peptide-MHC ligands. *Trends Immunol*. 2002;23(8):408–413.
60. Palacios EH, Weiss A. Function of the Src-family kinases, Lck and Fyn, in T-cell development and activation. *Oncogene*. 2004;23(48):7990–8000.
61. van der Merwe PA, Davis SJ. Molecular interactions mediating T cell antigen recognition. *Annu Rev Immunol*. 2003;21:659–684.

62. Artyomov MN, Lis M, Devadas S, et al. CD4 and CD8 binding to MHC molecules primarily acts to enhance Lck delivery. *Proc Natl Acad Sci U S A*. 2010;107(39):16916–16921.
63. Wang Z, Yang X, Chu X, et al. The structural basis for the oligomerization of the N-terminal domain of SATB1. *Nucleic Acids Res*. 2012;40(9):4193–4202.
64. Chang HC, Tan K, Ouyang J, et al. Structural and mutational analyses of a CD8alphabeta heterodimer and comparison with the CD8alphaalpha homodimer. *Immunity*. 2005;23(6):661–671.
65. Murakami N, Riella LV. Co-inhibitory pathways and their importance in immune regulation. *Transplantation*. 2014;98(1):3–14.
66. Baumeister SH, Freeman GJ, Dranoff G, Sharpe AH. Coinhibitory pathways in immunotherapy for cancer. *Annu Rev Immunol*. 2016;34:539–573.
67. Topalian SL, Drake CG, Pardoll DM. Immune checkpoint blockade: a common denominator approach to cancer therapy. *Cancer Cell*. 2015;27(4):450–461.
68. Xu-Monette ZY, Zhang M, Li J, Young KH. PD-1/PD-L1 blockade: have we found the key to unleash the antitumor immune response? *Front Immunol*. 2017;8:1597.
69. Garcia KC, Degano M, Stanfield RL, et al. An alphabeta T cell receptor structure at 2.5 A and its orientation in the TCR-MHC complex. *Science*. 1996;274(5285):209–219.

5

The Major Histocompatibility Complex

Dimitri S. Monos and Raja Rajalingam

A primary objective of the immune system is to protect the body against pathogens. The major histocompatibility complex (MHC) comprises a genomic region that has evolved to include many genes responsible for coordinating the immune response. It is named the histocompatibility complex because it was first identified as the site of numerous genes that determined whether transplanted tissue would be accepted or rejected. We now know that this region coordinates immunological functions far beyond those related to histocompatibility. The MHC region includes genes that determine both innate and adaptive immunity and thus influences responses to pathogens (viruses, bacteria, fungi, and parasites), transplantation, autoimmunity, cancer, vaccinations, certain drugs, and possibly other functionalities presently unknown. This chapter describes the genomic organization and immunological importance of the MHC with a special focus on the HLA (*h*uman *l*eukocyte *a*ntigen) genes. These HLA molecules present self and foreign peptides to T cells (Chapter 6) and thus play a central role in adaptive immunity. They also interact with receptors on the surface of natural killer (NK) cells (Chapter 12) and thus also play a role in innate immunity.

The terms HLA and MHC are often used interchangeably. However, in this chapter the term "MHC" is reserved for the broader genomic region, while the term "HLA" refers to the human class I and class II genes and their respective protein products.

CLINICAL RELEVANCE

- HLA molecules regulate antigen-specific immune responses by binding and then presenting pathogen-derived peptides to either CD4 or CD8 T cells.
- Certain HLA alleles are the major genetic determinants of susceptibility to many autoimmune diseases or drug hypersensitivity reactions because they can present specific self-peptides or small molecules (drugs) to T cells.
- HLA molecules play a key role in transplant rejection and appear to regulate placental development in pregnancy.
- Cancerous cells modify expression of their HLA genes in order to avoid immune recognition.

GENOMIC ORGANIZATION OF THE MAJOR HISTOCOMPATIBILITY COMPLEX

The human MHC region includes approximately 3.8 million base pairs (Mbp) of DNA on the short arm of chromosome 6 (6p21.3). It is defined as the region spanning from the gamma-aminobutyric acid type B receptor subunit 1 *(GABBR1)* gene on the telomeric side of the region to the Kinesin Family Member C1 *(KIFC1)* gene toward the centromere (ENSEMBL 86 GRCh38.p7 coordinates chr6: 29555629-33409924).[1] The functional MHC region may include additional downstream and upstream sequences totaling seven or more Mbp.

The classic 3.8 Mbp MHC region is the most gene-dense segment of human genome. It includes 158 protein-coding genes and 86 pseudogenes of unknown functionality (ENSEMBL 86 GRCh38.p7).[2] At least 65 (41%) of the coding genes are involved in innate and adaptive immunity.[2] The MHC is divided into three regions: class I, class II, and class III (Fig. 5.1). The class I region is at the telomeric end and includes the classical HLA class I genes (*HLA-A, -B*, and *-C*), the class I-related (like) genes *(MICA, MICB)*, the nonclassical HLA class I genes (*HLA-E, -F*, and *-G*) and four pseudogenes (*HLA-H, -K, -J*, and *-L*). The class II region occupies the centromeric end and contains the *DRA* and *DRB1* genes and, depending on the DR haplotype, one or none of the *DRB3*, *DRB4, DRB5* genes that code for DR52, DR53, or DR51 molecules, respectively; the *DQA1* and *DQB1* genes that encode the DQ molecule; and the *DPA1* and *DPB1* genes that encode the DP molecule. It also includes the *DM* and *DO* genes encoding the antigen-processing molecules DM and DO that are involved in the class II antigen presentation pathway, and the *TAP* and *LMP* genes encoding proteins that are involved in the class I antigen presentation pathway (Chapter 6). The class III region, interposed between class I and II regions, contains many immune and nonimmune genes. These include complement components, lymphotoxin, tumor necrosis factor, heat shock proteins, *NFKB, NOTCH4*, and 21-hydroxylase *(CYP21)*. Genes within the HLA class I and class II regions variably reflect the mechanisms of insertion, deletion, gene duplication, gene conversion, and mutation used by evolution to expand diversity of function. While the genomic organization of class I and class II genes is quite distinct, the derived molecules share highly similar structures, likely driven by the shared role of these molecules in presenting peptides to T-cell receptors (TCRs).

The MHC is also marked by extensive linkage disequilibrium (LD) between the class I *HLA-A, -B, -C* and class II *HLA-DR* and *-DQ*, but not *–DP*, genes. LD is the phenomenon whereby particular alleles of gene loci on the same strand of DNA are inherited together more often than expected by chance alone. Anthropological population studies suggest that the particular combinations of alleles of the different genes, as distant as they may be, provide a survival advantage. This may reflect functional interdependence of these alleles in antigen-specific immune responses.

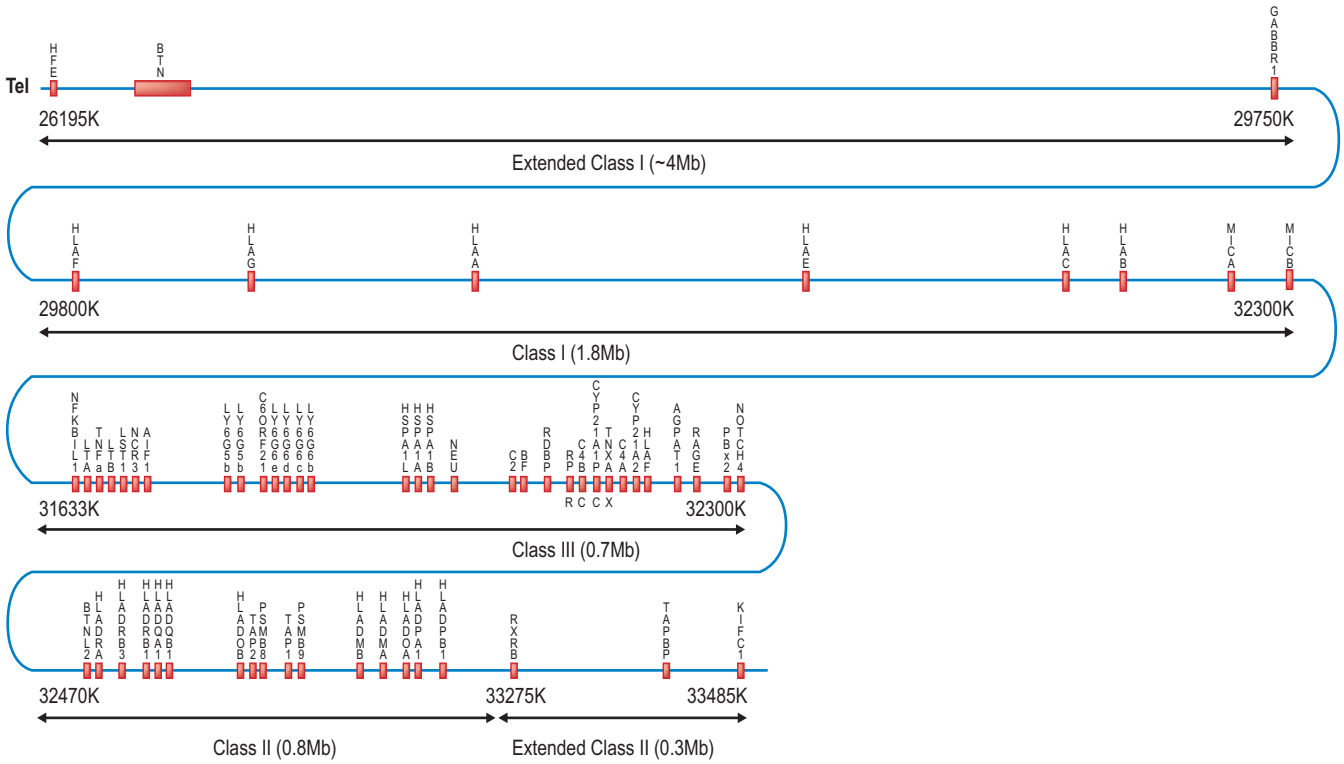


FIG. 5.1 Gene Map of the Extended Major Histocompatibility Complex. The core of the major histocompatibility complex consists of three major regions: class I, class III, and class II. The extended class I and extended class II regions of the complex are labeled. Sequence numbering begins at the telomere. The map depicts immune-related expressed genes as well as certain reference genes. The approximate locations of these selected genes near the start or end of the regions are indicated. (Modified from Beck S, Trowsdale J. The human major histocompatibility complex: lessons from the DNA sequence. *Annu Rev Genomics Hum Genet*. 2000;1:117–137.)

A particular combination of alleles of different loci in LD on the same strand of DNA is called a "haplotype." The frequency of a given haplotype varies among different populations, reflecting distant selection by pathogens, ethnic admixture, and drastic population reductions (genetic bottlenecks). LD is strongest between *HLA-B* and *-C* and between *HLA-DR* and *-DQ*, most likely due to their physical proximity. However, due to a hot spot of recombination between the *DQ* and *DP* there is no LD between *DP* and rest of the HLA genes even though *DQ* and *DP* are relatively proximal to each other in linear distance.

The haplotype is the unit of inheritance of the MHC from either parent. Each parent shares one haplotype with each of their children and typically differs from a child by one haplotype. Two siblings may share two, one, or no haplotypes, and thus range from being HLA-identical, through haplo-identical, to HLA-disparate. A parent is usually only haplo-identical with their child. Exceptions to this rule may occur in inbred populations, where both parents may share an identical HLA haplotype by descent. The HLA alleles originating from the maternal and paternal haplotypes are both expressed.

Ten years of genome-wide association studies (GWAS) have revealed a large number (884) of single nucleotide polymorphisms (SNPs) within the MHC that are associated with many (479) traits and diseases, establishing the MHC as the only region in the genome with this high density of SNPs that is associated with so many diseases.[3] The complexity of the region with its many insertions, deletions, duplications, and LD does not allow an easy dissection of disease-causing variants. However, recent developments in next-generation sequencing (NGS) of the entire MHC[4] will most likely advance our understanding of the principles underlying this complex genomic organization, how this complexity results in so many biological interdependencies, and how it contributes to the pathophysiology of the diseases associated with the MHC.

KEY CONCEPTS

Genomic Organization of the Major Histocompatibility Complex

- The highly complex MHC is associated with more diseases than any other genomic region of comparable size.
- The class I region contains the polymorphic *HLA-A*, *-B*, and *-C* genes, the less polymorphic non-classical class I *HLA-E*, *HLA-F*, and *HLA-G* genes, and the class I-related *MICA* and *MICB*.
- The class II region contains the *HLA-DR A* and *B*, *DQ A* and *B*, and *DP A* and *B* genes. It also contains the *TAP*, *LMP*, *DM*, and *DO* genes, which encode molecules that help process antigens into peptides that can bind class I and class II molecules.
- Genes within the MHC demonstrate extensive LD. A string of linked, polymorphic alleles is termed an MHC haplotype.
  - Haplotypes are preserved by means of natural selection, whose driving force is reproductive fitness.
  - Common haplotypes within a given population appear to reflect functional interdependencies between MHC gene alleles.
  - Different populations can exhibit different haplotypes.
- The HLA genes of the two chromosomes are both expressed.

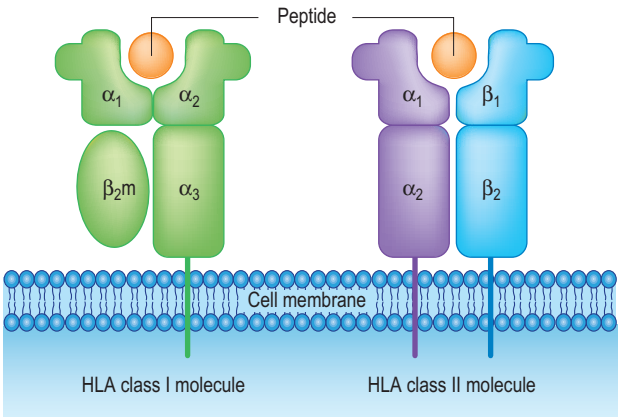


FIG. 5.2 Major Histocompatibility Complex (MHC) Class I and II Domain Organization. Although major histocompatibility complex class I and class II proteins have a different chain structure, the organization of their domains is extremely similar. Both class I and II molecules are expressed on the cell surface where they are accessible to T cells. Both have an outermost domain that contains a cleft where antigenic peptides are displayed. Two of the three class I α domains fold to create a domain with a peptide-binding cleft. The remaining α_3 domain helps support the peptide-binding domain and anchors the molecule to the cell membrane. The class I molecule also contains an extrinsic β chain, β_2-microglobulin ($\beta_2 m$), which is encoded by a separate, invariant gene. β_2-Microglobulin associates with the α_3 domain to support the antigen-binding domain created by the α_1 and α_2 domains. Class II molecules share a similar overall structure, but are the product of two genes, each containing two domains, one relatively constant that is proximal to the cell membrane and one highly variable that interacts with the peptide. *HLA*, Human leukocyte antigen.

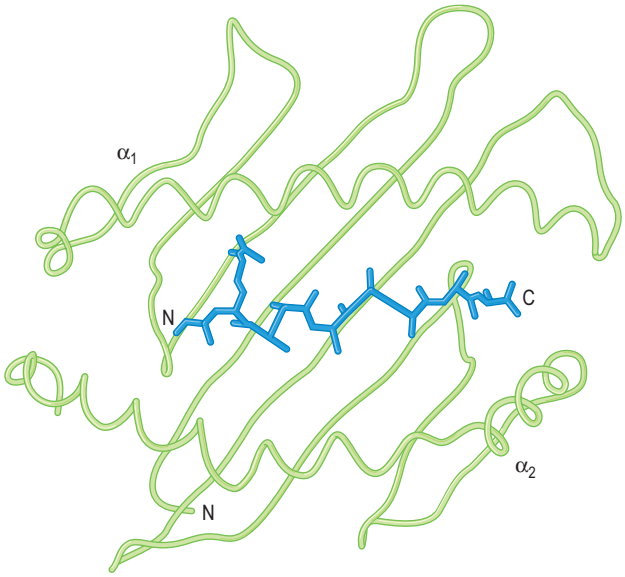


FIG. 5.3 The Three-Dimensional Structure of Human Leukocyte Antigen-B27. The α-helical margins of the peptide-binding cleft contain the bound peptide RRIKAITLK, which is oriented with its amino terminus to the left. There are extensive contacts at the ends of the cleft between peptide main-chain atoms and conserved human leukocyte antigen (HLA) side chains. The peptide amino and carboxyl termini are tethered to the cleft by hydrogen bonds and charge interactions. The peptide reciprocally stabilizes the three-dimensional fold of HLA-B27. The positively charged side chain of arginine in the P2 position of the peptide inserts into the B pocket, which contains a complementary negatively charged glutamic acid at its base. The resulting salt bridge is the dominant anchor for the peptide. Side chains P4, P6, and P8 make minor contributions to the interaction of the peptide with the HLA-B27 molecule. The central region of the peptide is left free to interact with a T-cell receptor. (Modified from Madden DR, Gorga JC, Strominger JL, Wiley DC. The three-dimensional structure of HLA-B27 at 2.1 A resolution suggests a general mechanism for tight peptide binding to MHC. *Cell*. 1992;70:1035.)

STRUCTURE AND FUNCTION OF THE HUMAN LEUKOCYTE ANTIGEN MOLECULES

The main function of both class I and class II HLA molecules is to bind peptides derived from self or nonself-antigens, and then traffic to the cell surface where these peptides can be displayed, or presented, for recognition by the appropriate T cells. Their structure has evolved to satisfy this particular requirement.

Classical Human Leukocyte Antigen Class I Molecules

The classical HLA-A, -B, and -C class I molecules consist of an α and a β chain. The α chain masses 45 kDa and is 362 to 366 amino acids long. It is encoded by the respective class I genes of MHC. The β chain, β_2-microglobulin (12 kDa), is encoded by its respective gene on chromosome 15. The α chain has three approximately 90-amino acid extracellular domains encoded by exons 2, 3, and 4 respectively; a transmembrane segment (approximately 25 amino acids) encoded by exon 5; and a C-terminal cytoplasmic end (approximately 30 amino acids) encoded by exons 6 and 7. β_2-Microglobulin, which is invariant, comprises the fourth domain (Fig. 5.2). The first two α domains (α_1 and α_2) are the most distal to the cell membrane. They combine to form a domain with a peptide-binding groove, or cleft, that is flanked by a surface that interacts with a TCR or a NK cell receptor.[5,6] The ends of the peptide-binding cleft are closed and fix the peptide's orientation. The sides of the peptide-binding cleft are composed of α helices and the floor is composed of symmetric strands of β-pleated sheet (Fig. 5.3). The α_3 domain and β_2-microglobulin create a combined structure that supports the peptide-binding domain and, along with the transmembrane domain of the α chain, attaches the molecule to the cell surface. Class I HLA molecules are ubiquitously expressed in all nucleated cells and in platelets. Expression of class I is significantly reduced on red blood cells and absent on sperm cells.

HLA class I molecules bind peptides derived from processed proteins of a pathogen or other self/nonself-peptides (Chapter 6). These peptides average nine amino acids long, and two or more of the amino acid side chains are used to anchor the peptide to pockets (see Fig. 5.3). Individual HLA class I alleles are generally distinguished by their own distinct pattern of peptide binding, as illustrated for selected HLA-B molecules in Table 5.1.[7] Among class I molecules, one or a few amino acid changes may considerably alter the binding properties of a binding pocket.

In a healthy non-endocytosing cell, HLA molecules are filled with a variety of peptides from self-molecules. The bound peptides are selected according to the binding motif of the particular

TABLE 5.1 Peptide-Binding Motifs Encoded by Different Human Leukocyte Antigen (HLA) Alleles Influence the Number of Peptides in a Protein That Can Be Recognized by a HLA Molecule (e.g., Human Immunodeficiency Virus [HIV] Envelope Protein)

Allele designation:	HLA-B*27:05	HLA-B*35:01	HLA-B*07:02
Peptide-binding motif	XRXXXXXX[KRYL]	XPXXXXXXY	XPXXXXXXL
Peptides from the HIV envelope protein able to bind to each allotype	IRGKVQKEY IRPVVSTQL TRPNNNTRK IRIQRGPGR SRAKWNNTL LREQFGNNK FRPGGGDMR WRSELYKYK KRRVVQREK ARILAVERY ERDRDRSIR LRSLCLFSY TRIVELLGR CRAIRHIPR IRQGLERIL	None	DPNPQEVVL KPCVKLTPL RPVVSTQLL SPLSFQTHL IPRRIRQGL
Number of peptides bound	15	0	5

Single-letter amino acid codes are used. *X*, Any amino acid; *R*, arginine; *K*, lysine; *Y*, tyrosine; *L*, leucine; *P*, proline, etc.
HLA, human leukocyte antigen.

allele. Even during viral infection or upon pathogen phagocytosis, the number of nonself-peptides may not be high. Together, the MHC class I and its peptide create a complex ligand that serves as the target of the TCR on the T-cell surface (Chapter 4).

The expression of class I molecules is upregulated by α-, β-, and γ-interferons, granulocyte-macrophage colony-stimulating factor (GM-CSF), and certain other cytokines (Chapter 14). Class I expression is governed by a regulatory element located ~160 nucleotides upstream from the initiation site of the class I gene. This site binds a number of regulatory factors, including those induced by interferons.

MICA and MICB

Within the class I region are *MICA* and *MICB* (MHC class I chain-related protein A and B). While both MICA and MICB are members of the class I family, neither associate with β_2-microglobulin nor bind peptides.[8] MICA and MICB are expressed as "danger signals" by virus-infected or otherwise stressed cells. Both are ligands for the activating NKG2D molecule *(KLRK1)*, a member of killer cell lectin-like receptors (KLR)[9] that appears on memory-effector T cells and NK cells and provides a signal to activate effector cytolytic responses.

Nonclassical Human Leukocyte Antigen-E, -F, and -G

The HLA nonclassical molecules E, F, and G are less polymorphic, have a limited tissue distribution, and have different functions than classical HLA class I molecules.[10] HLA-E primarily presents self-peptide to the TCR of CD8 T cells. The diversity of these self-peptides is limited and includes the leader peptide of classical class I HLA molecules.

The binding of **HLA-E** to inhibitory receptors (e.g., CD94/NKG2A) is an important part of the surveillance mechanism for missing self. In tumor cells, loss of class I expression results in a survival advantage for the particular tumor cell. In the absence of class I expression, HLA-E molecules no longer form a complex with intracellular class I leader peptides. As a result, HLA-E molecules are not expressed on the cell surface and the inhibitory signals to the NK cells are removed. This licenses the NK cell to kill the tumor target. Thus, HLA-E exists at the interface of innate and adaptive immunity.

HLA-F has a small binding cleft that does not contain peptide and its functions are not well understood. It mainly resides intracellularly and rarely reaches the cell surface.

HLA-G has limited tissue distribution and is primarily expressed by placental trophoblast cells, thymus, cornea, and some erythroid and endothelial precursor cells. HLA-G has a peptide groove, binds a nonamer peptide, and is recognized by the leukocyte immunoglobulin-like receptors (LILR-1 and LLIR-2) and a killer cell immunoglobulin-like receptor (KIR). In melanoma, HLA-G expression can be used by the tumor cells to avoid immunosurveillance by flooding the local microenvironment with soluble HLA-G and compromising the function of immune cells. The expression of HLA-G in the chorionic villi suggests a role in the maintenance of pregnancy. The mechanism appears to involve production of soluble forms of HLA-G. They appear to have an inhibitory role on the immune cells of the mother. Unique among other HLA molecules, HLA-G exists in different isoforms. Of these, four are expressed on the cell membrane, and three others exist as soluble forms. The functional significance of these isoforms is not known.

Class II Human Leukocyte Antigen Molecules

Class II HLA molecules are selectively expressed by professional antigen-presenting cells (e.g., macrophages, dendritic cells, B cells, and activated T cells) (Chapter 6). The HLA class II molecules are heterodimers that consist of two transmembrane glycoprotein α (34 kDa) and β (29 kDa) chains. Together, the α and β chains form a structure that is similar to HLA class I.

Unlike class I, both the α and β chains are encoded by genes within the MHC. Each of the two chains is composed of two extracellular domains. *DR, DQ*, or *DP* α *chain* includes α_1 and α_2 domains that are encoded by exons 2 and 3 of the *DRA, DQA*, or *DPA* gene. *DR, DQ*, or *DP* β *chain* includes β_1 and β_2 domains that are encoded by exons 2 and 3 of the *DRB, DQB*, or *DPB* gene. The α_1 and β_1 domains form the binding groove of the class II HLA molecule and are highly variable. Unlike class I where the peptide-binding domain is encoded by α_1 and α_2 domains in the same gene, trans arrangement of α and β chains derived from the two different haplotypes of the same or even different isotypes permits combinatorial polymorphism in class II. The α_2 and β_2 domains are proximal to the membrane and have limited polymorphisms (see Fig. 5.2).

Although the structure of the peptide-binding cleft in class II is homologous to that of class I, there are several distinct differences that have major functional consequences. Among the most important of these differences are those of length and cleft structure. The binding cleft of class I is closed, limiting peptide length, whereas the binding cleft of class II is open, which permits the peptide to extend on both sides of the class II molecule. Thus, the majority of peptides interacting with class II molecules have a length of greater than 13 amino acids, whereas class I prefers peptides of nine amino acids.

The peptide is bound to the class II molecule through the side chains of the peptide amino acids, which interact with five different polymorphic pockets within the cleft. Loading of the HLA class II molecules with peptides takes place primarily within the endosomes, where the HLA molecule interacts with endocytosed and phagocytosed extracellular antigens (Chapter 6). To prevent binding of intracellular peptides in the class II pocket, when the MHC molecule traffics through the endoplasmic reticulum it interacts with a protein termed invariant chain (Ii). Invariant chain is a trimer. Each of its subunits binds noncovalently with an HLA class II molecule. The MHC:invariant chain complex also interacts with another chaperon protein called calnexin. Upon release of calnexin, the class II molecule moves either directly into the late endosomal MHC class II compartment (MIIC) or is cycled to the cell surface where it is then internalized into MIIC. Once in the endosomal environment, invariant chain is degraded by proteases, including cathepsin S and L. It then leaves a fragment of peptide known as the class II-associated invariant chain peptide (CLIP). Upon dissociation of the CLIP peptide from the class II binding cleft within the endosome and with the assistance of HLA-DM, relevant exogenous peptide associates with the class II molecule prior to transport of the stable HLA class II–peptide complex to the cell surface.

Nonclassical Human Leukocyte Antigen -DM and -DO

The nonpolymorphic nonclassical class II molecules **HLA-DM** and **HLA-DO** are exclusively expressed in endosomes and regulate peptide binding to the classical HLA class II molecules.[11]

Proteasome Elements Within the Class II Region

The products of four genes in the class II region are involved with processing and loading peptides onto class I molecules (see Fig. 5.1). **PSMB8** and **PSMB9** are proteasome subunits generating peptides from the breaking down of proteins. **TAP1** and **TAP2** transport the peptides from the cytoplasm to the endoplasmic reticulum. The presence of these genes, which are related to the functioning of HLA class I molecules, in the midst of genes encoding the HLA class II molecules likely contributes to strong LD within the MHC. Allelic forms of genes in the class I region appear to require the presence of allelic forms in the class II region. Functional interdependencies promote joint transmission across generations.

Principles of Peptide Presentation

The mechanism by which HLA class I and class II molecules present peptides became clear when the structures of these two molecules were determined. A simplified cartoon of the domain structure of MHC class I and class II proteins is depicted in Fig. 5.2. A more intricate ribbon structure of the actual class I molecule interacting with the TCR is presented in Chapter 4. For both class I and class II, the peptide-binding structure takes the shape of a β-pleated floor with two α-helix walls. The peptide lies within the groove created by these structures (Figs. 5.3 and 5.4).

Each HLA molecule, whether class I or class II, binds a single peptide, but the same HLA molecule has a significant degree of promiscuity and can bind thousands of different peptides. Each of the binding grooves is composed of individual polymorphic pockets that dictate the binding of different peptides. Although the mode of TCR docking on HLA molecules is globally conserved, the shapes and chemical properties of the interacting surfaces found in these complexes are so diverse that no fixed pattern of contact has been recognized even between conserved TCR residues and conserved side chains of the HLA α helices.[12] Indeed, of the amino acid side chains not bound to the HLA, only two or three are typically bound to the clonotypic TCR. This limited contact yields considerable TCR plasticity, which has the important evolutionary implication of freeing the HLA molecule and the peptide–HLA complex from the strict stereochemical constraints usually imposed in receptor–ligand interactions. The consequence of TCR plasticity and this unusual receptor–ligand interaction has been the evolutionary development of a uniquely large number of different genes that encode various HLA structures, each of which is able to bind and present a different range of peptides to the same clonotypic TCR.

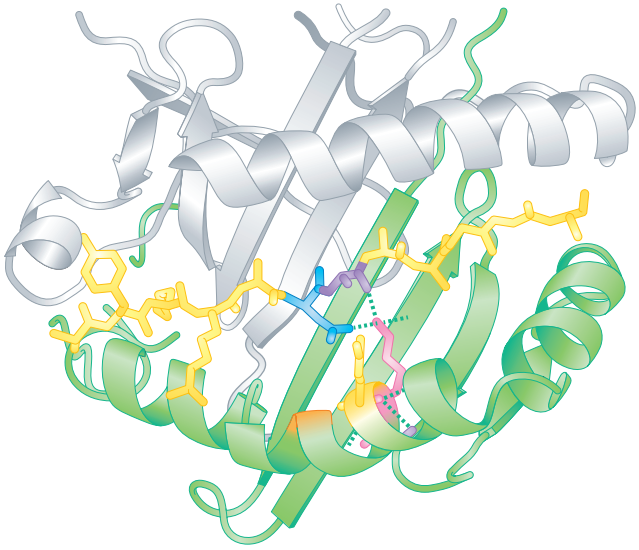


FIG. 5.4 Structure of a Human Leukocyte Antigen (HLA) Class II–Peptide Complex. The structure was prepared using PyMol from published coordinates. The HLA molecule is largely shown as a ribbon, while the peptide is a stick diagram. The peptide-binding groove is delimited by α helices. The upper helix is encoded by the α chain, and the lower helix by the β chain. β-pleated sheets form the saddle-like floor. Side chains are depicted on the β chain at positions 70 and 71, a region involved in specifying the side-chain pocket P4. This pocket binds the fourth side chain of the peptide contained within the HLA molecule. The side chains shown are, respectively, glutamine and lysine, which form part of the "shared epitope" structure associated with susceptibility to rheumatoid arthritis. The lysine is shown forming hydrogen bonds with the peptide antigen.

KEY CONCEPTS

Structure of the Human Leukocyte Antigen Molecules

- HLA class I molecules are involved in both innate (NK cells) and adaptive (T cells) immunity.
- Class I and class II HLA molecules share structures that enable them to bind peptides and present them to TCRs.
- Peptide binding to the HLA molecule is influenced by allele-specific pockets within the HLA binding cleft that interact with peptide amino acid side chains.
  - The HLA–peptide complex is recognized by the TCR of the T cell.

Selection by Self-Peptides in the Thymus

Peptides derived from external antigens, including pathogens, are typically absent during the formation of the T-cell repertoire in the thymus (Chapter 9). Self-peptide–HLA complexes are thus used as surrogates for selecting, or training, individual T cells to recognize nonself-pathogen peptides.[13] For T cells, "immunologic self" is the set of self-peptides and self-MHC molecules that select the TCR repertoire in the thymus. The two-step selection process begins with positive selection and survival of cells whose TCRs interact with the self-peptide–HLA complex. T cells with receptors that fail to recognize any self-peptide–HLA complex are eliminated. This is followed by a negative selection step whereby T cells with high-affinity interactions with self-peptides–HLA complex are eliminated. Negative selection promotes survival of T cells with lower-affinity interactions to self-peptide–HLA complexes. The surviving cells are released from the thymus and into the periphery. Selecting self-peptides expressed in the thymus constitute the T-cell recognition component of an individual's adaptive immune system. This patterning of TCR recognition on self-peptides presented by self-MHC molecules is critical to allorecognition and the avoidance of autoimmunity.

Evolutionary Considerations Driving the Separate Functions of Class I and Class II

One basic task of the T cell is to protect the body from two major types of pathogens: viruses that would commandeer the replicative machinery of a cell (Chapter 25), and bacteria that replicate autonomously and often extracellularly (Chapter 27). These two types of pathogens present very different challenges to the immune system. To terminate viral infection, a cell harboring virus has to be killed by a cytotoxic CD8 T cell (Chapter 12), whereas a bacterium can be eliminated by being phagocytized by a macrophage that has been selectively activated by a CD4 helper T cell (Chapter 11). The necessity of distinguishing between whether the presence of a pathogen peptide should elicit a killer or helper T-cell response is presumed to be the evolutionary drive that resulted in the creation of two specialized forms of HLA molecules, class I and class II (see Fig. 5.2).[12,14]

The specialized antigen processing and presentation intracellular machinery used to load class I molecules offers a means to reflect at the cell surface the molecular profile of antigens within the cell. This allows HLA class I molecules to screen for the presence of an intracellular viral infection. Recognition of the HLA class I–peptide complex is through the TCR of a CD8 T cell, which primarily reacts to the detection of an inappropriate intracellular antigen (i.e., a virus) by cytotoxic activity. In contrast, class II peptide loading occurs in coordination with phagocytosis and lysosomes. HLA class II molecules thus offer a means by which the immune system can be informed of the presence of extracellular antigens, such as bacteria. The recognition of the HLA class II–peptide complex is through the TCR of a CD4 T cell, which leads to the activation of helper T cells and the immune response that results.

Among the evolutionary strategies used for viral survival, some virally encoded genes decrease the expression of the HLA class I surveillance system, which would otherwise alert the immune system to the presence of an infected cell (Chapters 12 and 25).[15] This attempt to escape surveillance by downregulation of HLA class I is countered by the extensive interaction of class I molecules with various NK receptors expressed on NK cells or T-cell subsets. These interactions provide a mechanism for detecting decreases in HLA class I expression, which is termed recognition of "missing self."[16]

KEY CONCEPTS

Human Leukocyte Antigen Molecule Function

- An HLA molecule binds a peptide to create a peptide–HLA complex that serves as a ligand for a clonotypic TCR. The trimolecular peptide–HLA–TCR complex triggers the activation and proliferation of the T cell in an adaptive immune response.
  - HLA class I A, B, and C molecules are expressed on the surface of virtually all nucleated cells.
  - HLA class II DQ, DR, and DP molecules are constitutively expressed on B cells, professional antigen-presenting cells (APC), thymic epithelial cells, and activated T cells.
- The *immunological self* is the set of self-peptides and self-HLA molecules that select the TCR repertoire in the thymus. They constitute the T-cell recognition component of an individual's adaptive immune system.
- During an adaptive immune response, T cells recognize nonself-peptide–HLA complexes and become activated to either initiate an immune response (CD4 helper T cells) or recognize a target (CD8 cytotoxic T cells).
- Through thymic selection, the TCR can adapt to recognize a very large variety of peptide–HLA structures.
  - The plasticity of the recognition permits evolution of a large number of HLA genes encoding duplicated or alternative peptide-presenting molecules with specificity to bind different peptides.
  - The diversification of peptide-presenting structures fosters the development of different T-cell repertoires with completely different recognition properties. This thwarts the possibility that a pathogen will be able to evolve a way to bypass recognition.

GENERATION AND SELECTION OF POLYMORPHISMS; BIOLOGICAL CONSEQUENCES

The hallmark of HLA molecules, both class I and class II, is their extensive polymorphism. HLA polymorphism observed in different human populations is far greater than any other polymorphism observed in any other part of the human genome. This is a direct reflection of their role in the immune response. Pathogens characterized by different proteins and peptides, either in different epidemics or endemic to regions, account for much of the evolutionary drive responsible for the large number of alternative gene forms and their regional diversity across the human race. An individual with an adaptive immune system based on HLA molecules that effectively bind peptides derived from common pathogens is much more likely to have an effective response against that common pathogen. This results in selection of individuals with a particular HLA allele.

A *genetic polymorphism* implies that alleles of a gene are present at a frequency greater than expected from random mutation as a result of selection for diversity. In the case of the HLA genes, there is no preponderant wild-type allele, which would be an example of *balancing selection*. Instead, virtually all alleles qualify as genetically polymorphic. These reflect prior successful selection events. HLA polymorphisms provide a major evolutionary survival benefit, since they equip the species with a large number of very specific, but alternative, HLA molecules that differ in their binding pockets, are highly efficient in presenting different peptides, and select for different T-cell repertoires. A polymorphism that offers a survival advantage in a

given environment would eventually increase in frequency. This illustrates *frequency-dependent selection*, where the fitness of an individual bearing a particular allele increases, if it can manage an effective immune response to the particular pathogen.

Selection is a two-way street. It also operates on the pathogen, encouraging peptide variation. Variation in peptides drawn from common pathogens, and the introduction of novel pathogens with novel peptides, results in pressure on the species to create variation in HLA molecules among individual members of that species. The remarkably different frequency of the HLA alleles in different ethnic subsets tells the history of the successful adaptation of our ancestors' adaptive immune systems to a new environment with different pathogens, as well as bottlenecks resulting from migration and perhaps survival during periods of massive epidemics.

The evolutionary consequence of the diversification of genes encoding HLA molecules is seen at two levels. The first is at the level of the *individual* and is characterized by the presence of different HLA class I and class II loci, each of which codes for one or two different peptide-presenting HLA molecules for each locus. The second is at the level of the *population* and is evidenced by the development of a very large number of alleles at each locus, with each allele coding for alternative polymorphic gene forms and thus for various peptide-presenting allotypes, each of which has the potential to bind a different set of peptides. Duplication of HLA genes involved in peptide presentation is a genetic strategy that increases the range of peptide-presenting structures available to the individual, thus enhancing the variety of presented peptides that can be recognized and bound.[8]

KEY CONCEPTS

The Biological Significance of Polymorphisms; Why So Many?

- HLA class I and class II genes are extremely *polymorphic*.
  - Each HLA allele encodes molecules with different peptide-binding properties that influence the particular peptides recognized by the T cells.
  - The sequence of the HLA gene thereby determines the peptide recognition features of the adaptive immune response.
- HLA allelic polymorphisms are maintained by *frequency-dependent selection*, where the fitness of an individual bearing a novel allele increases as it can respond more effectively to certain pathogens.
- The multiple loci and numerous alleles per locus serve both the fitness of an individual and the survival of the species.
  - The polymorphism of the HLA system reflects the environmental/pathogen challenges to which a particular population has been exposed over evolutionary time.

HUMAN LEUKOCYTE ANTIGENS IN INFECTIONS, TRANSPLANTATION, AUTOIMMUNITY, AND CANCER

Human Leukocyte Antigen in Infections

The first line of defense during a pathogen infection is the triggering of innate immunity. The infectious agent and the foreign peptides generated from that agent then initiate an immune response involving immune cells and signals that subsequently induce adaptive immunity.

During the course of an infection, specialized APCs (dendritic cells, macrophages) are activated to take up antigen (Chapter 6). Increasing synthesis of class II coupled with presentation of a pathogen's peptides by class II to the immune system of the host activates CD4 T cells that recognize the HLA class II–peptide complex. This event triggers adaptive immunity. Eventually CD8 T cells recognize target cells infected with the pathogen by interacting with the HLA class I–peptide complex on their cell surface and the targets are eliminated, therefore containing the infection.

In response, pathogens have evolved mechanisms to overcome the specific attack by the host's immune cells. The first of these mechanisms is antigenic drift or shift, whereby the pathogen by minor (drift) or more substantive changes (shift) evade both humoral and cellular responses. These changes make the pathogen unrecognizable, as some of these new peptides do not form recognizable complexes with the HLA molecules of the host and therefore evade the T-cell responses. Another mechanism frequently adopted by viruses is to persist in vivo by not replicating until the immunity of the host is compromised. By not replicating, they avoid detection and they exist in a dormant state (latency). It therefore becomes evident that the infectivity of a microorganism reflects the interplay between several complex processes. These include the ability of the pathogen to create new molecular forms unrecognizable by the host and thus evade detection. These efforts by the pathogen to avoid immunity are then counterbalanced by molecular polymorphisms between HLA molecules that enable recognition of new molecular forms of the pathogen.[17]

Human Leukocyte Antigen in Transplantation

The large number of different HLA alleles greatly reduces the probability that two unrelated individuals will inherit an identical set of HLA alleles. In transplantation, two basic mechanisms of responses have been described (Chapter 89). The first involves the "direct" recognition of the peptide–HLA complex of the donor tissue by the T cells of the recipient. This is possibly through structural similarities of the HLA molecules of the donor that allow the TCR of the recipient to interact with the peptide–HLA complex. The second involves the "indirect" presentation of donor's HLA antigens processed by the recipient's APCs, generating peptides presented by the recipient's HLA molecules to the recipient's T cells. This indirect mechanism operates the same way as the presentation of a foreign antigen, whereby the HLA molecule is now the foreign antigen processed by the antigen-processing mechanisms of the recipient.

By using appropriate immunosuppressive agents and therapies, T-cell activation by the donor's HLA molecules after clinical transplant can be controlled. However, the major long-term problem is the presence of donor-specific antibodies that develop against mismatched HLA antigens. Controlling the antibody responses to mismatched HLA molecules has been very challenging and there is a need for continuous monitoring of their development. An approach holding some promise for the future is the utilization of regulatory T cells (Chapters 13 and 89), which have an important immunoregulatory role in all immune responses and can possibly induce transplant-specific tolerance.

Human Leukocyte Antigen in Autoimmunity

Selection of T cells on self-peptide presented by self-HLA allotypes in the thymus can predispose to autoimmunity. The inherent autoreactivity of the T-cell system can set the stage for the development of autoimmune diseases associated with the recognition of particular self-peptides, or peptides from external

KEY CONCEPTS

Human Leukocyte Antigen in Infections, Transplantation, Autoimmunity, and Cancer

- HLA and infectious agents participate in a balancing act: pathogens try to avoid the immune response and HLA alleles adapt to secure a robust immune response.
- Transplantation is an artificial system and the transplant is perceived by the immune response to be a foreign element.
  - Induction of tolerance is the objective.
  - Physicians promote immune nonresponsiveness to the transplant by pharmacologically manipulating the immune response.
  - Donor-specific antibodies are most commonly responsible for chronic rejections.
- Three features of the adaptive immune system can set the stage for pathogenic autoimmunity.
  - The TCR repertoire is selected by reactivity to self-peptides and self-HLA molecules.
  - The drive to genetic polymorphism generates alternative forms of peptide-binding HLA molecules that variably influence patterns of self- and nonself-reactivity.
  - Certain HLA allotypes bind particular self-peptides from critical target antigenic molecules that can predispose to autoimmune responses and disease.
- Oncogenesis is associated with the modification of patterns of antigen presentation by the cancerous cells to immune cells and by modification of immune cell responses to the cancerous cells.
- The cancer seeks to avoid immune surveillance and detection by the immune response.

antigens that mimic these self-peptides and are effectively presented by self-HLA.[18] Certain alleles encode HLA molecules that bind peptides from molecules expressed in sites favoring autoimmune recognition by T cells. These molecules become the target of the adaptive immune response. Together, features specific to certain sets of self-peptides and to certain self-HLA molecules can contribute to the progressive development of autoimmunity, and ultimately autoimmune disease.

Human Leukocyte Antigen in Cancer

Immune evasion is a critical process in tumor biology (Chapter 80). It is enabled by several mechanisms that include immunoediting, downregulation of HLA expression, secretion of immunosuppressive mediators, and expression of proteins that modulate immune checkpoints. Somatic mutation of HLA genes is a significantly frequent process in some tumor types. The strategies of immune evasion by cancer cells also include the silencing or aberrant expression of HLA class I and class II molecules, events that have often been associated with high-grade malignancy and metastatic potential in a variety of human cancers.[19]

In patients with solid tumors, HLA-G can contribute to a tumor-escape mechanism that favors cancer progression, and blocking strategies have been proposed to counteract it. Conversely, HLA-G can inhibit proliferation of malignant B cells due to the interaction between HLA-G and its receptor ILT2, which mediates negative signaling on B-cell proliferation. Thus, treatment of some malignancies can benefit by blocking HLA-G, whereas in others HLA-G induction can counteract tumor progression.[20]

The concept of developing cancer-specific immunotherapies involving tumor-specific antigens presented by HLA molecules to T cells has been successfully tested in a number of tumors, including testicular cancer and melanoma. These T-cell immunotherapies require adoptive transfer of T cells that have been expanded ex vivo and transferred back to the patient. Another approach is the use of retroviral vectors to transfer tumor-specific TCR genes into the patient's T cells before reinfusion.[21] Even though HLA molecules are involved in these processes, histocompatibility testing is not necessary in these therapies because the original T cells derive from the patient. However, if the mechanism of immunotherapy involves neoantigens (epitopes of mutated proteins) from tumors presented by specific HLA alleles, such individualized therapy needs to take HLA alleles into account.

Human Leukocyte Antigen Class I Molecules Regulate Natural Killer Cell Responses

The principal function of HLA class I molecules is the presentation of peptides that are expressed either by host-genome normally or dysregulated by tumorigenesis, or by foreign-genome derived from the infecting viruses and intracellular parasites. Cytotoxic CD8 T lymphocytes (CTLs) recognize specific HLA class I–peptide complexes via their TCRs and lyse if the HLA-laden peptides are derived from virus or tumor. If the cell-surface expression of HLA class I is downregulated as a consequence of some viral infection or tumor transformation, NK cells recognize and kill these HLA class I diminished abnormal cells.

NK cells are innate lymphoid cells programmed to kill target cells without a prior antigen "priming" period, as required for CTLs. Therefore, CTLs and NK cells serve as complementary killer cell components that control early immune responses to infection and tumor transformation. NK cells and CTLs are derived from a common lymphoid progenitor. They share several common features in development, morphology, cell-surface phenotypes, cytokine secretion, and lytic mechanism. However, they differ substantially in the tools used to recognize target cells: CTL use unique TCRs that are highly specific to self-HLA class I–peptide complex, while NK cells express an array of polymorphic receptors that bind self-HLA class I molecules.

Unlike TCRs that are generated by somatic genetic recombination, NK cells use a complex and sophisticated repertoire of activating and inhibitory receptors that are calibrated to ensure self-tolerance while exerting early assaults against viral infection and tumor transformation.[22] Human NK cell receptors include KLR, LILR, and KIR. The KIRs are the key receptors of human NK cells. Fourteen KIRs are identified: KIR2DL1, 2DL2, 2DL3, 2DL4, 2DL5, 3DL1, 3DL2, 3DL3, 3DS1, 2DS1, 2DS2, 2DS3, 2DS4, and 2DS5.[23] They are characterized by two or three immunoglobulin-like domains (2D or 3D; D denotes domain) in their extracellular portion and by either a long (L) or a short (S) cytoplasmic tail. Long tails are characteristic of the inhibitory forms and contain immunoreceptor tyrosine-based inhibition motifs (ITIMs) that trigger signals to switch off NK cell response (Chapters 3 and 12). The short-tailed activating KIRs lack ITIMs. Nevertheless, they have a positively charged amino acid residue in the transmembrane region that allows the interaction with the DAP-12, an adopter chain containing immunoreceptor tyrosine-based activation motifs (ITAMs), which trigger signals activating NK cell response.

The KIR gene family consists of 16 highly homologous genes clustered at the leukocyte receptor complex on chromosome 19.[24] Seven of them encode inhibitory KIRs (3DL1-3, 2DL1-3,

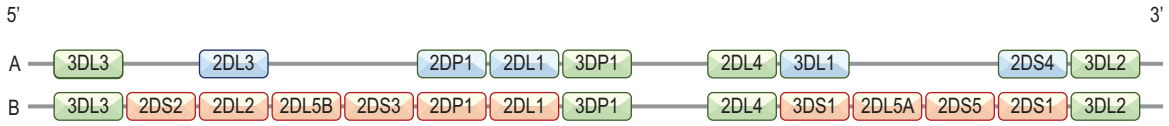


FIG. 5.5 Killer Cell Immunoglobulin-Like Receptor Haplotypes Vary in Gene Composition. Map of two core killer cell immunoglobulin-like receptor (KIR) haplotypes that differ quantitatively as well as qualitatively in gene content. Each box represents a KIR gene. The A haplotype comprises a set of fixed number of genes, most of which encode inhibitory KIR receptors. The B haplotype comprises more genes that are variable and mostly encode for activating KIR receptors. The framework genes conserved on both haplotypes are shown in *green*. A single chromosome may contain KIR genes belonging to either A or B haplotypes. All KIR genes are polymorphic at the nucleotide sequence level.

2DL5), six encode activating KIRs (3DS1, 2DS1-2DS5), one encodes a KIR that can trigger both inhibitory and activating signals (2DL4), and two are pseudogenes (2DP1 and 3DP1). The KIR genome displays a high degree of diversity determined by the variability in KIR gene content between haplotypes. The KIR haplotypes are broadly classified into two groups, A and B (Fig. 5.5). Group A haplotypes have a set of nine fixed genes (KIR3DL3-2DL3-2DP1-2DL1-3DP1-2DL4-3DL1-2DS4-3DL2) encoding for four inhibitory KIRs, 2DL1, 2DL3, 3DL1, and 3DL2, specific for four major HLA class I ligands, C2, C1, Bw4, and A3/A11 respectively. In contrast, the group B haplotypes are variable both in numbers and combinations of KIR genes, and comprising several genes (2DL2, 2DL5, 2DS1, 2DS2, 2DS3, 2DS5, 3DS1) that are not part of the A haplotype. While group A haplotypes contain only KIR2DS4 as an activating KIR, group B haplotypes contain up to five activating KIRs—KIR2DS1, 2DS2, 2DS3, 2DS5, and 3DS1. Random combinations of different gene-containing KIR haplotypes during familial inheritance produce substantial KIR genotype diversity in humans. The KIR A and B haplotypes are present in all human populations, but their frequencies vary considerably. In Africans and Caucasians, the A and B haplotypes are relatively in equal distribution, suggesting a balancing selection. Conversely, the A haplotype is overrepresented in Northeast Asians (Chinese, Japanese, and Koreans), while the B haplotype occurs most frequently in the natives of India, Australia, and America.[25]

In addition to haplotype diversity, each KIR gene exhibits sequence polymorphism, which is generally higher for the inhibitory KIRs than the activating KIRs. The KIR is considered to be second only to HLA in polymorphisms. These polymorphisms can influence cell-surface expression and alter receptor specificity and avidity, signal transduction, and cytokine secretion. The synergistic combination of allelic polymorphism and variable gene content individualize KIR genotypes to an extent where unrelated individuals almost always have different KIR types.

HLA-C is the prominent ligand for inhibitory KIR receptors. Half of the HLA-C allotypes have a lysine residue at position 80 that recognizes KIR2DL1. The remaining half of the HLA-C allotypes have asparagine 80 that binds to KIR2DL2 and 2DL3. KIR3DL1 binds to the Bw4 serological epitope, defined by amino acid residues 77 to 83, present on 40% of the HLA-B allotypes and certain HLA-A allotypes (HLA-A 23, 24, 25 and 32). KIR3DL2 binds certain HLA-A allotypes, such as A3 and A11. Very little is known about the ligands for the activating KIRs. Certain activating KIRs display a high degree of sequence homology with the corresponding inhibitory KIR in their extracellular Ig domains, and therefore activating KIRs would be expected to display a binding specificity similar to their inhibitory counterpart.

During NK cell development, interaction of inhibitory KIR receptor with cognate HLA class I ligand sets the functional threshold for NK cells, a process called "licensing."[26] Given that KIR genes at chromosome 19q13.4 and HLA genes at chromosome 6p21.3 are polymorphic and display significant variations, the independent segregation of these unlinked gene families produce diversity in the number and type of KIR-HLA combinations inherited in individuals,[27] which could potentially influence the health and disease status of a given individual. Consistent with this theory, combinations of certain KIR-HLA genes have been associated with diseases as diverse as autoimmunity, immune deficiency, infection, cancer, and reproductive failure.[28]

KEY CONCEPTS

Human Leukocyte Antigen Class I Molecules Regulate Natural Killer Cell Response

- NK cells arbitrate both innate and adaptive immunity. They are implicated in control and clearance of malignant and virally infected cells, and regulation of adaptive immune responses.
- Human NK cells use variable inhibitory and activating KIR receptors to discriminate between healthy and unhealthy cells.
  - Inhibitory KIRs recognize distinct HLA class I molecules and trigger signals that stop NK killing.
  - Activating KIRs presumably recognize determinants associated with infections and tumors, and trigger signals that activate NK killing.
- The effector function of a given NK cell depends upon the number and type of receptors it expresses and ligands it recognizes on the targets.
- Genes encoding KIRs and HLA ligands are located on different chromosomes and vary in number and gene type.
  - Independent segregation of KIR and HLA genes results in variable KIR-HLA combinations in individuals
  - Variable combinations influence an individual's immunity and susceptibility to a diverse array of diseases including autoimmunity, immune deficiency, infection cancer and reproductive failure.

HUMAN LEUKOCYTE ANTIGEN AND DISEASE ASSOCIATIONS

A large number of studies have established strong associations between certain diseases and individuals carrying particular HLA alleles. However, the mechanisms underlying HLA-disease associations remain unclear.

Hypotheses generated to explain these associations can be grouped into two general categories. The first invokes LD between a particular disease-associated HLA allele and another neighboring genomic element on the haplotype that is actually disease causative. Examples include hereditary hemochromatosis where association with HLA-A alleles is due to mutations in the *HFE* gene that is in LD with *HLA-A*, and congenital adrenal hyperplasia where association with *HLA-B* is due to linkage to a neighboring *CYP 21B* gene allele that causes 21-hydroxylase deficiency.

A second category implicates antigen presentation by the HLA allele. This category deals with diseases that have a strong immunological component. It has been hypothesized that inappropriate immune reactivity to some self-antigens can reflect aberrant T-cell repertoire selection, immune cross-reactivity with foreign antigens, immune attack of "altered self"-antigens, or differences in the expression levels of certain HLA alleles that secondarily influence the course of infections or cancer. MHC cusp theory represents an alternative hypothesis that HLA molecules promote disease due to their auxiliary allele-specific, yet antigen presentation-independent, biological effect.[29]

While many of these associations lie within the highly polymorphic HLA genes,[1] GWAS using SNP markers have established that the MHC region as a whole, and not only the HLA genes, harbors many SNPs associated with a large number of traits/diseases. Indeed, up to 90% of autoimmune disease variants have been located within noncoding regions of the genome.[30] It is therefore possible that disease-association elements may lie not only within the HLA genes but also dispersed within the rest of the MHC.

One possible genomic element that can be located within noncoding regions of the MHC and yet has a significant regulatory role is microRNA (miRNA). A search for functional genomic elements within the noncoding regions of the MHC genes revealed 12 miRNA, including hsa-miR-6891 (miR-6891) that is encoded by intron 4 of *HLA-B*.[31] Thus some, and perhaps many, diseases associated with specific MHC elements, whether HLA alleles or not, may involve noncoding RNAs (miRNAs or long noncoding RNAs) with important biological functions of a regulatory nature (Chapter 19).

Below is a compilation of selected diseases with strong HLA allele associations in different populations. A more extensive list of diseases and reference materials can be found elsewhere.[32–34]

Ankylosing Spondylitis

One of the more extraordinary observations in the MHC field was made in 1973 when the frequency of HLA specificity HLA-B27 was found to be 95% in patients with ankylosing spondylitis (AS), a disease characterized by arthritis affecting the spine and pelvis (Chapter 58). This observation implicated HLA-B*27 in the pathogenesis of AS and propelled the field of HLA and disease associations.[35] HLA-B*27:02 and B*27:05 demonstrate the highest degree of association, making genetic testing useful over and above serological testing. Even though the association of AS to B27 is among the strongest genetic associations with a common disease, the mechanism of action remains uncertain. Twin studies have confirmed that susceptibility to AS is genetically determined. HLA-B27 is found in 8% to 10% of the population, but only a minority of carriers develop the disease. Family studies suggest that less than 50% of the overall genetic risk is due to HLA-B27. A number of GWAS studies have demonstrated that non-HLA genes, including interleukin-23 receptor (IL-23R) and the protein-cleaving enzyme endoplasmic reticulum aminopeptidase 1 (ERAP1), also play a role.

B27 testing can be an instructive component of the diagnostic work-up of AS. Due to the chronicity of the disease and its gradual debilitating nature, a presumptive diagnosis based on B27 carrier status allows institution of treatment early in the disease when patients may have minimal symptoms.

Narcolepsy

Narcolepsy is a long-term neurological disorder characterized by irresistible daytime sleep attacks. These "sleep attacks" can occur at any time, and during any activity. Narcolepsy affects approximately 1 in 2000 people. Often those affected have low levels of the neurotransmitter hypocretin (also known as orexin). Hypocretin is a neuropeptide hormone that is responsible for controlling appetite and sleep patterns. Even though the cause of narcolepsy is unknown, the disease is believed to be of an autoimmune nature.

Family studies have shown that genetic heritability plays a role in narcolepsy. Twin studies show that only 25% to 30% of twins are concordant for the disease, again implicating environmental or other epigenetic events. The *HLA-DQB1*06:02* allele on the *DRB1*15:01-DQA1*01:02-DQB1*06:02* haplotype has been shown to be one of the most important predisposing genetic factors, with 85% to 95% of narcolepsy patients carrying this haplotype.[36] Conversely *DQB1*05:01* and *DQB1*06:01* have a protective effect. The protective associations of these two DQB1 alleles with narcolepsy may provide an insight to the molecular mechanisms for the differential associations of DQB1*06:01 and DQB1*06:02, as the size of P4 pocket of DQB1*06:02 is larger than the DQB1*06:01. This difference possibly influences the binding of larger residues in the DQB1*06:02 allele, which may explain the opposite effect these two alleles have on narcolepsy. Homozygosity for *HLA-DQB1*06:02* increases the risk for narcolepsy as compared to heterozygous persons, as does heterozygosity for HLA-*DQB1*03:01/DQB1*06:02*.

HLA testing for *DQB1*06:02* in narcolepsy is a useful aid to diagnosis. However, as instructive as the association may be, it is not specific as there are many narcolepsy patients without *HLA-DQB1*06:02* and many individuals with *HLA-DQB1*06:02* who do not have narcolepsy.

Type 1 Diabetes

Type 1 diabetes (T1D) is also known as insulin-dependent diabetes mellitus (IDDM) (Chapter 71). This is a disease in which the body fails to maintain normal glucose levels because of the destruction of insulin-producing pancreatic islet cells. The disease is characterized by infiltration of immune cells (CD4 and CD8 T cells) into the islets of the pancreas and by autoantibody production. When over 90% of an individual's beta cells are destroyed, clinical symptoms ensue.

Twin studies have shown that the concordance rate for the disease is 30% to 50%. This suggests that other factors, including environmental triggers (such as diet and viral infection) and epigenetic changes, may be involved. The major heritable risk of T1D comes from the HLA system (about 50%). More than 90% of Caucasian patients with T1D carry the haplotypes *DRB1*03:01, DQA1*05:01, DQB1*02:01* or *DRB1*04:01, DQA1*03:01, DQB1*03:02*. Patients heterozygous for these haplotypes carry a greater susceptibility risk. The critical residues are thought to be position 52 on the DQα chain and position 57 on the DQβ chain. The presence of arginine at position 52 on the DQα and the absence of aspartate on DQβ are strongly associated with

T1D. Conversely, in Caucasian populations resistance to T1D is conferred by *DQA1*01:02*, *DQB1*06:02*. Besides the contribution of the HLA, GWAS studies have identified a number of other genomic regions associated with the development of T1D.[37]

HLA typing is useful as an aid to diagnosis of T1D. Considering that the islet destruction by the autoimmune processes is progressive, associated with the presence of autoantibodies, HLA typing of siblings of T1D-affected patients may help assess risk for the non-symptomatic sibling.

Rheumatoid Arthritis

Rheumatoid arthritis (RA) is a chronic disease characterized by inflammation of the synovial lined joints leading to joint deformation and disabilities (Chapter 53). The presence of autoantibodies like rheumatoid factor and anti-cyclic citrullinated antibodies (ACPAs) are largely responsible for the autoimmune classification of this disease. It is a multifactorial disease that involves both environmental and genetic factors. RA prevalence in the general population is less than 1%. Studies with monozygotic twins show a 12% to 15% concordance rate for the disease, underlying genetic heritability, but also other factors such as environmental triggers or epigenetic components. Multiple genetic loci have been shown to contribute to the risk of developing RA. Of these, the HLA class II DRB1 is the most important and contributes 30% to 50% of the overall genetic susceptibility risk.

The HLA DRB1 alleles associated with RA share common sequences at positions 70 to 74 of the beta chain.[38,39] This has led to the shared epitope hypothesis. Amino acids in these positions influence both peptide binding and contact between HLA and TCR. HLA-DRB1 alleles associated with RA have any of the following sequences: QKRAA, QRRAA, RKRAA, and RRRAA. In Fig. 5.4, the yellow-colored residue in the α-helical ribbon is glutamine and the magenta residue is positively charged lysine. Hydrogen bonding to two side chains of the peptide is shown. The region around position 70 is involved in the formation of a peptide side-chain binding pocket that binds the fourth amino acid side chain contained within the HLA molecule. The presence of a negatively charged residue at position 71 or 74 removes susceptibility for RA. The presence of two alleles of this group increases susceptibility and favors development of more severe disease.[40]

GWAS have led to the identification of over 100 additional loci associated with RA. Among these is the protein tyrosine phosphatase, nonreceptor type-22 *(PTPN22)* gene, which codes for an inhibitor of T-cell activation. The majority of these additional loci are expression quantitative trait loci (eQTLs) in which genetic variants regulate the level of transcription.

Multiple Sclerosis

Multiple sclerosis (MS) is a complex neurodegenerative disease in which myelin sheath degradation is caused by the immune system (Chapter 66). Based on family and twin studies, the disease has been shown to have a large genetic component. The *HLA-DRB1*15:01*, *DQA1*01:02, DQB1*06:02* disease susceptibility haplotype accounts for up to 35% of the risk of developing the disease. A number of GWAS studies have identified more than 100 additional candidate genomic regions conferring risk, including cell adhesion, leukocyte activation, apoptosis, Janus kinase (JAK)-STAT signaling, nuclear factor-κB (NF-κB) activation, and T-cell activation and proliferation.[41] Although HLA typing is of minimal diagnostic value for MS, genetic testing may provide insights into the mechanism of the disease.

Celiac Disease

Celiac disease (CD) is an autoimmune disorder of the small intestine caused by a combination of genetic and environmental factors (Chapter 75). The disease is characterized by diarrhea and weight loss, among other symptoms. Monozygotic twins demonstrate 90% concordance, indicating a strong genetic component.[42] HLA genes contribute to about 40% of the genetic risk. GWAS studies implicate additional genomic regions. The environmental disease-triggering factor comes from a component of wheat gluten, the protein gliadin (family of closely related proline- and glutamine-rich proteins). CD is a lifelong condition. The only effective treatment is a gluten-free diet.

The implicated HLA molecules are the class II antigens DQ2 and DQ8. The DQ2 molecule mostly associated with CD is encoded by the *HLA-DQA1*05:01-DQB1*02:01* alleles, with a small proportion encoded by the *DQA1*02:01-DQB1*02:02* genotype. The DQ8 molecule associated with CD is *DQA1*03-DQB1*03:02*. Approximately 90% of patients with CD express HLA-DQ2, with the remaining 10% mostly expressing HLA-DQ8. Deaminated by transglutaminase, negatively charged gluten peptides bind strongly to HLA-DQ2 and -DQ8 to present an HLA-gluten peptide complex that activates CD4 T cells. The immune response also includes the development of antibodies against gluten and auto-antibodies to endogenous tissue transglutaminase.

As a complement to histology, genetic testing for HLA-DQ can help confirm the diagnosis in patients not known to be tissue-transglutaminase-antibody positive.

DRUG HYPERSENSITIVITY AND PHARMACOGENOMICS

Severe cutaneous adverse reactions to drugs include syndromes such as Stevens-Johnson syndrome/toxic epidermal necrolysis and drug reaction with eosinophilia and systemic symptoms or drug-induced hypersensitivity syndrome (Chapter 50).[43] Although their incidence is very low, they are severe, life-threatening adverse drug reactions with mortality rates as high as 5% to 12.5%. The associations reported between drug hypersensitivity and specific HLA alleles has been a recent finding and has led to the possibility that hypersensitivity reactions may be predictable and preventable. Drugs associated with immunologically mediated drug-induced hypersensitivity include the anticonvulsant carbamazepine and the antiretroviral agents nevirapine and abacavir. Regulatory agencies have issued relevant and informative pharmacogenomics guidelines: (http://www.fda.gov/Drugs/ScienceResearch/ResearchAreas/Pharmacogenetics/ucm083378.htm).

Carbamazepine

Carbamazepine is an aromatic amine anticonvulsant used for the treatment of epilepsy and other seizure disorders, trigeminal neuralgia, and bipolar disorder. Approximately 10% of patients develop mild cutaneous adverse reactions. Carbamazepine has been shown to be associated with class I *HLA-B*15:02* or *A*31:01*.

Nevirapine

Nevirapine is a non-nucleoside reverse transcriptase inhibitor widely prescribed for HIV-1 infection. Hypersensitivity HLA class I and II associations have been described for *DRB1*01:01, B*35:05, Cw8*, and *B*14:02*.

Abacavir

Abacavir belongs to the family of nucleoside reverse transcriptase inhibitors and is used for the treatment of HIV-1 infection. Two recent abacavir studies have shown that 100% of patients who develop abacavir drug hypersensitivity carry *the HLA-B*57:01* allele. Even though not all of *HLA-B57:01*-positive patients develop hypersensitivity, this predictive value supports the use of *HLA-B*57:01* typing prior to initiating abacavir treatment.

KEY CONCEPTS

Human Leukocyte Antigen and Disease Associations

- In multifactorial immune disorders, HLA alleles frequently confer a higher risk than other genomic factors.
- Most commonly the HLA molecules are directly involved in the disease process.
- Some associations reflect inheritance of an HLA haplotype where the HLA gene is in LD with other, causative, non-HLA genes.
- In some cases, the HLA molecule, the associated peptide, and the TCR are sufficient for disease development.
- In others, the HLA molecule may be necessary, but not sufficient for disease development.
- Twin studies showing less than 100% disease concordance suggest that in addition to the HLA genes, environmental of metagenomic modifications are also likely involved in many HLA-associated disorders.

METHODS OF DETECTING HUMAN LEUKOCYTE ANTIGEN POLYMORPHISMS. THE HUMAN LEUKOCYTE ANTIGEN TYPING PROBLEM

Since the discovery of the HLA genes over 50 years ago, there has been a concerted effort to properly categorize and characterize these very polymorphic genes. Our understanding of the complexity and polymorphic nature of the HLA genes has been substantially improved as the technologies for characterizing these genes have improved.

Serological and cellular testing in the 1960s (antibody and mixed lymphocyte culture [MLC]) was supplemented by two-dimensional electrophoresis and restriction fragment length polymorphism (RFLP) analysis in the 1970s and early 80s. The development of the polymerase chain reaction (PCR) in the mid-1980s revolutionized our understanding of these genes. Methods utilizing sequence-specific oligonucleotide probes (SSOP or SSO) and sequence-specific primers (SSPs) provided the means for further evaluating the highly variable sequence motifs within the HLA genes. Sanger sequencing-based typing (SBT) in the 1990s advanced tissue typing and transplantation genetics by providing an unprecedented molecular view of HLA polymorphism in the context of exonic variation. Most recently, NGS provides entire HLA gene characterization and haploid sequence determination.[44]

To meet the growing demand, clinical HLA typing over the past decade has transitioned from a combination of serological and DNA-based methods to more direct, faster, affordable, and informative DNA-based techniques. Even though serological typing may continue to have some clinical or research-based testing in determining the expression of the HLA molecule at the cell surface (a function that DNA-based testing cannot always verify), direct DNA-based typing techniques have all but replaced serological methods in routine HLA typing.

DNA-Based Typing Techniques: Sequence-Specific Oligonucleotide Probes, Sequence-Specific Primer, and Sequencing-Based Typing

The techniques primarily in use today in clinical immunogenetics laboratories are SSO, SSP, and SBT. The genomic regions analyzed are usually exon 2 and 3 of class I and exon 2 of class II genes. However, this rather limited genomic characterization generates many typing ambiguities.

SSO interrogates polymorphic differences using panels of individual DNA oligo probes that differentially hybridize to the target of interest. The probe either perfectly matches or mismatches the target's polymorphic sites. Drawing upon the sequence database of HLA alleles, the hybridization pattern of the oligoes is compared to an expected pattern and is interpreted as an HLA type.

SSP uses panels of specific primer sets that overlap polymorphic sites. Perfectly matched primers produce an amplification product while mismatched primers do not. The pattern of amplification from multiple primer sets determines the HLA allele.

SBT amplifies and sequences specific gene regions, usually exons, through a process of polymerase-based extension of specific sequencing primers. It uses fluorescently labeled nucleotides that indicate allelic differences base by base.

Next-Generation Sequencing and Its Potential Impact on Human Leukocyte Antigen Typing

Protocols utilizing NGS technology are on the rise as they provide the means for the complete characterization of these genes and the elimination of ambiguities in a cost-efficient manner.[45] Regardless of the platform (Illumina MiSeq; Thermo Fischer Ion Torrent; Pacific Biosciences, Oxford Nanopore), these systems resolve a number of technological barriers that continue to hamper existing molecular techniques, such as inflexibility in typing practices, discovering novel alleles, and the inability to easily resolve phase ambiguities. While HLA typing by NGS has already been adopted by several labs, it is likely that this new method will transform the way HLA typing is performed in the coming years.

KEY CONCEPTS

The Resolution of the Human Leukocyte Antigen Typing Problem

- Currently, HLA typing is primarily performed through DNA-based methodologies.
- The coming dominant methodology most likely will be single molecule DNA sequencing (NGS) the length of each class I or class II gene.

HUMAN LEUKOCYTE ANTIGEN NOMENCLATURE

The HLA genes are very polymorphic (over 27,000 named and close to 30,000 sequence entries that have not been named yet) and expected to further increase, approaching hundreds of thousands, possibly millions. This has led to the development of comprehensive systems for their naming.

The WHO Nomenclature Committee for Factors of the HLA system undertook the first systematic approach for the naming of the HLA alleles in 1968. The HLA naming convention has undergone substantial iteration as earlier naming conventions were unable to address the growing numbers and complexity of alleles (i.e., A*02 and B*15 have more than 100 alleles). The most recent nomenclature was introduced in 2010. Colons (:) were added into the allele names to act as delimiters of the separate fields (field separator). Thus, each HLA allele name has a unique number corresponding to up to four sets of digits separated by colons (Figs. 5.6 and 5.7).

The first field following the asterisk in the allele name (**XX**:xx:xx:xx) describes the allele family and generally corresponds to the serological assignment carried by the allele. HLA typing defined only at the first field is often referred to as "low resolution typing." The second field following the first colon (xx:**XX**:xx:xx) is assigned sequentially as new alleles are determined (e.g., 01, 02, 03,….101, etc.). Together, these two fields (**XX**:**XX**) indicate one or more nucleotide substitutions that change the HLA protein coding sequence and are often referred to as "high resolution typing." Indeed, the Harmonization of Histocompatibility Typing Terms Working Group recently defined a high-resolution typing result "as a set of alleles that encode the same protein sequence for the region of the HLA molecule called the antigen-binding site and that exclude alleles that are not expressed as cell-surface proteins."[46] The third field (xx:xx:**XX**:xx) is for designating synonymous nucleotide substitutions within the coding sequence that do not change the amino acids of the protein, while the fourth field (xx:xx:xx:**XX**) identifies sequence polymorphisms in introns, or in the 5′ and 3′untranslated regions.

All alleles receive at least a four-digit name, which corresponds to the first two sets of fields. At the end of the allele name, specific characters have been added (N=null, L=low expression, S=secreted, C=cytoplasm, A=aberrant, Q=questionable) to

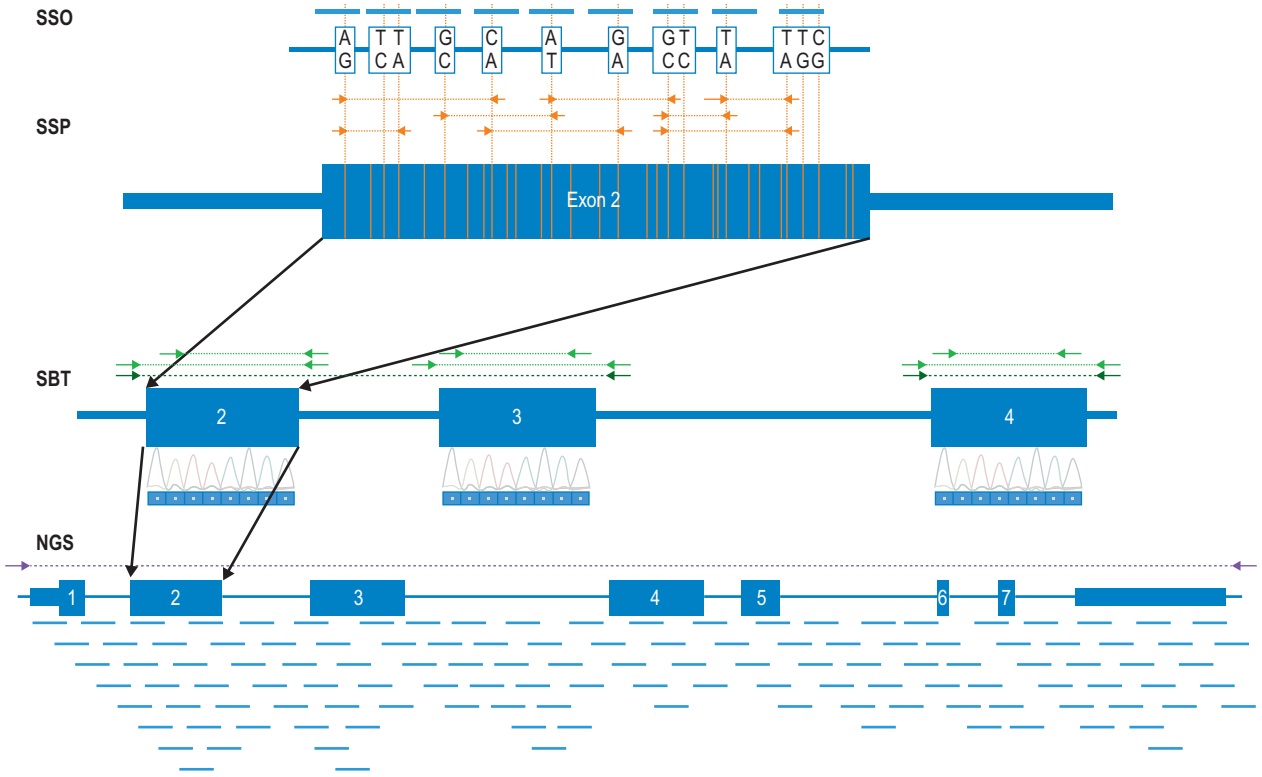


FIG. 5.6 Examples of Molecular Human Leukocyte Antigen Typing Techniques and Their Methods of Interrogating the Human Leukocyte Antigen Genes. For any given human leukocyte antigen (HLA) gene (*dark blue rectangles*), sequence-specific oligonucleotide probes (*SSOs*) of approximately 20 bp (*light blue lines*) can provide single-nucleotide resolution of haplotype differences (polymorphic differences, *red lines* in exon 2). This requires a complex panel of oligonucleotide probes to discern differences between HLA alleles. This probe set is static and therefore cannot adjust to novel alleles. Sequence-specific primers (*SSPs*) *(orange arrows)* can provide haplotype- or allele-specific resolution of nucleotide differences and additionally provide some level of phasing between polymorphic sites. As with SSOs, these oligonucleotide sets are complex and static, limiting their flexibility. Sequencing-based typing (*SBT*) provides whole-exon information on the polymorphic content of the HLA allele (amplification primers *[dark green]* and sequencing primers *[light green arrows]*) but cannot discern phasing, as this method generally does not rely on allele-specific primers for amplification as a first step. Next-generation sequencing (*NGS*) provides whole-gene amplification (amplification primers, *purple arrows*) and detection of polymorphic content for any HLA allele (known or unknown) and provides significant phasing between polymorphic sites that are within the read lengths of the system being used (usually between 200 and 1000 bp). This is accomplished through the alignment of thousands of short overlapping reads that are combined to form a single consensus sequence *(blue lines)*.

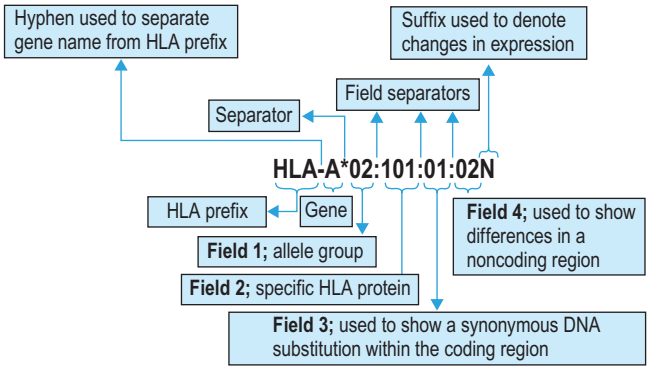


FIG. 5.7 Human Leukocyte Antigen *(HLA)* Nomenclature. (Courtesy Steven G. E. Marsh, Anthony Nolan Research Institute, London, United Kingdom.)

designate unique characteristics for an allele such as whether a protein is not expressed (i.e., HLA-A*24:09N) or whether the expression of the protein is unclear (i.e., HLA-A*32:11Q).[44] For ambiguous allele strings, the codes "P" and "G" were introduced. A group of alleles having nucleotide sequences that encode the same protein sequence for the peptide-binding domains (exon 2 and 3 for HLA class I and exon 2 only for HLA class II alleles) are designated by an upper case "P" that follows the 2-field allele designation of the lowest numbered allele in the group. For example, *HLA-A*01:01:01:01, HLA-A*01:01:01:03*, or *HLA-A*01:37* could be named *HLA-A*01:01P*.

A group of alleles that have identical nucleotide sequences across the exons encoding the peptide-binding domains (exons 2 and 3 for HLA class I and exon 2 for HLA class II) were named after the first allele in the sequence and given a code of "G" as a suffix. The upper case "G" follows the first three fields of the allele designation. For example, *HLA-A*01:01:01:01, HLA-A*01:01:01:03*, or *HLA-A*01:37* could be named *HLA-A*01:01:01G*. More details regarding HLA nomenclature can be found in the site http//hla.alleles.org.

In order to manage and to have access to the sequences of the ever-growing number of alleles, the IMGT/HLA Database project was initiated in 1997 as part of a European collaboration. The database is an invaluable resource, as it provides detailed DNA sequences and protein sequences for all known HLA alleles. It is also interactive as it incorporates tools for data retrieval and analysis so the user can select what segments of the gene/molecule to examine and compare among different alleles. It can also be used for new data submission.

ON THE HORIZON

- Advanced technologies for detailed characterization of the whole MHC should clarify and define functional interrelationships between MHC genes and the genomic elements responsible for many MHC-associated diseases.
- Computational approaches for the accurate definition of peptide-binding properties of individual HLA alleles will enable prediction or manipulation of the trimolecular complex of HLA–peptide–TCR in order to control responses to infectious diseases, autoimmunity, transplantation, vaccine design, and cancer.
- Immunotherapies for cancer involving neoantigens and HLA alleles will be individualized for precision.
- Understanding the genomic organization of the exceedingly complex MHC will most likely reveal and teach us important lessons relevant to the organization and operation of the rest of the genome as well.

FUTURE LEARNING AND RESOURCES

This chapter provides only a limited sketch of this fascinating, but complex, topic. The reader is referred to the *HLA Facts Book* for a more detailed and very accessible presentation, though slightly out of date. There are also a number of websites with extremely useful information. Four stand out in terms of their utility and the curated quality of the information. The IMGT/HLA Database contains all MHC sequences and has a variety of sequence alignments of different alleles as well as specialized sequence searches (http://www.ebi.ac.uk/imgt/hla/index.html and http://hla.alleles.org). The NCBI maintains dbMHC, which includes several components of the International Histocompatibility Working Group (IHWG) that are of interest. Among these are the anthropology database that contains HLA class I and class II allele and haplotype frequencies in various human populations (http://www.ncbi.nlm.nih.gov/projects/mhc/). Information about the genes and the genetic organization of the MHC is contained in several sites, but perhaps the most comprehensive and comprehensible is that using the Entrez search engine (http://www.ncbi.nlm.nih.gov). A comprehensive database of MHC ligands and peptide motifs is located at http://www.syfpeithi.de.

REFERENCES

1. Shiina T, Hosomichi K, Inoko H, Kulski JK. The HLA genomic loci map: expression, interaction, diversity and disease. *J Hum Genet*. 2009;54(1):15–39.
2. Aken BL, Ayling S, Barrell D, et al. The Ensembl gene annotation system. *Database (Oxford)*. 2016;2016 baw093.
3. Clark PM, Kunkel M, Monos DS. The dichotomy between disease phenotype databases and the implications for understanding complex diseases involving the major histocompatibility complex. *Int J Immunogenet*. 2015;42(6):413–422.
4. Dapprich J, Ferriola D, Mackiewicz K, et al. The next generation of target capture technologies—large DNA fragment enrichment and sequencing determines regional genomic variation of high complexity. *BMC Genomics*. 2016;17:486.
5. Kelley J, Walter L, Trowsdale J. Comparative genomics of natural killer cell receptor gene clusters. *PLoS Genet*. 2005;1(2):129–139.
6. Parham P. Immunogenetics of killer cell immunoglobulin-like receptors. *Mol Immunol*. 2005;42(4):459–462.
7. Lund O, Nielsen M, Kesmir C, et al. Definition of supertypes for HLA molecules using clustering of specificity matrices. *Immunogenetics*. 2004;55(12):797–810.
8. Beck S, Trowsdale J. The human major histocompatability complex: lessons from the DNA sequence. *Annu Rev Genomics Hum Genet*. 2000;1:117–137.
9. Lanier LL. On guard—activating NK cell receptors. *Nat Immunol*. 2001;2(1):23–27.
10. Gomez-Prieto P, Parga-Lozano C, Rey D, et al. HLA-G, -F and -E: polymorphism, function and evolution. In: Mehra NK, ed. *The HLA Complex in Biology and Medicine: A Resource Book*. Jaypee Brothers Medical Publishers Ltd. New Delhi, India; 2010:159–174.
11. Karlsson L. DM and DO shape the repertoire of peptide-MHC-class-II complexes. *Curr Opin Immunol*. 2005;17(1):65–70.
12. Housset D, Malissen B. What do TCR-pMHC crystal structures teach us about MHC restriction and alloreactivity? *Trends Immunol*. 2003;24(8):429–437.
13. Stefanova I, Dorfman JR, Tsukamoto M, Germain RN. On the role of self-recognition in T cell responses to foreign antigen. *Immunol Rev*. 2003;191:97–106.
14. Trowsdale J, Parham P. Mini-review: defense strategies and immunity-related genes. *Eur J Immunol*. 2004;34(1):7–17.

15. Lilley BN, Ploegh HL. Viral modulation of antigen presentation: manipulation of cellular targets in the ER and beyond. *Immunol Rev*. 2005;207:126–144.
16. Raulet DH. Missing self recognition and self tolerance of natural killer (NK) cells. *Semin Immunol*. 2006;18(3):145–150.
17. Murphy K, ed. *Janeway's Immunobiology*. 8th ed. : Garland Science, New York, NY, USA; 2011.
18. Winchester R. The genetics of autoimmune-mediated rheumatic diseases: clinical and biologic implications. *Rheum Dis Clin North Am*. 2004;30(1):213–227. viii.
19. Vinay DS, Ryan EP, Pawelec G, et al. Immune evasion in cancer: mechanistic basis and therapeutic strategies. *Semin Cancer Biol*. 2015(35 suppl):S185–S198.
20. Rouas-Freiss N, Moreau P, LeMaoult J, Carosella ED. The dual role of HLA-G in cancer. *J Immunol Res*. 2014;2014:359748.
21. Wurz GT, Kao C-J, DeGregorio MW. Novel cancer antigens for personalized immunotherapies: latest evidence and clinical potential. *Ther Adv Med Oncol*. 2016;8(1):4–31.
22. Lanier LL. NK cell recognition. *Annu Rev Immunol*. 2005;23:225–274.
23. Parham P, Moffett A. Variable NK cell receptors and their MHC class I ligands in immunity, reproduction and human evolution. *Nat Rev Immunol*. 2013;13(2):133–144.
24. Wilson MJ, Torkar M, Haude A, et al. Plasticity in the organization and sequences of human KIR/ILT gene families. *Proc Natl Acad Sci USA*. 2000;97(9):4778–4783.
25. Rajalingam R, Du Z, Meenagh A, et al. Distinct diversity of KIR genes in three southern Indian populations: comparison with world populations revealed a link between KIR gene content and pre-historic human migrations. *Immunogenetics*. 2008;60(5):207–217.
26. Kim S, Poursine-Laurent J, Truscott SM, et al. Licensing of natural killer cells by host major histocompatibility complex class I molecules. *Nature*. 2005;436(7051):709–713.
27. Du Z, Gjertson DW, Reed EF, Rajalingam R. Receptor-ligand analyses define minimal killer cell Ig-like receptor (KIR) in humans. *Immunogenetics*. 2007;59(1):1–15.
28. Khakoo SI, Carrington M. KIR and disease: a model system or system of models? *Immunol Rev*. 2006;214:186–201.
29. Holoshitz J. The quest for better understanding of HLA-disease association: scenes from a road less travelled by. *Discov Med*. 2013;16(87):93–101.
30. Farh KK-H, Marson A, Zhu J, et al. Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature*. 2015;518(7539):337–343.
31. Ladewig E, Okamura K, Flynt AS, et al. Discovery of hundreds of mirtrons in mouse and human small RNA data. *Genome Res*. 2012;22(9):1634–1645.
32. Mehra NK. *The HLA Complex in Biology and Medicine: A Resource Book*. Jaypee Brothers Medical Publishers Ltd. New Delhi, India; 2010.
33. Lechler R, Warrens A, eds. *HLA in Health and Disease*. 2nd ed. Academic Press, London, United Kingdom; 2000.
34. Tiwari JL, Terasaki PI. *HLA and Disease Associations*. Springer-Verlag, Berlin, Germany; 1985.
35. Schlosstein L, Terasaki PI, Bluestone R, Pearson CM. High association of an HL-A antigen, W27, with ankylosing spondylitis. *N Engl J Med*. 1973;288(14):704–706.
36. Mignot E. Genetics of narcolepsy and other sleep disorders. *Am J Hum Genet*. 1997;60(6):1289–1302.
37. Bradfield JP, Qu H-Q, Wang K, et al. A genome-wide meta-analysis of six type 1 diabetes cohorts identifies multiple associated loci. *PLoS Genet*. 2011;7(9):e1002293.
38. Winchester R. The molecular basis of susceptibility to rheumatoid arthritis. *Adv Immunol*. 1994;56:389–466.
39. Goronzy JJ, Weyand CM. Rheumatoid arthritis. *Immunol Rev*. 2005;204:55–73.
40. Michou L, Croiseau P, Petit-Teixeira E, et al. Validation of the reshaped shared epitope HLA-DRB1 classification in rheumatoid arthritis. *Arthritis Res Ther*. 2006;8(3):R79.
41. Hussman JP, Beecham AH, Schmidt M, et al. GWAS analysis implicates NF-κB-mediated induction of inflammatory T cells in multiple sclerosis. *Genes Immun*. 2016;17(5):305–312.
42. Greco L, Romino R, Coto I, et al. The first large population based twin study of coeliac disease. *Gut*. 2002;50(5):624–628.
43. Cheng C-Y, Su S-C, Chen C-H, et al. HLA associations and clinical implications in T-cell mediated drug hypersensitivity reactions: an updated review. *J Immunol Res*. 2014;2014:565320.
44. Detrick B, Hamilton RG, Schmitz JL, eds. *Manual of Molecular and Clinical Laboratory Immunology*. 8th ed. ASM Press, Washington, DC, USA; 2016.
45. Duke JL, Lind C, Mackiewicz K, et al. Determining performance characteristics of an NGS-based HLA typing method for clinical applications. *HLA*. 2016;87(3):141–152.
46. Nunes E, Heslop H, Fernandez-Vina M, et al. Definitions of histocompatibility typing terms: Harmonization of Histocompatibility Typing Terms Working Group. *Hum Immunol*. 2011;72(12):1214–1216.

6

Antigens and Antigen Presentation

Harry W. Schroeder Jr. and Robert R. Rich

ANTIGENS

By the late 19th century, "antibodies" were the hypothesized molecular entities that mediated specific immune memory, could neutralize toxins, and could form precipitates when mixed with the molecular species that induced their formation. In almost all cases, evidence for the presence of such antibodies required the prior exposure of responding individuals to the very substances (or ones closely related, as in the case of toxoids) with which the antibodies reacted. This *specific* relationship of inducing agent and antibody, which is the basis of antigen-specific *adaptive immunity*, led to the concept of an *antigen* as that molecular entity that could induce, in the blood of exposed individuals, the formation of antibodies specific for it. By inventing the concept of the specific *receptor*, with a specificity analogous to the lock-and-key model of enzymes, Paul Ehrlich could explain the specificity of antibodies in molecular terms of a reciprocal interaction between a receptor and its binding partner (*ligand*).[1] This explanation defines a fundamental relationship. An "antigen" is any molecule that binds specifically to the antigen-binding domain of an "antigen receptor" (be it antibody or T-cell receptor—TCR [Chapter 4]). This concept of "antigen" is useful for understanding the molecular basis of recognition and binding and for the practical use of the scientist or pharmacologist who is focused on the use of the antigen/antigen receptor interaction to detect, or purify, or neutralize a molecule of interest.

At the level of the organism, the concept of antigen focuses on its other critical property; that it *can induce* the proliferation and differentiation of the lymphocytes that bear its specific receptor and that it can promote the production of the antibodies that bind it. This view of "antigen" is that of the vaccinologist who wants to induce effective immunity to an organism expressing that antigen, or of the clinician who seeks to understand why a patient does or does not respond to a particular allergen, self-antigen, or tumor antigen. More than a century later, explaining the antigenicity of antigens (i.e., the response at the cellular and organismal level to a given molecule) is one of the fundamental issues in immunology. Why some or all individuals may fail to respond adequately to some pathogen or tumor antigens, why others respond detrimentally to our own self-antigens (autoantigens) or antigens present in tissue grafts (alloantigens), how we can improve the efficacy of vaccines, and how we can prevent or treat autoimmune and tissue-graft-related diseases (graft-*vs.*-host disease [GVHD] and graft rejection) all remain open questions.

Advances in the domain of *innate immunity* have led to the need to distinguish between the antigens that elicit adaptive immunity and the ligands that bind innate immune receptors (Chapter 3). Innate receptor ligands are often described as exhibiting patterns or motifs characteristic of a microbial class or physiological condition. To capture these notions, Janeway[2] and Matzinger[3] coined the terms "pathogen-associated molecular patterns" (PAMPs) and "danger signals," respectively. A prototypical innate ligand is lipopolysaccharide (endotoxin), which is produced by many bacteria and serves as a ligand for Toll-like receptor 4 (TLR4) (Chapter 3).[4] Another is dsRNA, which is an obligatory intermediate in RNA virus replication and serves as a ligand for TLR3. Both have features that are characteristic of a pathogen class. However, a variety of ligands bind to TLR4 that have no obvious "motif" shared with LPS. Moreover, danger signals also do not exhibit an obvious "motif" but instead can be merely characteristic of a physiological state. For example, the receptor for advanced glycan intermediates, RAGE, is also a receptor for HMGB1, a nuclear transcription factor released to function as an inflammatory cytokine by macrophages.[5] Thus, it is problematic to define innate receptor ligands in terms of intrinsic properties. Moreover, ligands for innate receptors can also serve as antigens for adaptive antigen receptors. Consequently, the conceptual difference between antigens and innate receptor ligands depends not on the intrinsic properties of the ligands, but on the properties of the receptors to which they bind.

It has become a cottage industry to reveal "bridges" between innate and acquired immunity, blurring the initial clear-cut distinctions between the two. Moreover, we see a continual expansion of the function of innate receptors as registrars of states of internal stress (Matzinger's "danger") rather than primarily as monitors of external threats (Janeway's "stranger"). It is thus useful to see the categories of innate and acquired immune receptors as a continuum rather than as essentially distinct and in need of "bridges."

At one extreme, purely innate receptors are expressed constitutively among tissues and over time. They are present in the "ground state" of the organism and thus are innate. They function like most other receptors of the body to respond homeostatically to perturbations in the internal milieu of the organism, particularly to an experience of physiologic stress to immunologic homeostasis, but also to a molecular threat of stress as flagged by microbial products.

At the other extreme, the acquired receptors of the adaptive immune system are pleomorphic rather than unimorphic and inducible rather than constitutive. In the case of the two acquired immune receptors defined in jawed vertebrates, antibodies and TCR, the induction is mediated by irreversible changes in the DNA sequence encoding them (Chapter 4).[6] The function of acquired immune receptors (antigen receptors) is to record exposure to an inciting antigen and thus mediate specific immune memory, which permits the "faster, stronger" reaction of a secondary immune response. In particular, many T- and

B-cell responses are accompanied by the rapid proliferation of those cells expressing unique and specific antigen receptors (clonal selection and expansion) (Chapters 7, 9, and 10).

In so-called "bridge" mechanisms, we see aspects of short-term memory (often called "priming") affected by innate mechanisms as well as acquired mechanisms that fail to exhibit memory. For example, pre-exposure to activation of certain Toll-like receptors can lead to enhanced responses through the same or other innate receptors over a period of half a day. Conversely, some T- and B-cell responses exhibit a strong "primary" response to antigen without evidence of an enhanced memory response. Moreover, priming through innate receptors serves as a critical mechanism for enhancing certain aspects of acquired immunity.

This chapter is organized around five themes: *antigen*, how antigens are manipulated by cellular and enzymatic machinery to permit recognition by antigen receptors (*antigen acquisition, processing*, and *presentation*), and, finally, the *antigen-presenting cells* (APCs) themselves. A central function of APC is to present antigens to antigen-receptors on lymphocytes (signal 1) but also to provide costimulatory signals (signal 2) and regulatory signals (signal 3) to those lymphocytes.

Antigens in the sense of *ligand* are defined tautologically as the ligands for antigen receptors (Fig. 6.1). This definition includes the *acquired* antigen receptors found on the surface of B cells (B-cell receptors [BCRs], also known as membrane-bound immunoglobulin [mIg]) and TCR. The closely related category of lymphocytes, natural killer cells, also bear receptors that can recognize antigens, but unlike B cells and T cells, these receptors are encoded in the germline rather than being acquired during development (Chapters 2 and 3).[7]

It is important to distinguish the antigen receptors just described from class I and class II major histocompatibility complex (MHC) *antigen-presenting molecules* (Chapter 5). MHC molecules bind short peptides (*oligopeptides*) and certain other molecules and present them to the TCR on T cells and, in some cases, to innate immune receptors on natural killer cells. MHC molecules themselves are innate receptors in that they are encoded in the germline, and their expression is regulated homeostatically.

The antigen bound by a particular antigen receptor is sometimes called its *cognate antigen*. This makes sense because of the allied concept that the great majority of lymphocytes express only a single antigen receptor due to the mechanism of allelic exclusion, with singular specificity for its own cognate antigen (Chapter 4). This concept is useful even if up to 5% of lymphocytes actually express more than one receptor. Because lymphocytes retain expression of that singular receptor when they divide, we can identify clones of that recognize the same cognate antigen. In turn, lymphocyte clones that bear different antigen receptors can recognize different aspects of the same cognate antigen.

Whether or not a particular molecule will serve as cognate antigen for any receptor depends on many factors. Because of the stochastic mechanisms used to generate the antigen-binding site of forming antigen-receptors (e.g., N region addition [Chapter 4]) and the relatively short life span of most naïve lymphocytes, there is a real possibility that many potential antigens, especially those present at low concentrations, will never encounter a cognate receptor in the lifetime of the individual. Alternatively, the antigen might be sequestered within the cell or body in such a way as to escape detection by lymphocytes.

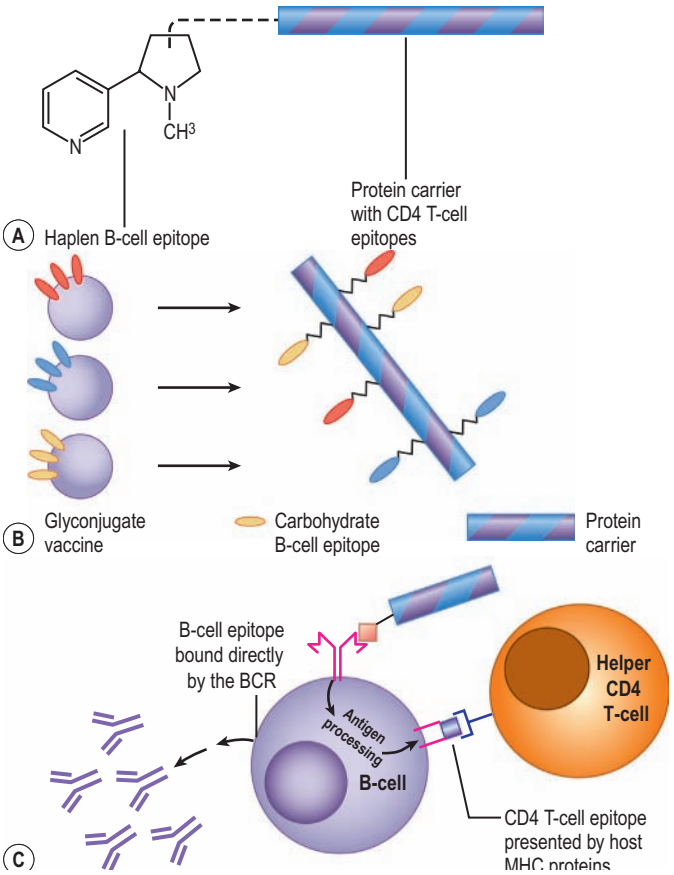


FIG. 6.1 Haptens, Carriers, and Two Kinds of Antigens. (A to C) The antigen-binding site of an antibody binds an antigen through the latter's epitope: this is the biochemical sense of antigen used in ELISA, flow cytometry, and western blot analysis. Haptens are self-conjugating antigen moieties that can modify epitopes and provide new binding specificities. Haptens and many antigens by themselves are not immunogens, the second sense of "antigen." Immunogens (complete antigens) are processed by antigen-presenting cells to reveal T-cell epitopes presented by MHC molecules. *MHC,* Major histocompatibility complex.

Such antigens are sometimes called *cryptic antigens*. Many *self-antigens* are recognized by developing T cells in the thymus or developing B cells in the primary or secondary lymphoid organs, causing the clonal deletion or the development of anergy in responding lymphocytes (Chapter 10). Only a few of these self-antigens escape these tolerance mechanisms and threaten to become disease-causing *autoantigens*.

The term *immunogen* refers to "antigen" in the classic, second sense of an antigen that, when used to immunize, stimulates an immune response to itself. Likewise, an *allergen* is an antigen that stimulates an allergic response.

Facets of antigens, such as their role as ligands for receptors and as inducers of antibodies, are given special terms. For example, "hapten," "epitope," and "determinant" refer to molecular structures that physically engage the antigen receptor (see Fig. 6.1). Antigens that are not also immunogens are "incomplete antigens," whereas immunogens are also called "complete antigens."

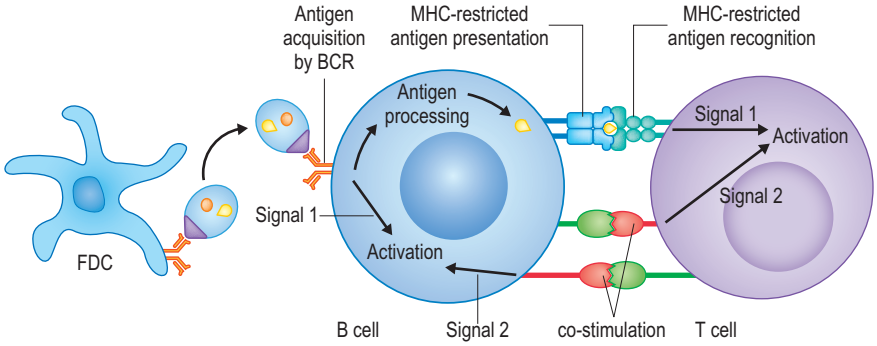


FIG. 6.2 Antigen Presentation. Follicular dendritic cells in germinal centers present antigens bound to local antibodies stored in their surface iccosomes. B cells acquire antigens through their BCR and present processed peptide epitopes via MHC molecules to T cells. T cells recognize antigens presented by MHC molecules on antigen-presenting cells. B and T cells receive signal 1 through the BCR and TCR, respectively. *BCR*, B-cell receptors; *FDC*, follicular dendritic cell; *MHC*, major histocompatibility complex.

Finally, we draw an important distinction between the terms "immunogen" and "adjuvant." Many immunogens are inactive unless mixed with an adjuvant. Adjuvants, such as alum (a form of aluminum hydroxide), have two critical functions in vaccines (Chapter 87): a depot effect and an immunostimulatory effect. As a depot, the adjuvant allows retention of the antigen in the tissue in order to provide steady stimulation. As an immunostimulant, it promotes the acquisition, processing, and presentation of antigen by APC.

Antigens for Antibodies

Antibodies are classically defined as soluble immunoglobulins (Ig). They reside in the blood and lymph fluids, and they permeate the tissues. When imbedded in the membrane of the producing B cells, immunoglobulins serve as the BCRs for antigen. They also have an essentially invariable constant region (termed Fc), which is largely responsible for its biological effects.

Antibodies can bind *via* their Fc domains to Fc receptors (FcRs) or other moieties (e.g., complement receptors) on a variety of other molecules (Chapter 8). Antibodies bind antigens through their highly variable *antigen-binding V domains* that are located at the *N*-termini of the heavy and light chains. BCRs can signal the presence of antigen to an antibody-producing B cell, and antibodies can signal the presence of antigen to cells expressing Fc or complement receptors. Immunoglobulins can also mediate *antigen acquisition* by B cells or FcR-positive APC by receptor-mediated endocytosis (Fig. 6.2).

Immunoglobulins are heterodimeric molecules composed of two heavy (H) chains and two light (L) chains. The *antigen-binding site* of immunoglobulin is formed by the juxtaposition of three highly variable intervals on the H chain and three hypervariable intervals on the L chain (see Chapter 4). The structure of this site can be in the form of a knob, a shallow groove, or a deep pocket. The latter can accommodate molecular structures as small as a single sugar molecule and as large as an oligosaccharide or oligopeptide of six or seven residues. These minimal structures on the antigen that actually bind to antibodies are termed the *epitopes*. Antigens can be much larger—for example, as large as a protein, virus, or bacterium - and can be viewed as collections of epitopes.

Epitopes can be formed by a string of contiguous residues of a protein or other polymer or by a set of non-contiguous residues that are juxtaposed in the three-dimensional structure of the parent antigen. The latter are called *conformational* epitopes because they are present in the antigen only when it is properly folded. They can be destroyed if the protein is denatured as, for example, on a Western blot. Conformational epitopes are typically found on the surface of native proteins and are often important for neutralizing antibodies, which must detect the epitope on a three-dimensional antigen surface. Linear epitopes for antibodies are usually available only when the protein is denatured, as in a Western blot, or if they are present in external loops of a protein. Transitory epitopes can be created when a protein undergoes conformational changes, such as when a protein is undergoing folding or unfolding, when an epitope is exposed by alterations in the structure, or when an epitope is displayed by the association between two different proteins.[8]

The term *hapten* comes from a Greek word meaning "hold" and is drawn from the dye industry, where the word refers to the ability of dyes to *hold fast* to fabrics despite washing. A hapten is the smallest chemical moiety of an epitope that can bind effectively to the antigen-binding site of an antibody and is usually used in relation to the "hapten-carrier" concept. Naturally occurring haptens include contact-sensitizing metals such as nickel and plant products such as urushiol, the toxin from poison ivy (Fig. 6.3).

Haptens, as exemplified by the small molecule trinitrophenol, are not immunogens for two reasons. First, by themselves, they form few electrochemical contacts with the antibody, and so their binding strength is usually very low. Second, they are so small that they cannot be subdivided to produce multiple epitopes, a feature critical to immunogenicity. Haptens are typically monovalent and so do not themselves cross-link BCRs. When chemically conjugated to proteins, they can become multivalent and, by modifying self-peptides, they can create epitopes for T cells.

KEY CONCEPT

Antigens for B cells

- Immunogens contain
  - epitopes that bind to the antigen-binding sites of antibodies
  - Class II epitopes for T helper cells
- Haptens can have almost any chemical nature.
- Epitopes on native proteins are usually amino acids discontinuous in the primary sequence but juxtaposed and found on the surface of the folded molecule.

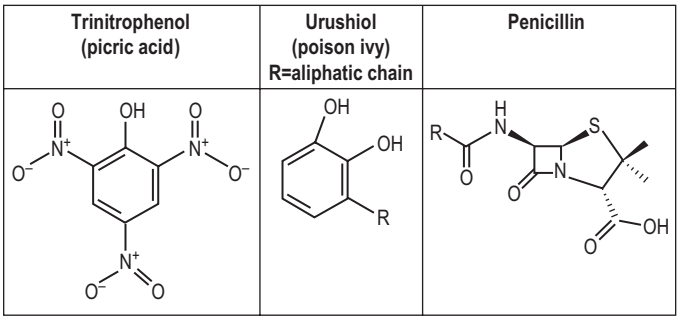


FIG. 6.3 Some Examples of Haptens

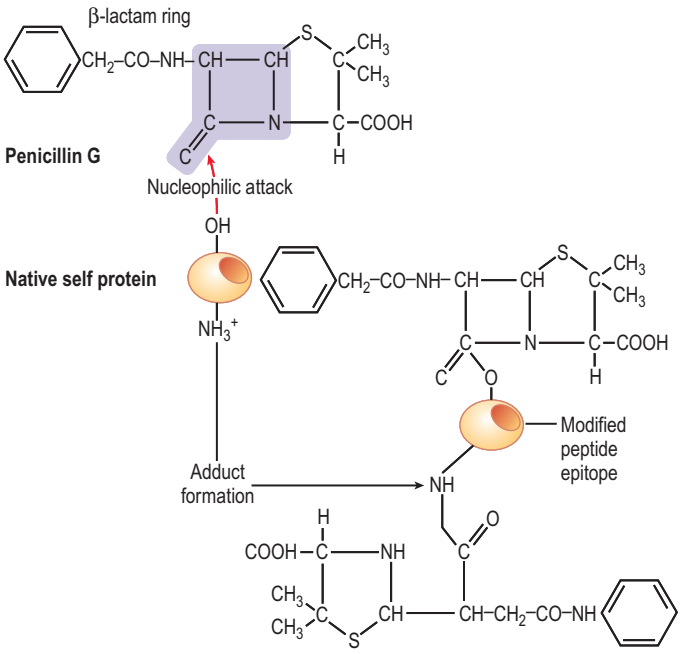


FIG. 6.4 Sensitizing Agents Such as Penicillin Create Neoantigens by Forming Covalent Adducts With Self-Proteins. Penicillin allergies involve both antibodies against penicillin and T-cell responses to penicillin-modified self-proteins. The same chemical reaction that allows penicillin to inhibit peptidoglycan formation in bacteria leads to adduct formation of cellular proteins. Nucleophilic attack by penicillin G *(upper left)* on the β-lactam ring *(shaded)* opens the ring and creates an adduct *(lower left)* with serines and lysines. The lactam adducts can be presented to B cells as modified self-proteins or processed for presentation by MHC molecules to T cells as lactam-conjugated self-peptides.

Immunogens for B cells can be placed into either T cell-dependent (TD) or T-independent (TI) categories. TI antigens come in two flavors. Type 1 TI antigens can stimulate antibody production from even neonatal B cells in the absence of MHC-restricted T-cell help. The prototype of TI antigens is lipopolysaccharide (LPS), also known as *endotoxin*, derived from the cell wall of bacteria. LPS drives polyclonal responses of mouse B cells because it is an activating ligand for the Toll-like receptor TLR4 (Chapter 3) found on B cells. LPS stimulates the proliferation of B cells with LPS-specific antigen-receptors, and it can induce class-switching by B cells and thus produce IgG or IgA. Human B cells normally do not express TLR4 and are unresponsive to LPS. However, TLR4 and LPS-responsiveness is inducible in human B cells by ligands for other TLR molecules, and patients with Crohn disease do express functional TLR4.[9]

Type 2 TI antigens stimulate antibody production from mature but not neonatal B cells. These antigens, which include the ABO blood group antigens, are typically polysaccharides or glycolipids with repeating epitopes.[10] These can cross-link multiple BCR on a single B cell and thereby activate it. B cells activated in this manner usually need help from T cells to undergo class-switching. Thus, antibodies against ABO antigens and other Type 2 TI antigens are of the IgM class. Since IgM cannot be transported across the placenta by the IgG transporter, FcRn (Chapter 8), this explains why incompatible ABO blood groups rarely present a problem for the fetus or newborn.

T-dependent antigens contain epitopes recognized by T cells as well as for B cells (see Fig. 6.1). B cells endocytose the parent antigen through the BCR, then process and present peptide epitopes to helper T cells within germinal centers.[11] The T and B cells provide co-stimulation for each other, inducing class-switching and somatic hypermutation of immunoglobulin genes in the B cells (Chapter 7).[12] Most T-dependent antibody responses quickly switch from initial IgM to IgG, IgA, or IgE. The simultaneous and cooperative responses of T cells and B cells in lymph node *germinal centers* facilitate immune responses of both cell types.

T-dependent antigens can be modeled by the hapten-carrier concept, in which B-cell responses to the hapten require help from T cells responding to the epitopes within a carrier protein (see Fig. 6.1).[13] Experimentally, carrier proteins are typically foreign proteins. However, self-antigens can also serve as carriers in clinically relevant examples. For example, both urushiol, the active ingredient of poison ivy, and penicillin (see Fig. 6.3) readily form covalent adducts with cellular proteins (Fig. 6.4). These haptenated self-proteins constitute *neoantigens*, which in this case elicit both T- and B-cell allergic responses.

Haptens are actually not the smallest moieties that can be recognized specifically by antibodies. *Determinants* are the molecular structures that actually *determine specificity*, and these can be as small as a single-side chain of an amino acid. For example, antibodies can be generated that recognize the difference between a serine and a phosphoserine as part of a larger protein. In this case, the determinant is the phosphate group. A single nitrate group is the determinant differentiating the two haptens dinitrophenol and trinitrophenol.

Using appropriate carriers and adjuvants, antibodies can be raised *in vivo* against almost any chemical moiety. There seem to be only two limitations to what can be an antigen in the sense of ligand. The antigen must be perceived as "foreign" in some sense to the responding animal, and it must provide a surface to which an antibody can form an electrochemical attachment. It is often possible to trick the immune system into generating autoreactive antibodies, and many of these arise in autoimmune reactions. Similarly, bacteriophage libraries of recombinant antibodies (*phage display libraries*)[13] can be screened for almost any chemical specificity without regard for self/non-self discrimination.

The notion that "anything goes" is powerful theoretically but has its practical limitations. On the one hand, the powerful adjuvants and immunization protocols that can be used in experimental animals cannot be used in humans, so that vaccinologists

are still frustrated by their inability to reproducibly stimulate effective immunity to many antigens. On the other, a variety of substances are effectively and fortunately hypoallergenic or hypoantigenic, permitting their use in cosmetic and implantable devices. These substances are non-proteinaceous and so lack T-cell epitopes. They are also chemically inert and thus unable to haptenate proteins. Finally, though sometimes polymeric, they seem to lack a suitable chemical surface for forming high-affinity bonds with antibodies. For example, despite efforts to find contrary evidence, polysiloxane ("silicone") is immunologically inert even in experimental models.[14] This compound is used in contact lenses, breast implants, and other medical devices. These practical comments do not exclude the theoretical possibility that some individuals might generate antibodies reactive with some hypoantigenic substances.

Carbohydrate Antigens

Polysaccharides tend to be type 2 T-independent (TI) antigens.[10] Examples include the pneumococcal capsular polysaccharides targeted by pneumococcal vaccines and the human ABO blood group antigens.[15] The latter are a "family" of related antigens expressed by most tissues and by many kinds of commensal bacteria. The polymorphic *ABO* locus encodes a glycosyltransferase that functions as a haptenating enzyme to modify the H antigen, a polysaccharide found on many different glycoproteins and glycolipids. The common *O* allele is functionally silent so that only the H antigen is generated. Both active enzymes (alleles *A* and *B*) transfer a uridylate diphosphate (UDP)-charged sugar to glycoproteins and glycolipids. The *A* and *B* enzymes use UDP-*N*-acetyl-galactosamine and UDP-galactose, respectively, as sugar donors.

Individuals of genotypes *AA*, *BB*, or *AB* express enzyme A only, B only, or both A and B antigens as self-antigens and are tolerant of them and of bacterial antigens with the same sugars. In contrast, *O*-type individuals make neither *A* nor *B* antigens. As a result, both antigens appear foreign to *O* individuals, and exposure to common bacteria induces both anti-*A* and anti-*B* antibodies. Similarly, *A*-type individuals (genotype *AO* or *AA*) make anti-B antibodies, and B–type individuals (*BO* or *BB*) make anti-A antibodies. These systems are called "blood type," referring to the *type of antibodies the person does not make*. The antigens, however, are found in all tissues, including red blood cells. Infants are exposed to the Type 2 TI A and B antigens through environmental exposure soon after birth. Anti-A and anti-B antibodies consequently begin to accumulate early in life (though use diagnostically is not valid during the first year) and appear as "natural antibodies"[16] of the IgM isotype.

Antigens as Ligands for T-Cell Receptors

Most epitopes recognized by TCR are short peptides generated from proteins through *antigen processing*. Consequently, as peptides rather than native proteins, epitopes recognized by T cells are generally linear. The biochemical definition of antigen indicates that epitopes recognized by TCR should bind directly to them. However, although direct binding of the epitope to TCR can be demonstrated in rare cases, the affinity of these interactions is typically too weak for measurement and probably too weak for biological effects. The weak affinity of a TCR for its epitope reflects, in part, thymic education.[17] which has several mechanisms to select against T cells that recognize epitopes directly (Chapter 9). Instead, epitopes for TCR first bind tightly to MHC molecules (class I for CD8 T cells and class II MHC for CD4 T cells), and it is the molecular complex of MHC plus peptide that is recognized by the TCR (Chapter 5).

Conventional oligopeptide epitopes may include modified amino acids such as phosphoserine or sugar residues. Hapten-modified self-peptides are major determinants of allergic responses to non-protein environmental agents, such as metals, cosmetics, and antibiotics. However, certain class Ib (or "non-classical") MHC molecules (Chapter 5) seem to be specialized for recognizing non-peptide antigens. For example, CD1 presents certain lipids, both endogenous and bacterial, to invariant natural killer (iNKT) cells[18]; MR1 presents unidentified bacterial ligands to mucosa-associated invariant TCR (MAIT) T cells[19]; and HLA-E presents peptide sequences to CD8 T cells and NK cells.[20] With rare exceptions, MHC molecules do not discriminate between self and foreign peptides. Indeed, the vast majority of epitopes bound by MHC molecules are self-epitopes.

The binding site for peptides in both class I and class II MHC molecules is a deep cleft that interacts with the peptide backbone and two or three of the side chains.[21] The latter are considered *anchor* residues and define the *binding motif* of the MHC molecule. The need for terminal anchoring severely limits the length of epitopes for class I MHC molecules to seven to ten amino acids, though longer peptides can sometimes bind by looping out central residues. The short length of peptide epitopes for class I molecules is enforced by a closed-ended binding cleft that binds both amino and carboxyl termini (Fig. 6.5). In class II molecules, both ends of the binding cleft are open. While, theoretically, this allows longer peptides to bind, in practice, the degradation of the ends of these peptides that usually occurs during antigen processing causes the lengths of most peptides presented by class II molecules to be limited to 15 to 20 amino acids.

Three factors generally control whether a given peptide will be recognized by T cells as antigen: foreignness, binding affinity, and antigen processing (Chapter 10). Roughly half of all patho-

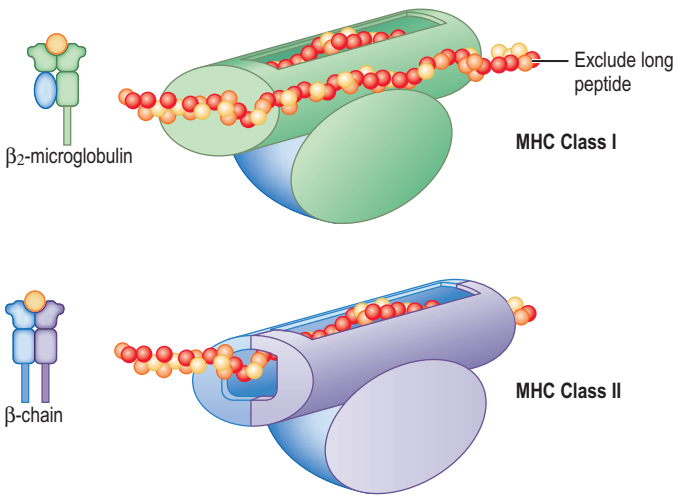


FIG. 6.5 Peptide Binding by MHC Class I and II Molecules. Class I molecules are usually closed at both ends. The peptide termini must interact with terminal sockets. Peptides that are too long must be cleaved *(arrows)* prior to entry into the binding site. The clefts of class II molecules are open at the ends, permitting the binding of longer peptides. *MHC*, Major histocompatibility complex.

gen peptides are identical to self-peptides and, except for cases of autoimmunity, are tolerated by the immune system. The binding affinity of an epitope for most class I MHC molecules can be predicted fairly well.[22] However, naturally occurring proteins must be processed proteolytically (see below), and many sequences that could be good epitopes are either not generated effectively or are destroyed proteolytically.[23] For example, the chicken egg white protein ovalbumin contains three sequences that bind very tightly to class I MHC molecules in the C57BL/6 laboratory mouse, and each of these is a potent antigen on its own. However, only one of these is produced naturally from the parent protein. Identifying the mechanisms and rules governing the proteolytic production of epitopes remains a major challenge.

The fact that any given allelic form of MHC molecule recognizes two or three anchor residues with high specificity severely limits the "universe" of peptides that it can present to T cells. Thus, a given class I allele can recognize only about 0.01% of all possible octamer peptides. Two different MHC alleles may recognize distinct anchors and thus see radically distinct antigenic universes. Nonetheless, most proteins have few potential T-cell epitopes, and some proteins will not be antigenic in certain individuals simply because they do not produce a foreign peptide that binds with sufficient strength to any of the individual's MHC molecules.

Human populations have hundreds of allelic forms of class I and class II MHC molecules. The high polymorphism of MHC molecules has three important biological implications. First, most individuals from unrelated parents are heterozygous for each MHC locus, effectively doubling the size of their antigenic universe. Second, while many individuals may be unable to respond to common or novel pathogens, at least a few individuals are likely to be protected due in part to the promiscuity of their MHC molecules.[24] This may account for the prevalence of certain MHC haplotypes in populations exposed to malaria and for genetic vulnerabilities or resistance to HIV. Finally, epitopes detected by one individual may be invisible to the T cells of another person. Thus, subunit vaccines effective for some MHC haplotypes may be ineffective for others. For example, the hepatitis B surface antigen subunit vaccine is often ineffective in individuals homozygous for certain *HLA* haplotypes.[25]

Major Histocompatibility Complex Restriction

Recognition of peptides by TCR is said to be "MHC-restricted." This has two meanings, molecular and genetic. In both, "restriction" is seen as a limitation or condition on the ability of T cells to recognize their antigens. In the newer, molecular view, the ability of T cells to recognize their peptide epitopes is "restricted" by the MHC molecules that physically present the epitopes to them. Indeed, X-ray crystallography has shown that about 90% of the contacts made between TCR and the MHC:peptide complex are with the MHC molecule.[26] However, the older, genetic meaning of "MHC-restricted" is more severe. T cells can recognize their cognate epitope *only* when it is presented by a particular allelic form of MHC molecule. This property was revealed long ago because of the extraordinary genetic diversity (polymorphism) of most class I and class II molecules in most species, including humans. Most MHC alleles are very distinct from one another, and most of the differences are concentrated in the portions of the MHC molecules that bind peptide and TCR. As a result, different allelic forms of MHC differ both in what peptide they can bind and how they will be seen by the TCR. Due to thymic selection (Chapter 9), TCRs have specificity for both the particular peptide and a particular MHC molecule and are unable to recognize other combinations of peptide and MHC. Thus, unless they happen to be MHC-matched, antigen-specific T cells from one individual are unable to recognize even the same peptide presented by APCs from other individuals.

There are a number of potential benefits for the operation of MHC restriction and for why T cells recognize antigens presented by MHC molecules (Fig. 6.6). Antigen-processing increases the complexity of pathogen antigens by exposing epitopes not available on the surface of pathogens. The exquisite specificity of MHC molecules reduces the universe of detectable foreign peptides, but it also reduces the universe of detectable self-proteins, decreasing the risk of autoimmunity. The requirement that T cells do not respond unless activated by co-stimulation from the APC is enforced by anchoring MHC molecules on the APC. Thus, MHC-restriction forces a T cell to be restricted by antigen-presenting cells, again limiting the risk of pathogenic self-reactivity.

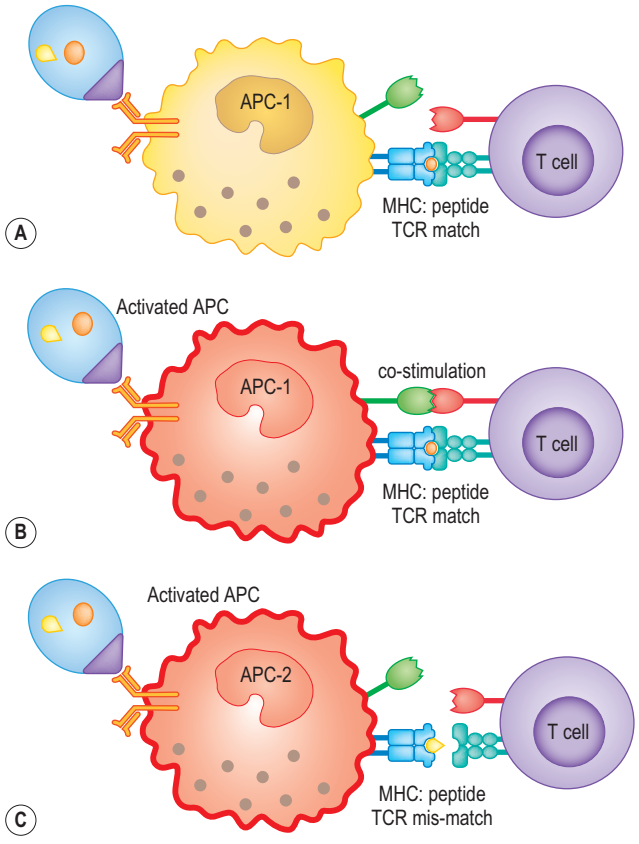


FIG. 6.6 MHC Restriction Carries Out Two Critical Functions. First, by presenting processed peptides derived from within proteins and pathogens, MHC molecules sample a broader antigenic landscape than antibodies, whose epitopes are surface oriented. Second, naive T cells respond to cognate epitopes only when presented by an activated APC (B) but not when the APC is resting (A). Experimentally observed MHC restriction results when an activated APC cannot present the proper MHC/peptide pair (C). *APC*, Antigen-presenting cells; *MHC*, major histocompatibility complex; *TRC*, T-cell receptor.

Alloantigens

Thymic selection ensures that T cells recognize non-self peptides in the context of self-MHC molecules. A clinically important exception to this specificity manifests as "alloreactivity".[27] If lymphocytes from imperfectly matched donors are mixed in a mixed-leukocyte reaction (MLR),[28] about 5% of the T cells from one donor cross-react with some of the MHC:peptide complexes of the other. Thus, a T cell specific for a peptide from cytomegalovirus (CMV) and restricted by the class I molecule HLA-A2 might recognize some other peptide presented by HLA-B7 from the other donor. Disparities in the *major histocompatibility complex antigens* (MHC molecules, also known as *major transplantation antigens*) between an organ donor and recipient are very likely to stimulate graft rejection because of the high probability of cross-reactivity. But even if the donor is perfectly matched for MHC molecules (as happens in one-fourth of full-sibling pairs), minor histocompatibility differences can also cause tissue rejection. The *minor antigens* are proteins that are polymorphic or of limited expression in a species and thus, in humans represent donor-recipient disparities in one or more of the roughly 30,000 different proteins encoded by the genome. A prominent example of a minor histocompatibility antigen is the H-Y antigen, encoded by the Y chromosome, which causes T cell-mediated rejection of male donor tissue by female recipients. The MLR is extraordinarily sensitive to the presence of alloantigens and is a good predictor of whether organ transplants will be accepted.

KEY CONCEPT

Properties of Superantigens

Defining Properties
- Presented and recognized as an unprocessed, native protein.
- Contact TCR and MHC molecules outside the traditional antigen-binding groove.

Specific Properties
- Selectively stimulate T cells expressing certain TCR Vβ chains
- TCR recognition is not MHC allele restricted.
- Stimulate both CD4 and CD8 T cells in an MHC class II-dependent manner.

Superantigens

"Superantigens" are microbial proteins that bind both class II MHC molecules and TCR, causing activation of the T cell.[29] Superantigens include certain bacterial toxins, such as staphylococcal enterotoxin A (SEA) and toxic shock syndrome toxin (TSST), and viral proteins, such as the mouse mammary tumor virus (MMTV) superantigen. Superantigens bind class II MHC molecules as intact macromolecules and bind outside of the peptide-antigen binding groove. Class II binding is independent of the specific MHC allele (though often with a preference for one class II isotype—DR, DQ, or DP). Each type of superantigen also binds TCR belonging to a characteristic subset of Vβ families and, in some cases, also makes contact with the TCR Vα chain. Nonspecific T-cell activation appears to be a bacterial strategy to avoid microbe-specific recognition. However, by stimulating large numbers of clones of a specific Vβ family, bacterial superantigens can also induce an overwhelming T-cell response with massive cytokine release leading, for example, to food poisoning and toxic shock syndrome.

ANTIGEN-PRESENTING CELLS

Cells that Present Antigens to B cells: Follicular Dendritic Cells

In many respects, B cells do not need antigen to be presented to them by any other cell: they express high-affinity antigen-receptors that have no contextual requirement for antigen-binding. However, engaging B cells with soluble, especially monovalent, antigens can tie up the BCR in a non-signaling and even tolerogenic mode. Efficient B-cell activation requires either a polyvalent antigen, as with TI antigens, or some form of additional mechanism that effectively cross-links the BCR and induces signaling. Such cross-linking can be mediated by immune complexes (assemblies of antigens and antibodies), antigens found on the surface of pathogens (thus rendered polyvalent), or antigens fixed by complement. Two complement receptors on B cells (CR1=CD35; CR2=CD21) bind fragments of C3 and C4 attached to antigens (Chapter 4).

Most B cells are activated in the limiting architecture of the germinal center, and clonal activation leads to nearly monoclonal responses in each center. In this context, a unique cell, the *follicular dendritic cell* (FDC; Chapter 2),[30] plays a special role in presenting antigens to B cells. FDCs represent less than 1% of the cells within a germinal center and are tightly associated with B cells and associated macrophages.

FDCs are stromal, not hematopoietic, in origin and are required for the formation of germinal centers. The presence and activity of FDCs in germinal centers require the presence of lymphotoxin, a member of the tumor necrosis family. FDC's play an organizing role by secreting the chemokine CXCL13, which binds to CCR5 on lymphocytes (Chapter 15), inducing the secretion of lymphotoxin.[31] FDCs trap and accumulate antigens through FcR and complement receptors (especially CR1, CR2, and C4bR) and store them in *iccosomes* (immune complex-coated bodies).[32] Antigen in iccosomes is stable for months.

The ability of FDCs to accumulate antibodies secreted by local B cells gives them a high-affinity trapping mechanism with a specificity cognate with those B cells. FDCs are thought to be critical for isotype switching, affinity maturation, and B-cell memory. Because C4b is a surrogate activator of CD40, FDC-presented antigen can drive these processes in the absence of CD40L; likely explaining the efficacy of many T-independent antigens. FDCs are non-phagocytic, do not express MHC class II or T-cell activation molecules such as B7/CD80/CD86, and do not present antigens to T cells. However, by providing antigen to B cells or germinal center dendritic cells, FDCs can activate germinal center T cells indirectly.

FDCs aid in the formation of tertiary lymphoid tissues in inflamed tissues and contribute to unwanted immune reactions, such as in chronic organ rejection and autoimmunity associated with vascular inflammation such as SLE. Fetal and maternal immune cells come into direct contact at the decidua, and the fetus can be considered an allograft to the mother. Uterine dendritic cells (DC) within the decidua have been implicated in the maintenance of pregnancy by inducing tolerance.[33] These observations suggest that manipulating FDCs may be therapeutically useful in controlling autoimmunity.

Cells that Present Antigens to T Cells

Cells that present antigens to and activate *naïve* T cells are called *professional Antigen-Presenting Cells* or, more commonly,

just APC.[34] Unlike class I MHC molecules, which are expressed on almost all nucleated cells in the body, class II molecules are expressed almost exclusively by professional APC. Among human cells, MHC class II and costimulatory molecules can also be expressed by activated T cells and inflamed endothelial cells. In addition, MHC class II is expressed by medullary and cortical epithelial cells in the thymus, where they play an important role during positive and negative selection.

Naïve T cells have a high threshold for activation and require costimulatory ligands found on professional APC. These include B cells, macrophages, mast cells and basophils, and both myeloid and plasmacytoid dendritic cells (DCs). Activated T cells have vastly reduced requirements for co-stimulation, and their responses are proportional to the level of MHC expression. For typical class I molecules, this means almost all nucleated cells are potential targets for CD8 T cells. However, because most cells do not express costimulatory molecules, they do not stimulate naïve T cells.

The different types of professional APCs have distinct but overlapping properties. Conventional (myeloid) DC (cDC) are by far the most potent in terms of their ability to present a wide variety of antigens to naïve T cells.[35] cDC in their immature form are constitutively tissue-resident phagocytes, actively engulfing and digesting any antigen in their vicinity. In this stage, they express few costimulatory or class II MHC molecules. Once activated by a "danger" signal, cDC mature rapidly. They stop phagocytosis and begin to express recently digested antigens through upregulated MHC molecules. They become mobile and follow a chemokine trail from the tissue to nearby draining lymphatics, through which they reach lymph nodes where they begin to present their antigens to T cells. Upregulated costimulatory molecules and cytokines allow the DC to activate naïve T cells and direct their differentiation.

The functions of cDC can be modulated. In the absence of full activation, they tend to be tolerogenic. Once activated by certain innate receptor ligands, cDC express IL-12 and drive a Th1 response (Chapter 11). However, if IL-10 is provided by some other cell type, cDC secrete little IL-12 and drive a predominantly IL-2 response. Other cytokine environments are able to skew cDC functions so that they drive Th-17 or Treg pathways.

Although cDC are extremely potent on a per-cell basis, the activity of other professional APC types should not be underestimated because they vastly outnumber the cDC. Macrophages are active phagocytes and can activate naïve T cells. B cells are not phagocytic but can internalize cognate antigens through their BCR. As a result, they are several orders of magnitude more sensitive than cDC to limiting antigen concentrations.[36] Plasmacytoid DC secrete very high levels of type I interferon but are not potent antigen-presenters. In contrast, mast cells are not phagocytic but may be specialized for presenting antigens acquired through FcR.

Antigen Acquisition

We have hinted that APCs acquire antigens through multiple mechanisms. From a topological perspective, antigens are either exogenous or endogenous. Exogenous antigens (i.e., antigens synthesized externally to the APCs) are acquired by the cells and its associated antigen processing apparatus chiefly through endocytosis. Endogenous antigens (i.e., antigens synthesized within the APC) are already "acquired" by the cell but are often in the wrong cellular compartment. Viral capsid proteins synthesized by an infected cell and inserted into the cell membrane can be internalized by endocytosis for antigen presentation. Autophagy is an important mechanism for digesting internal structures, including invasive bacteria as well as mitochondria and other organelles. It is part of the cellular physiology of most cell types. In APC, autophagy also mediates the presentation of internal antigens. However, most peptides presented by class I MHC molecules are "acquired" through a non-phagocytic mechanism involving ubiquitin, specialized proteases, and both peptide and protein transporters located in the membrane of the endoplasmic reticulum.

Pinocytosis, essentially very small-scale reversible endocytosis of the cell membrane, samples the fluid phase outside the cell.[37] This takes place in many cells and is a property of "ruffled" membrane edges found at the leading edge of mobile cells.

Receptor-mediated endocytosis through clathrin-coated pits internalizes many receptors and their cargo ligands. This is a property of most cell types and is a major mechanism for acquiring antigens for APCs. Typical FcR for IgG, such as found on mast cells, macrophages, and dendritic cells, internalize upon cross-linking (Chapter 8). These cells readily present antigens found in immune complexes. In contrast, B cells express an alternatively spliced form of this receptor that binds antibodies at the surface but does not internalize. Thus, as a group, B cells are not effective at presenting generic antigens acquired through FcR. However, the BCR itself readily internalizes after being cross-linked, and B cells are exquisitely adept at presenting their cognate antigen.

Phagocytosis is a mechanism for internalizing particles that may be as large as the cell itself. Initial engagement of multiple receptors causes a local deformation of the cell surface and a partial invagination of the cell membrane. This deformation, coupled with the triggering of specific receptors, leads to subcellular enzyme activity that changes the lipid composition of the membrane bilayer and remodels the cytoskeleton in the vicinity of the particle. This relaxes the membrane, allowing deeper invagination and creating the *phagocytic cup*. If the particle is small enough, the cup deepens until, when the particle is nearly engulfed, the outer edge of the cup closes like a purse-string, leading to membrane fusion and creating a new external surface and an intracellular vesicle. The endocytic vesicle undergoes successive fusion with other vesicles until fusion with the lysosome. Phagocytosis is a property of phagocytes (DC, macrophages, and neutrophils) and in these cells is accompanied by further inflammatory activation of the phagocyte.

In addition to professional phagocytes, other cell types can also phagocytose apoptotic cells, using a specialized set of receptors that detect apoptotic cells. Many cells can mediate this process, including those that do not possess large lysosomes. Phagocytosis of apoptotic cells by APC is often anti-inflammatory. In contrast, phagocytosis of microbes or necrotic cells, including cells undergoing secondary necrosis after apoptosis, is inflammatory.

Autophagy, found in even the simplest eukaryotes, is a major mechanism for mediating normal protein turnover, protection from intracellular pathogens, and resistance to starvation. It allows a cell to recycle its own organelles. Perhaps 50% of the peptides presented by class II MHC molecules are endogenous peptides acquired through autophagy. Autophagy also mediates presentation by class I MHC molecules of exogenous and some endogenous peptides.[38]

Three broad categories of autophagic mechanisms have been described. *Macroautophagy* mediates the engulfment of microbes or organelles through autophagosomes. *Microautophagy* closely resembles the vesicle fusion described for phagocytosis. *Chaperone-induced autophagy* allows vesicles, including lysosomes, to import proteins directly from the cytoplasm. These three forms of autophagy are linked by their reliance on a set of *ATG* (Autophagy) gene products. ATG8 and ATG12, both ubiquitin-like proteins, nucleate a phagophore attached to an invading microbe or targeted organelle, followed by rapid recruitment of a lipid membrane to seal the object into an autophagosome. Macroautophagy can be induced by bacterial products, such as LPS, and by drugs, including rapamycin. Autophagy can control the activity of specific cellular proteins, complicating the interpretation of many experiments, including NFκB, so that the role of ATG proteins may be very indirect.

Cross-presentation refers to the idea that an APC can present antigens produced by some other cell. Since "cross-presentation by class II MHC molecules" is the norm, the term is used most often to describe cross-presentation by MHC class I, where it is critical for activation of CD8 T cells by APC. Cross-presentation can also inhibit immune responses. For example, antigen-specific B cells present epitopes from their cognate antigens and thereby become targets for CTL. In this way, CTL can suppress B-cell responses to some antigens.

KEY CONCEPT

Antigen Processing and Presentation for Class I MHC

- Peptides presented on most cell types are synthesized endogenously.
- Peptides acquired through endocytosis and autophagy can be cross-presented
- Most epitopes are processed by proteasomes and enter the ER through the TAP transporter.
- Peptides of 8–10 amino acids are bound at both termini within the binding cleft.

ANTIGEN PROCESSING

T cells recognize their cognate antigens in the form of short peptides embedded in MHC molecules (Chapter 5). X-ray crystallography shows that approximately 90% of the molecular surface recognized by the TCR is the MHC molecule. *Antigen processing* excises these peptides from their parent protein antigens and loads them onto MHC molecules.[39] The processing can take place in different cellular compartments but loading takes place chiefly in specialized loading compartments. Class I MHC molecules load chiefly in the endoplasmic reticulum, though some loading might take place in endosomal compartments during *cross-presentation*. In contrast, specific mechanisms prevent class II loading in the ER but facilitate it in the specialized endosomal loading compartments.

Endogenous peptides presented by class I molecules derive from proteins made on the cell's own ribosomes. The chief mechanism for degrading large proteins into small peptides is the *proteasome*,[40] a macromolecular tubular structure containing multiple protease activities that cleave proteins into small fragments of 8 to 14 residues. There are at least four mechanisms that feed proteins into proteasomes. First, nascent cytoplasmic polypeptides that fail to fold properly are attacked by enzymes that attach chains of the protein *ubiquitin* to the protein targeted for destruction. Second, nascent proteins that are translocated into the endoplasmic reticulum endosome, but do not fold properly, can be exported back to the cytosol by a protein transporter called Sec61. This mechanism is called ERAD (endoplasmic reticulum-associated protein degradation). These two mechanisms are the most important and together represent the DRiP (defective ribosome products) model. The "DRiP" pathway is not simply an error-dependent pathway but rather represents a specialized mechanism of sampling open reading frames.[41] Third, properly folded mature proteins are ubiquitinated in the course of normal activity—part of the normal course of protein turnover. Finally, in autophagy and cross-presentation, proteins engulfed by phagocytes are partially degraded in lysosomes before transfer into the cytosol.

The DRiP mechanism provides a conceptual core for understanding how class I MHC molecules can sample internally generated antigens, a key requirement for targeting virus-infected or malignant cells. Degradation of mature proteins has a half-time of hours to months, but virus replication can take place in the order of hours. If MHC molecules were to sample antigens as they degrade through normal channels, surface representation would lag considerably behind internal processes. By sampling proteins during their synthesis, or if they fail to fold properly immediately after synthesis, DRiP mechanisms ensure that cytotoxic T cells receive a timely report of internal states.

All four class I MHC pathways make heavy use of the "UPS" (ubiquitin/proteasome) system for protein degradation found in all cells. UPS is initiated by heat-shock chaperone proteins recognizing a malfolded protein and inducing covalent tagging of the protein with a single copy of ubiquitin, a 76 amino acid protein. This triggers polyubiquitination in which succeeding ubiquitins are attached to the preceding ubiquitin. Polyubiquitin tails are recognized by the regulatory subunit of the proteasome, a multi-subunit cylindrical machine. Substrates are fed through the central channel and digested by proteolytic subunits, producing a residue of peptides roughly 8 to 14 amino acids long.

These products are substrates for the *transporter associated with antigen processing* (TAP), a member of the ATP-binding cassette (ABC) transporter family.[42] Heterodimers of TAP1 and TAP2 subunits form peptide pumps that burn ATP to drive peptides from the cytosol into the lumen of the ER. Without a ready source of peptides in the ER, class I MHC molecules are extruded into the cytosol by Sec61 to be recognized by the UPS. TAP mutations are involved in many cases of type I bare lymphocyte syndrome (BLS) (Chapter 34),[43] and many tumor cells lack class I expression due to mutations in their *TAP* genes (Chapter 5). There are several allelic forms in humans with modest differences in specificity. Cells that are not expressing TAP express low levels of certain MHC alleles and can be highly resistant to specific CTL. However, mice and humans lacking TAP are not profoundly immunodeficient and produce TAP-independent class I MHC-restricted CTL (Chapter 34). These findings suggest alternative mechanisms for loading class I MHC molecules are important, even if not dominant under normal circumstances.

Interferon-γ induces an alternative set of proteolytic and regulatory subunits for the proteasome to create the "immunoproteasome." This has altered specificity and activity and may favor the production of peptides suitable for transport by TAP or alternative pathways.

A minor subset of peptides enters the ER independently of TAP through the protein secretory pathway. Nascent polypeptides bearing a *signal peptide* are recognized and transported into the lumen by the *signal recognition particle* (SRP). The signal peptide itself is cleaved by a *signal peptidase*. By inserting the sequence of a CTL epitope behind a signal peptide, this pathway can be used to deliver the epitope directly into the MHC class I loading compartment. Peptides that are too long to bind MHC class I molecules may be retained temporarily in the ER by additional peptide-binding proteins such as BiP, pumped back into the cytosol by a non-TAP mechanism, or trimmed by cytosolic aminopeptidases

The luminal proteases appear quite efficient, apparently reducing the steady-state concentration of free antigenic peptides in the ER to very low levels. As a result of differences in protease specificities and/or kinetic effects, different epitopes can be carved out of the same protein, depending on whether it follows the proteasome/TAP or the secretory pathway into the ER. Finally, different peptides within a single protein can be degraded or protected at different rates, leading to the immunodominance of a subset of potential epitopes for a given MHC molecule.

The production of an immunodominant epitope from influenza A nucleoprotein (NP) illustrates how extra-epitope residues might affect non-proteasome, non-ubiquitin processing.[44] The optimal NP peptide in one mouse strain is the nonamer TYQRTRALV. The three *C*-terminal residues are efficiently removed from a related 12-mer peptide TYQRTRALVRTG. However, an 11-mer, TYQRTRALVTG, is impotent at producing the epitope. The terminal TG sequence represents a "block" to epitope production. All of these complexities in antigen processing complicate the prediction of which potential epitopes, identified on the basis of their ability to bind to class I molecules, will be immunodominant *in vivo*.

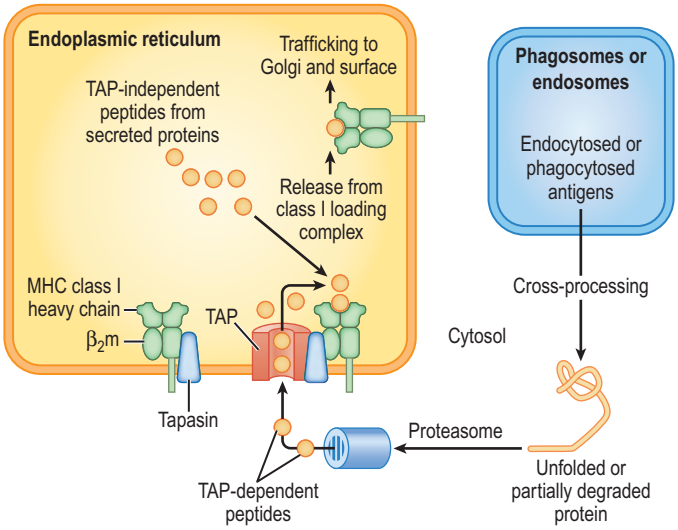


FIG. 6.7 Antigen Processing for MHC Class I. Two chief pathways for antigen processing intersect within the cytosol. Most endogenous antigens are synthesized on ribosomes and then processed by proteasomes into peptides, which enter the ER through the TAP transporter. A minor set of antigens are processed within the ER from proteins secreted into the ER. Professional antigen-presenting cells transfer endocytosed antigens into the cytosol for processing. Autophagy (not shown) can also process endogenous antigens. CD1, MR1, and probably some conventional class I MHC molecules can also acquire ligands in the endocytic compartment itself. *MHC,* Major histocompatibility complex; *TAP,* transporter associated with antigen processing.

Class I Major Histocompatibility Complex Trafficking

Nascent Class I MHC molecules are inserted into the ER membrane via the protein secretory pathway (Fig. 6.7). Nascent chains bind first to the membrane-bound chaperone calnexin until they begin to fold into association with β_2-microglobulin light chains (β_2m). Heterodimers of β_2m and heavy chain are released by calnexin and bind the soluble chaperone, calreticulin. This assembly engages a *class I loading complex*, which includes a 60 kDa thiol reductase, the TAP heterodimer, and another MHC-encoded protein, tapasin.[45] Tapasin retains class I molecules in the complex until they bind peptide. MHC molecules failing to attract peptides misfold and are exported by Sec61 to the cytosol. Only those that bind peptide are released by tapasin from the loading complex, migrate to the Golgi where they undergo glycan maturation, and then traffic to the cell surface for recognition by CD8 T cells.

KEY CONCEPT

Antigen Processing for Class II MHC

- Class II MHC is expressed constitutively only by professional APC (dendritic cells, macrophages, and B cells)
- Epitopes presented by professional APC are acquired mostly through endocytosis and autophagocytosis.
- Peptides (10–15 amino acids) often have terminal extensions, extending outside the antigen-binding groove.

Antigen Processing for Class II-Restricted T Cells

MHC class II molecules assemble in the ER where they associate with *invariant* chain, a 31 kDa protein that chaperones nascent αβ dimers of class II molecules into endosomes. A segment of the invariant chain, called CLIP, blocks class II molecules from binding peptides in the ER. CLIP may be removed along with the rest of the invariant chain once in the endosomes. Alternatively, a CLIP fragment may be left occupying the binding cleft; many class II MHC molecules traffic to the cell surface with CLIP embedded.

MHC class II molecules are loaded with peptides digested by lysosomes. Antigens acquired through endocytosis or autophagy are unfolded and partially degraded in endosomal and acidic lysosomal subcompartments by disulfide isomerase (which unlinks disulfide loops) and a variety of proteases. Most peptides presented by class II molecules are processed from parent proteins and loaded onto class II molecules within a specialized loading compartment of the endosomes. Within this general scheme, there are at least two pathways for epitope production, distinguishable in part by their dependence on the function of DM molecules (Fig. 6.8).[39] DM molecules (heterodimers of DMA and DMB subunits, which are homologous to MHC class II proteins) catalyze the exchange of CLIP for processed epitopes. In the DM-independent pathway, peptides are loaded onto class II molecules recycling from the cell surface in the absence of CLIP.

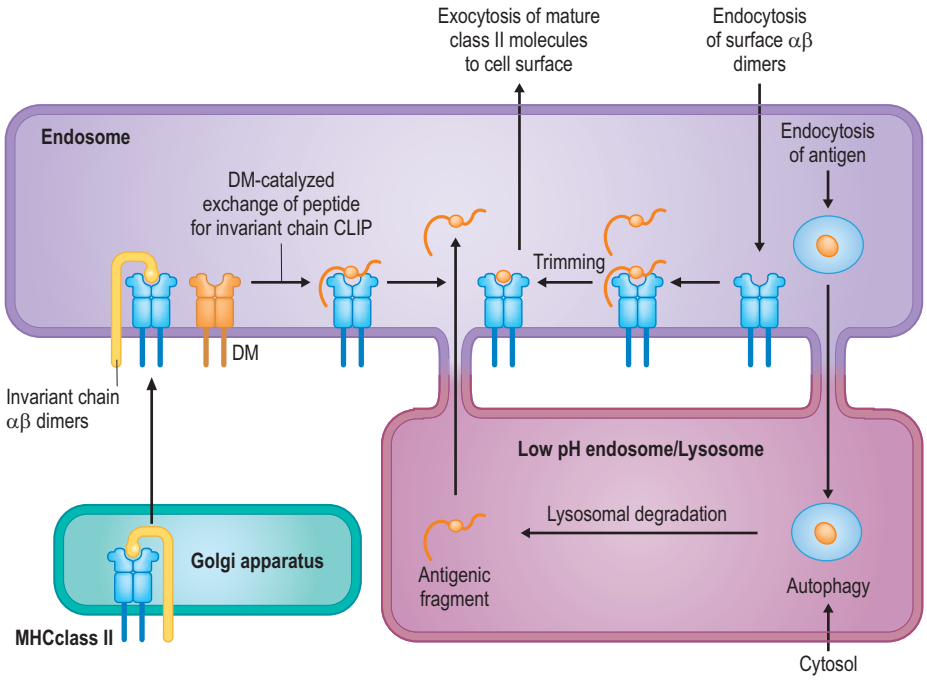


FIG. 6.8 Two Pathways for Loading Antigens Onto Class II MHC Molecules. Autophagy and endocytosis transfer endogenous and external antigens, respectively, into the endosomes. Nascent class II molecules are chaperoned to the endosomes from the Golgi by the invariant chain. The DM molecules catalyze the exchange of antigenic peptides for the invariant chain. Mature class II molecules recycling from the cell surface can acquire peptides in a DM-independent manner. Antigens, initially binding as polypeptides, are trimmed into oligopeptides in the endosomes and at the surface.

The initial ligand for binding class II molecules is a large, unfolded protein or protein fragment rather than the oligopeptide ultimately displayed at the cell surface. This MHC-polypeptide complex is the substrate for trimming exopeptidases. Endosomal aminopeptidases cannot cleave at proline residues; thus, prolines are found near the *N*-terminus of many mature epitopes. Trimming of the extra-cleft residues can continue even after the MHC–peptide complex has reached the surface through the activity of the membrane-bound surface enzyme aminopeptidase N.

As in the case of epitope formation for class I molecules, the initial conformation of the antigen can affect the production of specific linear epitopes. For example, vaccinating mice with synthetic peptides elicits class-II-restricted T cells specific for at least two HIV gp160 epitopes, of which only one is detected on infected cells. Similarly, prior denaturation or mutational destabilization of viral influenza hemagglutinin abolishes its ability to be processed for presentation to some Th clones.

Predicting Epitopes for T-Cell Receptor

The critical role of antigens for T cells in driving both T- and B-cell responses to antigen has fueled attempts to use antigen sequence information to predict T-cell epitopes.[22] The characterization of *binding motifs* for a large number of class I and a smaller number of class II molecules has facilitated computer-based algorithms for predicting potential epitopes from a linear protein sequence. These motifs reflect the chemical affinity of the binding cleft for various amino acid side chains. In many cases, it is possible to show that multiple alleles of class I or class II molecules recognize the same or very closely related epitopes. These groups of MHC alleles are called "supertypes" and can facilitate the prediction of epitopes for a large number of alleles. However, because of the complexity of endoprotease and exoprotease cleavage during the processing of epitopes for both class I and class II MHC molecules, it is difficult to predict whether a given epitope will actually be used *in vivo*.

ANTIGEN PRESENTATION

Once loaded with peptide, MHC molecules move to the cell surface, where they can be recognized by T cells. This is antigen presentation at the bare minimum and is sufficient for triggering effecter responses from pre-activated T cells. Presentation to naïve T cells requires additional factors and efficient activation of the TCR requires adhesion molecules. These additional processes are usually included under the rubric of "antigen processing." Antigen recognition/presentation of both naïve and activated T cells is mediated by the *immune synapse*,[46] although this structure might not be required in all cases.

The immune synapse represents a very close and tight association of APC and T cell resembling tight junctions and may even mediate trogocytosis, through which process peptide-loaded MHC molecules are transferred to the T cell itself (Chapter 4).[47] The synapse is initiated when the leading edge of a T cell engaged in amoeboid motion through a tissue meets a potential APC or target cell. Initial low avidity interactions between the T-cell integrins CD11a and VLA-four mediate a weak approximation of the two cellular surfaces. If MHC molecules on the APC present cognate antigen, the TCR will bind with maximum affinity. The co-receptors CD8 and CD4 bind to class I and class II MHC molecules, respectively (Chapter 5). Their cytoplasmic tails, already loaded with the protein kinase LCK, are swept into proximity with the cytoplasmic domains of the TCR chains, leading to phosphorylation of the latter and

initiating downstream signaling events. An immediate effect is the 1000-fold upregulation of the integrin avidities. The co-stimulatory receptor CD28 migrates to the field, engaging its ligands on the APC. At the surface, the synapse matures as TCR and CD28, and their ligands concentrate at the center of the synapse, surrounded by a ring of integrins. Internally, the ongoing signaling events cause an arrest of cell migration, with a re-organization of the microtubules to permit the trafficking of vacuoles to the synapse. Vacuolar contents are delivered into the intercellular space of the synapse, where, if the T cell is a CTL, they will induce apoptosis in the opposite cell.

Microbial Interference with Antigen Processing and Presentation

Considering the importance of class I-mediated immune responses in antiviral immunity, it is not surprising to find pathogens using a variety of mechanisms for subverting antigen processing. Proteins from several serotypes of human adenovirus, as well as HIV-1 tat protein, inhibit class I MHC transcription. Other viral factors inhibit class I MHC maturation in the endoplasmic reticulum. Proteins US2 and US11 from the human cytomegalovirus (CMV) target nascent class I molecules for destruction through ERAD. The CMV US6 and herpes simplex virus ICP47 proteins inhibit TAP function, indirectly starving MHC molecules of peptide. In contrast, the CMV US3 protein binds class I molecules that have already engaged peptides but retains them in the ER. The HIV-1 protein nef binds the intracellular tails of mature class I MHC molecules, targeting them for increased endocytosis and degradation. Other pathways are also affected by pathogens. For example, the protein ICP345 from the herpes simplex virus inhibits ATG6, a critical autophagy-initiating protein. Also, *Mycobacterium tuberculosis* suppresses the acidification of lysosomes in macrophages to create for itself an environment conducive to its own replication inside lysosomes.

CLINICAL RELEVANCE

Clinical Correlates of Antigen Processing

- The tautological and complementary interaction of antigens and their antigen receptors constitute the recognitive mechanism for acquired immunity to pathogens and tumors and autoimmunity towards the "self."
- Human genetic defects in these antigen processing pathways result in three types of bare lymphocyte syndrome (BLS).
- Structural features of certain antigens may predispose them to induction of allergic (IgE) responses by affecting antigen processing.
- Molecular mimicry and exposure of cryptic self-epitopes are mechanisms leading to autoimmunity.
- Tumor-specific neoantigens can be detected by the immune system, often leading to selection for tumor variants lacking the relevant MHC and/or antigen-processing functions.
- Tumor antigens are typically self-proteins that can be:
  - Products of tumor viruses
  - Over-expressed normal proteins
  - Mutated products of oncogenes
  - Clonally-expressed idiotypes of antigen receptors

CLINICAL RELEVANCE

If microbes can manipulate specific mechanisms of antigen processing and presentation, it should be obvious that human genetic variation in these pathways should also be able to contribute to immune competency or, through deficiency, to immune dysfunction. This may result in specific or global defects in antigen processing and recognition, such as the bare lymphocyte syndrome (BLS) (Chapter 34).[43] Type I BLS involves general loss of class I surface expression, typically the result of mutations in the TAP peptide transporter. Type II BLS reflects the loss of class II MHC expression. Defects in any of at least four different transcription factors can cause this disease. In type III BLS, defects in the RFX transcription factor depress the expression of both class II MHC molecules and the β_2-microglobulin light chain shared by all class I MHC proteins.

Allergens appear to be unusual antigens. It has been long known that allergens inducing delayed-type hypersensitivity are often drugs or environmental compounds able to form covalent adducts with self-proteins, thus generating neoantigens. Many allergens inducing immediate hypersensitivity are or are associated with proteases.[48] These proteases are thought to drive a Th2 response by T cells.

Self-antigens recognized by autoimmunity are so far chemically unremarkable. The principles that govern pathological self-recognition appear to be the same as for healthy self-tolerance or recognition of foreign antigens: the availability of particular processed peptides at appropriate times for tolerance induction of T-cell activation.

Two nonexclusive models have emerged that may begin to account for why certain individuals are predisposed to autoimmunity and why certain self-antigens are likely to become autoantigens (Chapter 51). The first is the concept of *molecular mimicry*.[49] According to this idea, exposure to sufficient doses of a pathogen-derived epitope that cross-reacts with a previously ignored or *cryptic* self-epitope can break self-tolerance to that epitope. A newer concept in autoantigenicity can explain why many autoantigens are proteins normally found intracellularly, where they are involved in nucleic acid and protein metabolism: small nuclear riboproteins, histones, and heat-shock proteins. This involves the observation (discussed earlier with regard to cross-presentation) that apoptotic bodies are efficiently recognized by dendritic cells, and that many intranuclear and intracellular antigens are exposed on the extraverted surfaces of apoptotic bodies. Thus, by inducing apoptosis, it is possible that certain predisposing infections can elicit autoimmune reactions to cryptic self-epitopes.

Finally, tumor-specific and tumor-associated antigens are typically self-proteins.[50] In rare human cases, such as a peptide derived from papillomavirus type 16, they can be encoded by tumor viruses. Some tumor antigens, such as carcinoembryonic antigen (CEA), prostate-specific antigen (PSA), or the MAGE proteins of melanomas, are normal proteins that are merely overexpressed by tumor cells. These can serve as diagnostic markers or as a target for tumor-selective immunity. The antigenic products of mutated tumor suppressor genes and other oncogenes, such as *HER2*, the retinoblastoma protein RB, and the breast cancer–associated antigen BRAC, are also called *neoantigens* and are potentially more specific targets for immunotherapy. The neoantigens expressed by these tumors arise as chance mutations during the many steps of carcinogenesis. The immune system itself provides a unique category of neoantigens that can be targets of immunotherapy, the clonally distributed products of rearranged antigen receptor genes – idiotypes – expressed by malignancies such as myelomas.

Translational Research in Antigen Processing and Presentation

ON THE HORIZON

- Improved definition of the cellular and molecular mechanisms controlling dendritic cell presentation of antigen to T cells and NK cells by classical and non-classical MHC class I molecules as relevant to antiviral and anti-tumor immunity.
- Understanding the role of autophagy to improve vaccine design.
- Clarification of the role of MHC class Ib molecules in the presentation of non-peptide antigens and in the regulation of immune responses.
- Solution to the puzzle of clonally expressed antigen receptors on NK cells.

Several areas are ripe for new insights with probable application to clinical immunology. These include a deeper understanding of peptide processing for presentation by class I MHC molecules, identification of antigenic ligands recognized by CD1- and MR1-restricted T cells, and a better understanding of the mechanisms that control antigen receptor function in NK cells. Although it is now widely accepted that dendritic cells present endocytosed antigens for presentation by MHC class I, the mechanisms and rules governing this process are not clear.

Likewise, autophagy mediates the processing of endogenous antigens in both MHC class I and class II pathways and is likely important in cross-presentation. Pathogens can both be cleared by autophagy but also can wrest control of autophagy for their own purposes, probably including redirection of antigen processing. We expect clarification of the role of autophagy in antigen processing to yield new tools for controlling infection and vaccine design.

The CD1 and MR1 class Ib MHC molecules present non-peptide ligands to specialized invariant NKT cells, and these appear to have important regulatory functions in suppressing autoimmunity, especially in the gut. Ligands for MR1 and for CD1 are poorly understood. This is an area of growth in the next several years and will likely prove important for understanding and treating autoimmunity and some infectious diseases. Finally, NK cells in mice and humans express novel species of antigen receptors in a clonal fashion. These receptors and their mechanism of clonal expression or clonal licensing have opened a new chapter in acquired immunity.

REFERENCES

1. Silverstein AM. *A history of immunology*. 2nd ed. New York: Academic Press; 2009. 07/17/2009. 552 p.
2. Medzhitov R, Janeway Jr. CA. Decoding the patterns of self and nonself by the innate immune system. *Science*. 2002;296(5566):298–300.
3. Matzinger P. The danger model: a renewed sense of self. *Science*. 2002;296(5566):301–305.
4. Bagheri M, Zahmatkesh A. Evolution and species-specific conservation of toll-like receptors in terrestrial vertebrates. *Inter reviews of immunol*. 2018;37(5):217–228.
5. Ibrahim ZA, Armour CL, Phipps S, Sukkar MB. RAGE and TLRs: relatives, friends or neighbours? *Mol Immunol*. 2013;56(4):739–744.
6. Ru H, Zhang P, Wu H. Structural gymnastics of RAG-mediated DNA cleavage in V(D)J recombination. *Curr Opin Struct Biol*. 2018;53:178–186.
7. Bashirova AA, Martin MP, McVicar DW, Carrington M. The killer immunoglobulin-like receptor gene cluster: tuning the genome for defense. *Ann Rev Genomic Hum Genet*. 2006;7:277–300.
8. Lewis GK, Fouts TR, Ibrahim S, Taylor BM, Salkar R, Guan Y, et al. Identification and characterization of an immunogenic hybrid epitope formed by both HIV gp120 and human CD4 proteins. *J Virol*. 2011;85(24):13097–13104.
9. Ganley-Leal LM, Liang Y, Jagannathan-Bogdan M, Farraye FA, Nikolajczyk BS. Differential regulation of TLR4 expression in human B cells and monocytes. *Mol Immunol*. 2010;48(1–3):82–88.
10. Mond JJ, Lees A, Snapper CM. T cell-independent antigens type 2. *Annu Rev Immunol*. 1995;13:655–692.
11. Luckheeram RV, Zhou R, Verma AD, Xia B. CD4(+)T cells: differentiation and functions. *Clin & develop immunol*. 2012;2012:925135.
12. Methot SP, Di Noia JM. Molecular Mechanisms of Somatic Hypermutation and Class Switch Recombination. *Adv Immunol*. 2017;133:37–87.
13. Britschgi M, von Greyerz S, Burkhart C, Pichler WJ. Molecular aspects of drug recognition by specific T cells. *Curr Drug Targets*. 2003;4(1):1–11.
14. Tugwell P, Wells G, Peterson J, Welch V, Page J, Davison C, et al. Do silicone breast implants cause rheumatologic disorders? A systematic review for a court-appointed national science panel. *Arthritis Rheum*. 2001;44(11):2477–2484.
15. Branch DR. Anti-A and anti-B: what are they and where do they come from? *Transfusion*. 2015;55 Suppl 2:S74–S79.
16. Panda S, Ding JL. Natural antibodies bridge innate and adaptive immunity. *J Immunol*. 2015;194(1):13–20.
17. James KD, Jenkinson WE, Anderson G. T-cell egress from the thymus: Should I stay or should I go? *J leuko biol*. 2018;104(2):275–284.
18. Vartabedian VF, Savage PB, Teyton L. The processing and presentation of lipids and glycolipids to the immune system. *Immunol Rev*. 2016;272(1):109–119.
19. Karamooz E, Harriff MJ, Lewinsohn DM. MR1-dependent antigen presentation. *Semin Cell Dev Biol*. 2018;84:58–64.
20. Joosten SA, Sullivan LC, Ottenhoff TH. Characteristics of HLA-E Restricted T-Cell Responses and Their Role in Infectious Diseases. *J Immunol Res*. 2016;2016:2695396.
21. Stevanovic S. Structural basis of immunogenicity. *Transpl Immunol*. 2002;10(2–3):133–136.
22. Mei S, Li F, Leier A, Marquez-Lago TT, Giam K, Croft NP, et al. A comprehensive review and performance evaluation of bioinformatics tools for HLA class I peptide-binding prediction. *Brief Bioinform*. 2019
23. Unanue ER, Turk V, Neefjes J. Variations in MHC Class II Antigen Processing and Presentation in Health and Disease. *Annu Rev Immunol*. 2016;34:265–297.
24. Manczinger M, Boross G, Kemeny L, Muller V, Lenz TL, Papp B, et al. Pathogen diversity drives the evolution of generalist MHC-II alleles in human populations. *PLoS Biol*. 2019;17(1):e3000131.
25. Singh R, Kaul R, Kaul A, Khan K. A comparative review of HLA associations with hepatitis B and C viral infections across global populations. *World j gastro: WJG*. 2007;13(12):1770–1787.
26. Housset D, Malissen B. What do TCR-pMHC crystal structures teach us about MHC restriction and alloreactivity? *Trends Immunol*. 2003;24(8):429–437.
27. Montgomery RA, Tatapudi VS, Leffell MS, Zachary AA. HLA in transplantation. *Nat Rev Nephrol*. 2018;14(9):558–570.
28. DeWolf S, Shen Y, Sykes M. A New Window into the Human Alloresponse. *Transplant*. 2016;100(8):1639–1649.
29. Krakauer T. Staphylococcal Superantigens: Pyrogenic Toxins Induce Toxic Shock. *Toxins (Basel)*. 2019;11(3).
30. Deshane J, Chaplin DD. Follicular dendritic cell makes environmental sense. *Immunity*. 2010;33(1):2–4.
31. Ruddle NH. Lymphotoxin and TNF: how it all began-a tribute to the travelers. *Cytokine Growth Factor Rev*. 2014;25(2):83–89.
32. Tew JG, Wu J, Qin D, Helm S, Burton GF, Szakal AK. Follicular dendritic cells and presentation of antigen and costimulatory signals to B cells. *Immunol Rev*. 1997;156:39–52.
33. Blois SM, Kammerer U, Alba Soto C, Tometten MC, Shaikly V, Barrientos G, et al. Dendritic cells: key to fetal tolerance? *Biol Reprod*. 2007;77(4):590–598.
34. Sprent J. Antigen-presenting cells. Professionals and amateurs. *Curr Biol*. 1995;5(10):1095–1097.
35. Collin M, Bigley V. Human dendritic cell subsets: an update. *Immunol*. 2018;154(1):3–20.

36. Adler LN, Jiang W, Bhamidipati K, Millican M, Macaubas C, Hung SC, et al. The Other Function: Class II-Restricted Antigen Presentation by B Cells. *Front Immunol*. 2017;8:319.
37. Palm W, Thompson CB. Nutrient acquisition strategies of mammalian cells. *Nature*. 2017;546(7657):234–242.
38. Virgin HW, Levine B. Autophagy genes in immunity. *Nat Immunol*. 2009;10(5):461–470.
39. Kelly A, Trowsdale J. Genetics of antigen processing and presentation. *Immunogenet*. 2019;71(3):161–170.
40. Budenholzer L, Cheng CL, Li Y, Hochstrasser M. Proteasome Structure and Assembly. *J Mol Biol*. 2017;429(22):3500–3524.
41. Dolan BP, Bennink JR, Yewdell JW. Translating DRiPs: progress in understanding viral and cellular sources of MHC class I peptide ligands. *Cellular and molecular life sciences: CMLS*. 2011;68(9):1481–1489.
42. Liu X. ABC Family Transporters. *Advances in experimental medicine and biology*. 2019;1141:13–100.
43. Shrestha D, Szollosi J, Jenei A. Bare lymphocyte syndrome: an opportunity to discover our immune system. *Immunol lett*. 2012;141(2): 147–157.
44. Yellen-Shaw AJ, Eisenlohr LC. Regulation of class I-restricted epitope processing by local or distal flanking sequence. *J Immunol*. 1997;158(4):1727–1733.
45. Praest P, Liaci AM, Forster F, Wiertz E. New insights into the structure of the MHC class I peptide-loading complex and mechanisms of TAP inhibition by viral immune evasion proteins. *Mol Immunol*. 2019;113:103–114.
46. Dustin ML, Choudhuri K. Signaling and Polarized Communication Across the T Cell Immunological Synapse. *Annu Rev Cell Dev Biol*. 2016;32:303–325.
47. Dopfer EP, Minguet S, Schamel WW. A new vampire saga: the molecular mechanism of T cell trogocytosis. *Immun*. 2011;35(2):151–153.
48. Porter PC, Ongeri V, Luong A, Kheradmand F, Corry DB. Seeking common pathophysiology in asthma, atopy and sinusitis. *Trends Immunol*. 2011;32(2):43–49.
49. Segal Y, Shoenfeld Y. Vaccine-induced autoimmunity: the role of molecular mimicry and immune crossreaction. *Cellular & molecular immunol*. 2018;15(6):586–594.
50. Freudenmann LK, Marcu A, Stevanovic S. Mapping the tumour human leukocyte antigen (HLA) ligandome by mass spectrometry. *Immunol*. 2018;154(3):331–345.

B-Cell Development and Differentiation

Harry W. Schroeder Jr., Andreas Radbruch, and Claudia Berek

B lymphocytes (B cells) are the sole source of antibodies. Each developing B cell creates its own unique immunoglobulin (Ig) that can be expressed on the cell surface as B-cell receptor (BCR) for antigen or secreted as an antibody (Chapter 4). During their development, B cells encounter different microenvironments (Chapter 2) that expose them to a variety of cytokines (Chapter 14), adhesion molecules (Chapter 16), and chemokines (Chapter 15).[1] These processes impose major bioenergetic demands on the cells.[2] Ultimately, however, the fate of the B cell is dependent on the interaction between Ig and antigen (Chapter 6).

EARLY B-CELL DEVELOPMENT BEGINS IN THE PRIMARY LYMPHOID ORGANS

B cells arise from multipotent hematopoietic stem cells that successively populate the yok sac, the endothelium of the major arteries, the fetal liver, and then the bone marrow.[3] Stem cell daughter cells give rise to lymphoid primed multipotent progenitors (LMPPs), which in turn can give rise to either myeloid or lymphoid cells. LMPPs then produce common lymphoid precursors (CLPs), which can generate T cells, B cells, natural killer (NK) cells, or dendritic cells. Final B-cell differentiation requires the exposure of CLP daughter cells to specialized microenvironments in the fetal liver and the bone marrow. These two tissues are the primary B-lymphoid organs. The shift from fetal liver to bone marrow begins in the middle of fetal life and ends prior to birth. The bone marrow continues to produce B cells until the end of life, although the rate decreases with age.

An intact and functional B-cell antigen receptor complex, which consists of membrane-bound immunoglobulin (mIg), the Igα and Igβ co-receptors, and ancillary signal transduction components, must be present in order for the developing B cell to survive (Chapter 7). The composition of the BCR is subject to intense selection. In the primary organs, hazardous self-reactive BCRs, as well as nonfunctional ones, can be culled by changing the L chain (receptor editing), by cell anergy,[4] or by apoptosis of the host cell. Survivors of this initial selection process are released into the blood and thence to the spleen, lymph nodes, and other secondary lymphoid tissues and organs where selection for specificity continues (Chapter 2).

B-cell differentiation (Fig. 7.1) is commonly presented as a linear process defined by the regulated expression of specific sets of transcription factors, Ig, and cell-surface molecules. Given the central role of the BCR (Chapter 4), initial developmental steps are classically defined by the status of the rearranging Ig loci. With the development of monoclonal antibody technology, analysis of cell-surface markers such as CD10, CD19, CD20, CD21, CD24, CD34, and CD38 (Fig. 7.2) has facilitated definition of both early and late stages of development, especially in those cases where Ig cannot be used to distinguish between cell types.[5] Of these, CD19, a signal transduction molecule expressed throughout B-cell development up to, but not including, the mature plasma cell stage[6] warrants special mention as the single best clinical marker for B-cell identity.

In practice, B-cell development is a more complex process than the simple, linear pathways depicted in Figs. 7.1 and 7.2. For example, proB cells typically derive from a common lymphoid progenitor, but they can also develop from a bipotent B/macrophage precursor. Thus, B-lineage subsets identified by one fractionation scheme may consist of mixtures of subsets identified by others. Consequently, when comparing patient findings to the literature, definition of the fractionation scheme used by the reference laboratory is essential.

Initial commitment to the B-cell lineage requires activation of a series of transcriptional and signal transduction pathways. At the nuclear level, the transcription factors PU.1, Ikaros, ID-1, E2A, EBF1, and PAX5 play major roles in committing progenitor cells to the B-cell lineage.[3] After lineage commitment has been established, however, it is generally accepted that the composition of the BCR controls further development.

Each B-cell progenitor has the potential to produce a large number of offspring. Some will develop into mature B cells,

KEY CONCEPTS

B-Cell Development in the Primary Lymphoid Organs

- Commitment to the B-cell lineage reflects differential activation of transcription factors that progressively lock the cell into the B-cell pathway
- B-cell development is typically viewed as a linear, stepwise process that is focused on the assembly and testing of immunoglobulin function, first in the fetal liver and bone marrow, and then in the periphery:
  - Failure to assemble a functional receptor leads to cell death.
  - Expression of a functional receptor subjects the B cell to antigen selection/
  - B cells with inappropriate specificities tend to be eliminated.
  - B cells responding appropriately to external antigen can develop either into immunoglobulin-secreting plasma cells or into memory cells,
- At the clinical level, B-cell development can be monitored by examining the pattern of expression of lymphoid-specific surface proteins.

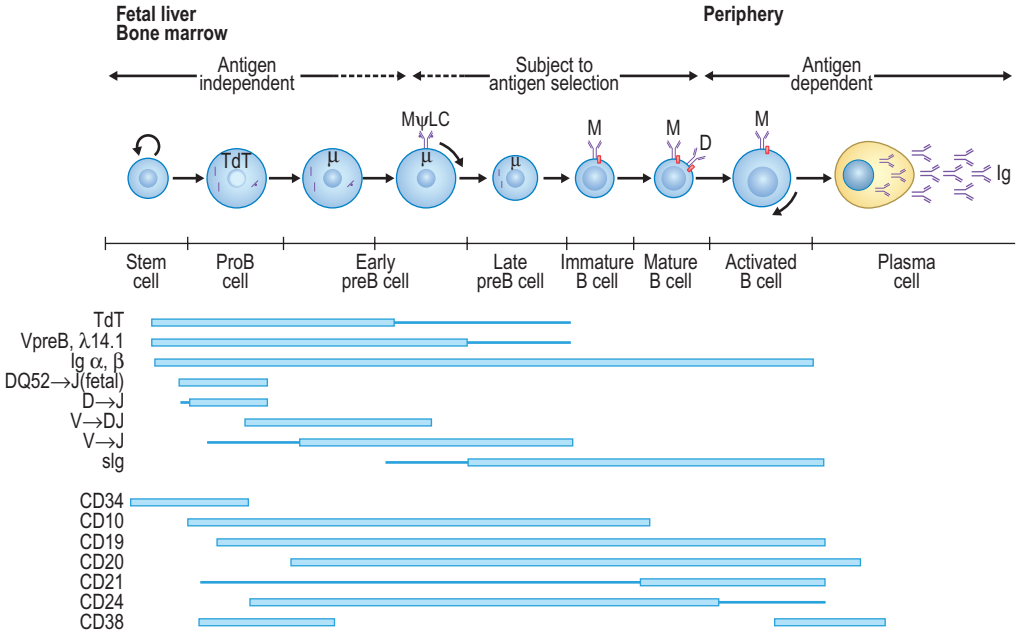
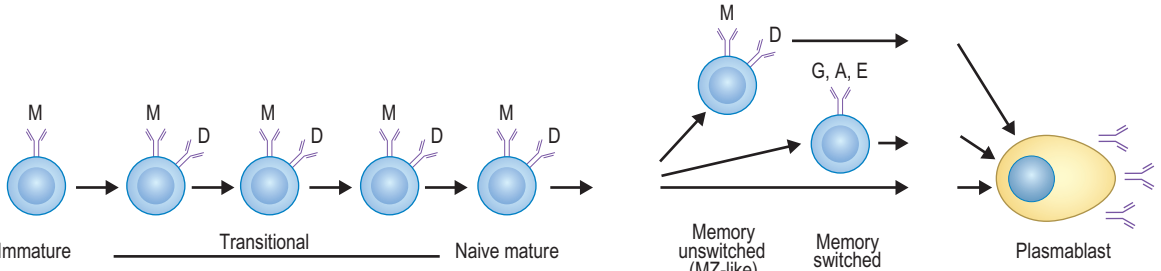


FIG. 7.1 Model of B-Cell Differentiation. B-cell development is typically viewed as a linear progression through different stages of differentiation. The various processes associated with the assembly of the B-cell antigen receptor complex and the expression pattern of surface molecules whose presence or absence are illustrated through use of bars. The various steps in immunoglobulin rearrangement and the pattern of expression of these surface molecules can be used to characterize stages in B-cell development.



	Immature	Transitional T1	Transitional T2	Transitional T3	Naive mature	Memory unswitched (MZ-like)	Memory switched	Plasmablast
IgM	++	+++	++	+	+	+	-	-
IgD	-	+	++	++	++	+	-	-
R123	+	+	+	Int	-	+	+	+
CD10	+	+	+/-	-/+	-	-	-	+/-
CD19	+	+	+	+	+	+	+	lo
CD24	+++	+++	++	+	+	+	-	-
CD27	-	-	-	-	-	+	+	hi
CD38	+++	+++	++	+	+	lo	lo	hi

FIG. 7.2 B-Cell Subsets in the Peripheral Blood. Peripheral blood B-cell subsets can be identified by differential staining for IgM, IgD, CD10, CD19, CD24, CD27, CD38, and use of the rhodamine dye R123, which is extruded by the ABCB1 transporter expressed in naïve, mature B cells and, to a lesser extent, T3 transitional B cells.[36] B cells that are $CD20^{+}CD27^{+}CD43^{+}CD70^{-}$ are currently the best candidates for the human B-1-cell counterparts,[5] but this is subject to change. In addition to these subsets found in the blood, germinal center B cells found in the spleen and lymph nodes are characterized as $IgD^{-}CD38^{++}$ $CD10^{+}$ $CD27^{+/-}$; marginal zone B cells are typically $CD27^{+}$ $CD21^{++}$ $CD23^{+/-}$ $CD1c^{+}$ IgD^{-}; and long-lived plasma cells found in the bone marrow, spleen, and tonsils are $CD138^{+}$ $CD20^{-}$ $CD38^{++}$.

and even fewer into long-lived memory B cells or plasma cells. Others, indeed the majority, will perish.[7] Most of the defined steps in this process of development represent population bottlenecks, developmental checkpoints wherein the developing B cell is tested to make sure that its BCR will be beneficial.[8] In the periphery, exposure to antigen is associated with class switching and hypermutation of the variable domains of the antigen receptor. A few select survivors earn long lives as part of a cadre of memory B cells. These veterans enable rapid engagement of antigens to which they have been previously exposed, providing experienced protection against repeat assaults.

Specialized microenvironments also play a role in peripheral B-cell development (Chapter 2), each of which enables the B cell to properly engage different types of antigens or venues of attack (see below). In the marginal zone (MZ), mature splenic MZ B cells await bacterial pathogens. In the lymphoid follicles, B cells reactive with a given antigen collaborate with follicular T cells and dendritic cells in order to maximize the immune response (Chapter 6). In the germinal centers (GCs) B cells use class switching and somatic mutation to modify and optimize the function and affinity of their Igs. And, underneath mucosal surfaces, B cells are primed to express IgA (Chapter 24).

Generation of a Functioning Antigen Receptor Is Key to the Viability of a B Cell

Ig rearrangement is hierarchical. In proB cells, heavy (H) chain $D_H \rightarrow J_H$ joining precedes $V_H \rightarrow DJ_H$ rearrangement (Chapter 4), followed by light (L) $V_L \rightarrow J_L$ joining in late-stage preB cells.

Production of a properly functioning BCR is essential for development beyond the preB cell stage. For example, function-loss mutations in RAG1/2 and DNA-dependent protein kinase (DNA-PKcs, Ku 70/80) preclude both B-cell and T-cell development (Chapter 34). Each proB cell faces the probability that only one of three possible splices will place the V_H and J_H in the same reading frame. The opportunity to try rearrangement on the second chromosome gives failing proB cells a second opportunity. Together, this provides the cell with five chances out of nine for initial survival ($1/3 + 1/3 \times 2/3$). In-frame, functional VDJ_H rearrangement allows the proB cell to produce μ H chains, most of which are retained in the endoplasmic reticulum. The appearance of cytoplasmic μ H chains marks initiation of the preB cell stage. These early preB cells tend to be large in size.

VpreB and λ14.1 [λ5], which together form the surrogate light chain (ψLC), and Igα and Igβ (Chapter 4) are constitutively expressed by proB cells. The first H chain quality control checkpoint tests for the ability of the μ H chain to associate with surrogate light chain to form a preB cell receptor. In addition to checking to see if the scaffolding (Frameworks) of the L chain can associate correctly with the scaffolding of the H chain, VpreB encodes a sensing site that can test the H chain antigen-binding site. Thus, the surrogate light chain functions as the first, and invariant, antigen to screen for antigen-binding characteristics.[9]

Successful formation of a stable preBCR facilitates termination of further H chain rearrangement *(allelic exclusion)*, which is followed by four to six cycles of cell division, a process associated with a progressive decrease in cell size. Late preB daughter cells reactivate RAG1 and RAG2 and begin to undergo $V_L \rightarrow J_L$ rearrangement. Successful production of a complete κ or λ light chain permits membrane-bound expression of IgM on the cell surface, which identifies the *immature* B cell. Immature B cells expressing self-reactive IgM antibodies may undergo repeated rounds of light chain rearrangement to lessen the self-specificity of the antibody, a process termed *receptor editing*.

Immature B cells that have successfully produced an acceptable IgM BCR extend transcription of the H chain locus to include the Cδ exons downstream of Cμ. Alternative splicing permits co-production of IgM and IgD. These now newly *mature* IgM^+IgD^+ B cells enter the blood and migrate to the periphery where they form the majority of the B-cell pool in the spleen and the other secondary lymphoid organs. The IgM and IgD on each of these cells share the same variable domains.

Tyrosine Kinases Play Key Roles in B-Cell Development

Signaling through the BCR is required for continued development. FLT3 (FLK2) is a receptor tyrosine kinase belonging to the same family as c-FMS, the receptor for colony-stimulating factor-1 (CSF-1). FLT3 ligand, which has homology to CSF-1, is a potent co-stimulator of early proB cells. In mice, targeted disruption of *flt3* leads to a selective deficiency of primitive B-cell progenitors.

Bruton tyrosine kinase is an important component of the phospholipase Cγ (PCLγ) pathway, which is used in BCR signaling. Deficiency of BTK function results in the arrest of human B-cell development at the preB cell stage and is the genetic basis of X-linked agammaglobulinemia (XLA) (Chapter 33). Inhibitors of BTK can be used to treat B-cell malignancies.[10]

BLNK is a SRC homology 2 (SH2) domain-containing signal transduction adaptor. When phosphorylated by SYK, BLNK serves as a scaffold for the assembly of cell-activation targets that include GRB2, VAV, NCK, and PLCγ. Loss of function mutations in BLNK can result in agammaglobulinemia due to the loss of preB and mature B cells.

Cell-Surface Antigens Associated with B-Cell Development

B-cell development is associated with the expression of a cascade of surface proteins, each of which plays a key role in the fate of the cell (see Fig. 7.1, Table 7.1). The timing of the appearance of each of these proteins can be used to further analyze the process of B-cell development. Several of these surface proteins, as well as their ligands and intracellular components, can be targeted for therapeutic interventions. Examples can be found in Table 7.2.

CD34 is a highly glycosylated type I transmembrane glycoprotein that binds to CD62L (L-selectin) and CD62E (E-selectin) and thus likely aids in cell trafficking (Chapter 16). It is expressed on a small population (1% to 4%) of bone marrow cells that includes hematopoietic stem cells. In mice deficient in CD34, the colony-forming activity of hematopoietic progenitors appears reduced. However, the hematopoietic profile of adult blood appears unaffected.

CD10, also known as neprilysin, neutroendopeptidase, and the common acute lymphocytic leukemia antigen (CALLA), is a type II membrane glycoprotein metalloprotease. CD10 has a short N-terminal cytoplasmic tail, a signal peptide transmembrane domain, and an extracellular C-terminal domain that includes six N-linked glycosylation sites. The extracellular

TABLE 7.1 Cell-Surface Proteins Active in Early B-Cell Development

Gene	Class or Alternative Name	Associated or Targeted Genes or Molecules	B-Cell Developmental Phenotype in Human or *Mouse* Associated with Disrupted Function of the Indicated Gene
B-Cell Receptor Complex			
μ chain	Immunoglobulin superfamily	κ, λ L chains, ψL chain, CD79 a,b(Igα, β)	AGM1: agammaglobulinemia and no B cells *Arrest at preB cell stage*
Immunoglobulin lambda-like polypeptide 1; IGLL1 (λ14.1, λ*5*)	Immunoglobulin superfamily	VpreB, μ H chain	AGM2: agammaglobulinemia and reduced B-cell numbers. *Arrest at preB cell stage*
VPREB1 (VpreB)	Immunoglobulin superfamily	λ14.1, μ H chain	*Arrest at preB cell stage*
CD79a,b (Igα, β)	Immunoglobulin superfamily, cytoplasmic ITAM motifs	H chain, LYN, FYN, BLK, SYK	AGM3 (CD79A); AGM6 (CD79B): Agammaglobulinemia and arrest at proB cell stage *Arrest at proB cell stage*
Other Cell-Surface Proteins			
CD10	Type II metalloproteinase	Hydrolyzes peptide hormones, cytokines	*Not expressed in murine B-cell progenitors*
CD19	Immunoglobulin superfamily	mIgM, PI-3 kinase, VAV, LYN?, FYN?	CVID3: panhyperglobulinemia, normal numbers of $CD20^+$ B cells in the blood
CD20	Four transmembrane domain surface molecule	B-cell Ca^{2+} channel subunit; indirectly interacts with LYN, FYN, LCK	CVID5: low IgG, normal IgM, variable IgA *20%–30% reduction in B-cell numbers*
CD21	Complement control protein	iC3b, C3dg, C3d, CD19, CD81, Leu 13, CD23	CVID7: Low IgG, reduced IgA, low normal IgM *Diminished T-cell-dependent immune responses, decreased germinal center formation, reductions in affinity maturation*
CD24	GPI-linked sialoglycoprotein	Ligand for P-selectin (CD62P)	A57V polymorphism associated with increased risk of multiple sclerosis *Deletion in mice leads to reductions in late preB and immature B-cell populations*
CD34	Type I transmembrane glycoprotein	Ligand for L-selectin (CD62L) and E-selectin (CD62E)	*Not expressed in murine B-cell progenitors*
CD38	Type II transmembrane glycoprotein ADP-ribosyl cyclase, cyclic ADP-ribose hydroxylase	ADP ribosylates proteins	*Diminished T-cell-dependent immune responses, augmented responses to T-cell-independent type 2 polysaccharide antigens*

TABLE 7.2 Selected Anti-B-Cell Biologicals and Consequences

Modality and Target	Consequences
Monoclonal antibodies directed against:	
BAFF	Preferentially blocks B cells overexpressing BAFF, including autoantibody-producing B cells in autoimmune diseases
CD20	Destruction of preB cells, B cells including follicular and memory B cells, and some short-lived plasma cells. Left intact are progenitor B cells and long-lived plasma cells
CD40	Blocks isotype switching, GC formation, memory B-cell generation, and class-switched antibody production
FcRn	Permits cellular IgG degradation and blocks IgG recycling, thereby lowering serum IgG levels in general, including autoantibodies
IL-6 receptor	Blocks IL-6 activation of intracellular kinases, reducing inflammation
Inhibitors targeting:	
Proteasomes	Causes cellular accumulation of misfolded or unfolded damaged protein, which is more common in plasma cells, leading to cell death

BAFF, B-cell activating factor of the tumor necrosis family; *GC*, general center; *IL*, interleukin.

domain contains 12 cysteines whose disulfide bonds help stabilize its zinc-binding pentapeptide motif, which is involved in its zinc-dependent metalloprotease catalytic activity. It can inactivate multiple physiologically active peptides such as endothelin-1, bombesin, bradykinin, or oxytocin. Inhibition of CD10 activity on bone marrow stromal cells enhances B-cell maturation. CD10 (CALLA) is used as a marker for preB acute lymphocytic leukemias and for certain lymphomas.

CD19 is a cell-surface glycoprotein of the Ig superfamily that is exclusively expressed throughout B-cell development from the proB cell stage up to the plasma cell stage (see Fig. 7.1). CD19 exists in a complex with CD21 (complement receptor 2 [CR2]: complementarity determining region [CDR]2), CD81 (TAPA-1) and Leu 13 (Chapter 4). With the help of CD21, CD19 can bind the complement C3 cleavage product C3d. The simultaneous binding of sIgM and CD19 to a C3d-antigen complex enables CD19 and the BCR to interact and thereby provides a link between innate and adaptive immune responses (Chapter 3). CD19–BCR interactions permit the cell to reduce the number of antigen receptors that need to be stimulated in order to activate the cell. Co-activation also reduces the threshold required for B-cell proliferation in response to a given antigen.

CD19 is generally considered as a positive B-cell signaling regulator.[11] Co-ligation of CD19 and surface Ig promotes calcium mobilization, enhances MAP kinase and Src PTK

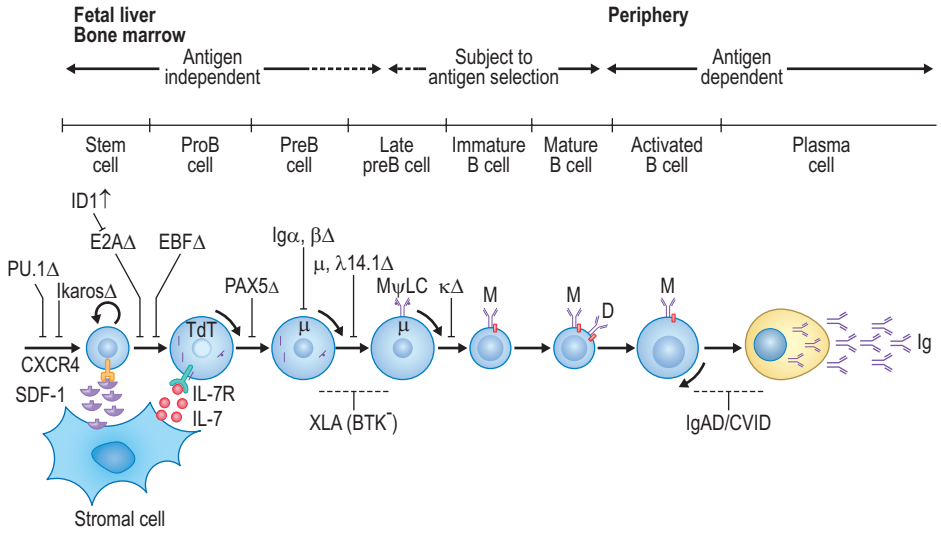


FIG. 7.3 Genes Involved in Early B-Cell Development. The stage of development at which abnormal function of a selected set of transcription factors, cytokines, chemokines, and signal transduction elements can influence B-cell development is illustrated. A *Greek delta* (Δ) or a *dash (–)* indicates a loss of function of the gene in question. An *upward arrow* (↑) indicates an increase in the function of the gene in question. *IgAD/CVID,* IgA Deficiency/Common Variable Immune Deficiency; *XLA,* X-linked agammaglobulinemia.

activation, and prolongs BCR signaling in lipid rafts. Patients deficient in CD19 have normal numbers of $CD20^+$ B cells in the blood but are panhypogammaglobulinemia and are susceptible to sinopulmonary infections (Chapter 33).

CD20 contains four transmembrane domains and cytoplasmic C- and N-termini. It is a member of the CD20/FcεRIβ superfamily of leukocyte surface antigens. Differential phosphorylation yields three forms of CD20 (33, 35, and 37 kDa). Activated B cells have increased fractions of the 35- and 37-kDa forms of the antigen. CD20 appears to function as a B-cell Ca^{2+} channel subunit and regulates cell cycle progression. It can interact directly with Major Histocompatibility Complex (MHC) class I and II molecules (Chapter 5), as well as members of another family of four transmembrane domain proteins known as the TM4SF (e.g., CD43, CD81, and CD82). It also appears to interact indirectly with LYN, FYN, and LCK. A patient, born of consanguineous parents, with CD20 deficiency has presented with normal numbers of B cells, reduced circulating memory B cells, low IgG with normal IgM and IgA, and reduced somatic hypermutation (SHM). Rituximab, a common monoclonal biologic approved for medical use in 1997, is directed against CD20. It is commonly used to treat certain autoimmune diseases and lymphoid cancers (Chapter 84).

CD21 (CR2) is a cell-surface protein that contains a small cytoplasmic domain and an extracellular domain consisting of a series of short consensus repeats termed complement control protein (CCP) domains. These extracellular domains can bind three different products of complement C3 cleavage, iC3b, C3dg, and C3d. When binding these products, CD21 acts as the ligand-binding subunit for the CD19/CD21/CD81 complex, tying the innate immune system to the adaptive immune response. Mice that lack CD21 exhibit diminished T-dependent B-cell responses. However, serum IgM and IgG are in the normal range. A patient lacking CD21 presented with low IgG, low normal IgM, normal IgA, normal responses to protein vaccination, but impaired responses to polysaccharide vaccines.

CD24 is a GPI-linked sialoprotein that serves as a ligand for P-selectin (CD62P). It is expressed on progenitor, immature, and mature B cells. Its expression decreases in activated B cells and is lost entirely in plasma cells. Monoclonal antibodies against CD24 inhibit human B-cell differentiation into plasma cells. In mice, CD24 is also known as HSA, or heat-stable antigen. Mice made deficient for CD24 show a leaky block in B-cell development with a reduction in late preB and immature B-cell populations. However, peripheral B-cell numbers are normal and no impairment of immune function has been demonstrated.

CD38 is a bifunctional enzyme that can synthesize cyclic ADP-ribose (cADPR) from nicotinamide adenine dinucleotide (NAD^+) and also hydrolyze cADPR to ADP-ribose. It is presumed that the enzyme exists to ADP-ribosylate target molecules. CD38 is expressed on preB cells, activated B cells, and early plasma cells, but not on immature or mature B cells or on mature plasma cells. Antibodies to CD38 can inhibit B lymphopoiesis, induce B-cell proliferation, and protect B cells from apoptosis. CD38 knockout mice exhibit marked deficiencies in antibody responses to T-cell-dependent protein antigens and augmented antibody responses to T-cell-independent type 2 polysaccharide antigens.

Transcription Factors and Epigenetic Mechanisms Controlling B-Cell Differentiation

Ultimately, B-cell development is a function of differential gene expression, including epigenetic regulation of specific B-cell-related genes and processes. Deficiencies in the function of the transcription factors that regulate lymphoid-specific gene expression can thus be expected to result in abnormal B-cell development (Fig. 7.3, Table 7.3).

PU.1 belongs to the erythrocyte transformation specific (ETS) family of loop-helix-loop (winged helix) transcription factors, which bind purine-rich DNA sequences. PU.1 regulates CD79a

TABLE 7.3 Nuclear and Cytoplasmic Factors Active in Early B-Cell Development

Gene	Class or Alternative Name	Associated or Targeted Genes or Molecules	B-Cell Developmental Phenotype in Human or *Mouse* Associated with Disrupted Function of the Indicated Gene
Transcription Factors			
PU.1	Loop-helix-loop (winged helix)	CD79a (Igα), μ H chain	*Arrest prior to the proB cell stage*
Ikaros	Zinc finger	RAG1, TdT, IL2R, VpreB, LCK	CVID13: progressive loss of B cells and serum immunoglobulins *Arrest prior to the proB cell stage*
Aiolos	Zinc finger	RAG1, TdT, IL2R	*Aging mice develop symptoms of systemic lupus erythematosus*
E2A	Basic Helix-loop-helix (BHLH)	RAG1, IgH, Igκ, TdT, EBF, PAX5	AGM8: Agammaglobulinemia, reduced numbers of $CD19^+$ B cells that lacked BCRs. *Arrest prior to the proB cell stage*
EBF	EBF/Olf helix-loop-helix (HLH)-like	CD79a (Igα), λ14.1, VpreB, PAX5	*Arrest prior to the proB cell stage*
ID-1	Helix-loop-helix	CD79a (Igα), λ14.1, VpreB, PAX5	*Arrest prior to the proB cell stage*
PAX5	Paired-domain	CD19, λ14.1, VpreB, BLK kinase, J chain, V_H promoters, Vκ promoters	Susceptibility to B-cell acute lymphoblastic leukemia 3. *Arrest at proB cell stage*
The Recombinase Complex			
RAG1, RAG2	Recombinase	Recombination signal sequences of immunoglobulin gene segments	Autosomal recessive SCID *Arrest at proB cell stage*
TdT	Nontemplated DNA polymerase	Coding ends of rearranging immunoglobulin gene segments	*Absence of N nucleotides, diminished production of pathogenic anti-DNA autoantibodies, loss of heterosubtypic immunity against influenza virus*
DNA-PK	DNA repair complex	Multimeric complex consisting of DNA-PKcs, Ku70, Ku80, which repairs double-stranded DNA breaks	SCID *Arrest at proB cell stage, original mouse SCID mutation identified as a loss-of-function mutation in DNA-PKcs*
Protein Tyrosine Kinases			
FLK2/FLT3	Class III receptor tyrosine kinase	GRB2, SHC	Activating mutations contribute to acute myeloid leukemia *Selective deficiency of primitive B-cell progenitors*
BLNK	SH2 adaptor protein	SYK, GRB2, VAV, NCK, phospholipase Cγ (PLCγ)	AGM4: Normal numbers of proB cells, absent preB, and B cells *Arrest at proB cell stage*
BTK	BTK/TEC protein tyrosine kinase	Phospholipase Cγ (PLCγ), SAB	XLA: X-linked agammaglobulinemia–arrest at preB cell stage *Xid: Impaired responses to T-independent antigens.*

(Igα), J chain, μ chain, κ chain, λ chain, RAG1, and terminal deoxynucleotidyl transferase (TdT), the enzyme responsible for *N* addition (Chapter 4). PU.1 requires the presence of other factors in order to activate or repress their target genes. It cooperates with PU.1 Interaction Partner (PIP, LSIRF, IRF4), c-JUN, and c-FOS. PU.1-deficient mice demonstrate defective generation of monocyte, granulocyte, and lymphocyte progenitors, indicating a role in the generation of MPPs as well as LMPPs. PIP-deficient mice lack GCs in peripheral lymphoid organs and exhibit defects in B-cell activation.

Ikaros and Aiolos belong to the same zinc finger transcription factor family. Ikaros is expressed in stem cells and mature lymphocytes. Aiolos is only expressed after commitment to the B-cell lineage. Ikaros can generate several isoforms by means of differential splicing. Isoforms differ in their DNA-binding patterns, tendency to dimerize, and nuclear localization. Ikaros binds TdT, λ14.1 (λ5), VpreB, and LCK genes.

The *E2A* locus encodes two basic helix-loop-helix transcription factors that represent two alternately spliced products, E12 and E47. Targets for E2A include RAG1 and TdT. Although the functions of E12 and E47 overlap; E47 appears to play a greater role in driving TdT and RAG2, whereas E12 is a better activator of EBF and PAX5 and thus helps commit developing cells to the B-cell lineage.

ID-1 has a helix-loop-helix domain but lacks a DNA-binding domain. Thus, it can function as a dominant negative factor, inhibiting the function of helix-loop-helix transcription factors, such as E2A. ID-1 is expressed only in proB cells.

EBF, or early B-cell factor, is a helix-loop-helix like transcription factor. It is expressed at all stages of differentiation except plasma cells and is critical for the progression B cells past the early proB cell stage.

PAX5 is a paired-box, or domain, transcription factor that, among the progeny of hematopoietic stem cells, is expressed exclusively in cells of the B-cell lineage. PAX5 has both a positive and a negative effect on B-cell differentiation. The presence of PAX5 prevents early B-lineage progenitors from transiting into other hematopoietic pathways. Downregulation of PAX5 allows upregulation of BLIMP1 and plasma cell differentiation. PAX5 downregulation in plasma cells permits expression of genes typically expressed in macrophages and neutrophils.

A number of epigenetic modifiers affect numerous aspects of B-cell development.[12] These include polycomb complexes 1 and 2, histone deubiquitinases, histone acetyltransferases, histone deacetylases, and DNA methyltransferases. For example, histone modifications, DNA methylation, DNA looping, and

noncoding RNAs all play crucial roles in V(D)J recombination, classswitching, and SHM.

MicroRNAs, Long Noncoding RNAs, and B-Cell Development

MicroRNAs (miRNAs) are a class of small, noncoding RNAS that downregulate target genes at a post-transcriptional level.[13] These RNAs are derived from longer transcripts by the sequential action of RNA polymerase II, the nuclear nuclease Drosha, and the cytosolic nuclease Dicer. Mature miRNAs are incorporated into the multiprotein RNA-induced silencing complex (RISC), which repress target mRNAs by either inducing mRNA cleavage or mRNA degradation, or by blocking mRNA translation. miRNAs that play a role in both early and late B-cell development include miR-150, miR-155, and miR-17-92. Abnormal function of these miRNAs can contribute to oncogenesis and immune dysfunction.

On the other end of the spectrum, long noncoding RNAs (lncRNA) are transcripts with lengths exceeding 200 nucleotides that are not translated into protein. These transcripts can be cell-stage specific and they can interact with DNA, RNA, proteins, or combinations thereof. In so doing they can modulate three-dimensional chromatin structure, protein scaffolding, interfere with transcriptional gene regulation, and influence posttranscriptional modifications.[14] Mutation of the DNA that encodes these transcripts can have oncogenic or tumor suppressive functions.

Modulation of B-Cell Development by Chemokines, Cytokines, and Hormones

Stromal cells provide the microenvironment for B-cell development and differentiation (see Fig. 7.3). For example, the chemokine CXCL12, also known as preB cell growth-stimulating factor and as stromal cell-derived factor-1 (PBSF/SCF-1), promotes proB cell proliferation. Mice with a targeted disruption of this gene exhibit impaired B lymphopoiesis in fetal liver and bone marrow, and fail to undergo bone marrow myelopoiesis.[15] CXCL12 binds to CXCR4 signaling, which can activate diverse G-protein-coupled receptor (GPCR) pathways and lead changes in cell migration (Chapter 16), adhesion, and transcriptional activation. In IgM^+ IgD^+ mature B cells, CXCR4 associates with surface IgD.

Although in mice interleukin 17 (IL-17) plays an essential role in B-lineage differentiation, in humans IL-7 has a minimal proliferative effect on human B-cell progenitors. Nevertheless, IL-7 enhances CD19 expression, which plays an important role in BCR signal transduction (Chapter 4). IL-7 treatment of human proB cells leads to a reduction in the expression of RAG1, RAG2, and TdT, thereby modulating Ig gene segment rearrangement.

Interferons-α and -β (IFN-α/β) are potent inhibitors of IL-7-induced growth of B-lineage cells in mice.[16] The inhibition is mediated by apoptotic cell death. One potential source of IFN-α/β is bone marrow macrophages. Another macrophage-derived cytokine, IL-1, can also act as a dose-dependent positive or negative modulator of B lymphopoiesis.

Systemic hormones also regulate lymphopoiesis.[17] A role for sex steroids is suggested by the reduction in preB cells during pregnancy. Estradiol can also alter later stages of B-cell development, promoting expansion of the MZ compartment. Prolactin appears to enhance production of both MZ and follicular B cells. *Pit-1* transcription factor deficient mice do not produce growth hormone, prolactin, or thyroid-stimulating hormone. These dwarf mice exhibit a defect in B-cell development that is correctable by the thyroid hormone thyroxine.

KEY CONCEPTS

B-Cell Development in the Periphery

- T-independent activation of naïve B cells results in terminal differentiation into short-lived plasma cells.
- T-dependent activation of B cells:
  - Induces germinal center formation, permitting somatic hypermutation and classswitch recombination.
  - Results in differentiation into high-affinity memory B cells and plasma cells secreting high-affinity antibodies.
  - Generates long-term humoral immune protection.
- The longevity of plasma cells is supported by highly specialized survival niches in the bone marrow.
- T-follicular helper cells control late B-cell differentiation by cell-bound ligands and secreted cytokines.
- Activated B cells control T-cell development by presentation of antigen and co-stimulation.

Organization of the Peripheral Lymphoid Organs

During fetal development, the primary and secondary lymphoid organs follow an organized process of construction. This process involves multiple factors that play various roles in the development, maintenance, and function of these organs. The resultant compartmentalization of the lymphoid organs facilitates the efficiency and regulation of immune responses.

Each secondary lymphoid organ has a preferred route of entry for its constituent B cells.[18] For example, while most lymphocytes enter the spleen through the bloodstream, they enter lymph nodes and Peyer patches through high endothelial venules. Their migration and tissue-specific homing is strictly controlled by chemokines (Chapter 15). Dendritic cells, macrophages, and B cells transport antigens from peripheral sites of entry into the follicle within which circulating lymphocytes survey available antigens. Activation of T cells and B cells by the same antigen (cognate antigen recognition) permits initiation of a T-cell-dependent immune response.

The Spleen

In the secondary lymphoid organs, T cells and B cells are segregated into clearly defined areas (Chapter 2). The pattern observed in the white pulp of the spleen is illustrated in Fig. 7.4. It is within the white pulp that antigen-dependent B-cell activation occurs and where these activated cells subsequently undergo further differentiation. The marginal sinus, a lymphoid structure lined with a mucosal addressin cell adhesion molecule-1 (MAdCAM-1) expressing endothelium (Chapter 16), provides entry for lymphocytes, macrophages, and dendritic cells into the splenic tissue.[19] B cells in the adjacent MZ handle the incoming antigen and transport it in a continuous shuttle to the primary follicles, where it is captured and stored by follicular dendritic cells (FDCs). FDCs are stromal cells that present antigen to B cells (Chapter 6). In contrast to other types of dendritic cells, FDCs do not process antigen. Instead, FDCs have abundant complement receptors and Ig Fc receptors that allow them to accumulate antigen in the form of immune complexes within the B follicle. Antigen presentation by FDCs is crucial for maintenance of B cells as well as for their activation and differentiation (see below).

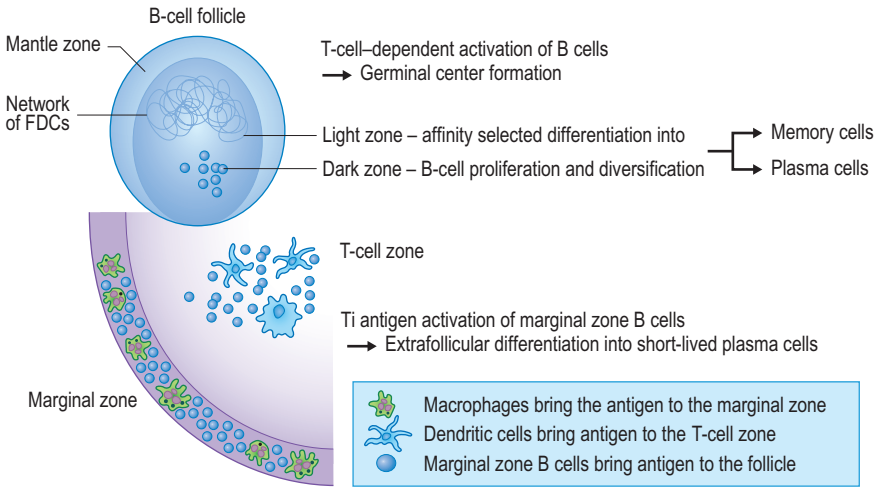


FIG. 7.4 T-Cell and B-Cell Compartments Within the White Pulp of the Murine Spleen. Most splenic T cells can be found in a compartment that surrounds the central arterioles. The compartment is also referred to as the periarteriolar lymphocyte sheath. Most splenic B cells are found in the adjacent follicles, where they are embedded in a network of follicular dendritic cells *(FDCs)*. In addition, there are B cells in the marginal zone, which is located next to the marginal sinus and marks the border of the white and the red pulp. The figure shows a secondary follicle where a dark and a light zone (germinal center) can be distinguished. Primary, follicular B cells are pushed back and form the mantle zone.

B-CELL DEVELOPMENT IN THE PERIPHERY

The life span of mature B cells expressing surface IgM and IgD appears entirely dependent on antigen selection. After leaving the bone marrow, unstimulated cells live for only a few days. Deletion of the transmembrane/intracellular domains of the BCR leads to loss of mature B cells, which indicates that signaling through the BCR is essential for their survival. As originally postulated by Burnet's "clonal selection" theory, a B cell is rescued from apoptosis by its response to an antigen that fits its BCR.

The reaction to antigen leads to activation, which can then be followed by diversification. The nature of the activation process is critical.[19] T-cell-independent stimulation of B cells induces differentiation into short-lived plasma cells with a limited ability to carry out class switching. T-dependent stimulation adds additional layers of diversification, including mutation (SHM) of the variable domains, class switch recombination (CSR) (Chapter 4), and differentiation into memory B cells or into long-lived plasma cells.

B-Cell Activating Factor of the Tumor Necrosis Family and a Proliferation-Inducing Ligand Play Key Roles in the Development of Mature B Cells

Emigrating, recently arisen B cells leave the bone marrow and continue their maturation in the periphery. They demonstrate progressively higher levels of IgD expression with a commensurate reduction of IgM. The splenic environment plays a key role in this maturation process. Immigrant splenic B cells pass through two transitional stages, termed transitional stages 1 (T1) and 2 (T2). Passage through this checkpoint requires the interaction of soluble B-cell activating factor of the tumor necrosis family (BAFF), with its receptor, BAFF-R, which is expressed primarily on B cells. Death signals triggered through interaction of the BCR with self-antigen can be counterbalanced by stimulation of BAFF-R, which enhances expression of survival factors such as Bcl-2 and at the same time downregulates pro-apoptotic factors. Only a minority of these cells successfully make the transition, as this differentiation step is a crucial checkpoint for controlling self-reactivity.[20,21]

BAFF and a second TNF family member APRIL (A Proliferation-Inducing Ligand) are essential factors for the development and long-term maintenance of B cells.[20] As activated B cells develop into plasma cells, BAFF-R is downregulated while the receptors TACI (transmembrane activator and calcium-modulator and cyclophilin ligand interactor) and BCMA (B cell Maturation Antigen) are upregulated. Mutations in TACI and BAFF-R have been associated with hypogammaglobulinemia (Chapter 33). In contrast to BAFF-R, these members of the TNF-R family bind both BAFF and APRIL. APRIL can induce isotype switching in naïve human B cells and—more important—it is a crucial survival factor supporting the longevity of plasma cells.

Marginal Zone B Cells

Functionally and developmentally distinct subsets of mature B cells exist. In the spleen, follicular B cells have a key role in adaptive immune responses that take up to a week to engage. In contrast, MZ B cells lie at the interface between the initial innate and the delayed adaptive immune response[22] with a response that can be measured in hours to a few days. The ability of MZ B cells to rapidly respond to encapsulated bacteria by differentiating into antigen-specific plasma cells helps keep such infections under control. The MZ in the spleen does not fully mature until after the age of 2, which explains why, in addition to individuals lacking a spleen due to a congenital defect or trauma, young infants commonly show a poor response to blood-borne infections with encapsulated bacteria.[23]

B1 B Cells

In addition to "conventional" MZ and follicular B cells (B-2 subset), a small population of innate B cells exists. These B1 B cells are tissue resident cells and are mainly found in the body's

cavities such as the peritoneal cavity.[24] B1 B cells do not participate in adaptive immune responses. Instead, they respond to Toll-like receptor (TLR) (Chapter 3) rather than BCR signaling. B1 cells have been best studied in mice. They preferentially develop from distinct progenitors that comprise a majority of B cells in fetal life. Accordingly, virtually all B cells in mouse fetal liver and 40% to 60% of B cells in fetal spleen are B-1 cells. Later in development, B-1 cells comprise less than 10% of the splenic IgM^+ B cells. Unlike B2 B cells, the B1 subset has self-renewal capacity, which requires autophagy.

Plasma cells originating from B1 B cells secrete IgM antibodies. These natural antibodies (NAb) are found in every serum and provide an important and immediate defense against many infectious organisms. Self-reactivity of these IgM antibodies appears to play a major role in tissue homeostasis.

The frequent presence of CD5 on chronic lymphocytic leukemia (CLL) B cells and their tendency to produce poly- and self-reactive antibodies led many to conclude that CLL was a leukemia of human B-1 cells. When it became clear that CD5 was not a definitive marker for B-1 cells in humans, considerable effort was expended searching for this elusive subset. Currently B cells that are $CD20^+$ $CD27^+$ $CD43^+$ $CD70^-$ appear to be the best candidates.

DIFFERENTIATION AND THE RESPONSE TO ANTIGEN

T-Independent Responses

Unlike T cells, which require presentation of antigen by other cells, B cells can respond directly to an antigen as long as antigen is able to cross-link the antibodies on the B-cell surface. Such antigens, especially those that by nature cannot be recognized by T cells (e.g., DNA or polysaccharides), can induce a B-cell response independent of T-cell help.[19] Depending on the cytokine milieu, the B cells may even undergo class switch, although the range of switching options appears to be restricted. B cells that are activated by antigen alone do not take part in a germinal center reaction (see below).

T-Dependent Responses

Activated B cells express both MHC class I and class II molecules (Chapter 5) on their cell surface. They can thus present both intracellular and extracellular antigens to CD4 T helper and CD8 T cytotoxic lymphocytes. Their role as antigen-presenting cells is enhanced when they present peptides to T cells derived from the same antigen that is bound to their BCR (Chapter 6). Cognate recognition of the same antigen by both a B cell and a T cell permits each of these cells to reciprocally activate the other. In particular, T-cell-activated B cells express the co-stimulatory molecules CD80 and CD86 (Fig. 7.5), which are in turn required for activation of T cells via CD28, and for inactivation by CD152 (CTLA-4). B cells also express the co-stimulator CD40, which interacts with CD40L expressed on T cells. Antigen presentation by B cells is a further requirement for the full differentiation of T cells into T follicular helper cells (Tfh).[19,25]

As Tfh provide cytokines such as IL-4 and IL-21, they have a crucial role in the formation of GCs (see below) and the differentiation of antigen-activated B cells into memory B cells and long-lived plasma cells.[26] These cytokines control divergent outcomes: upregulation of the transcription factor BCL-6 required for the development of GC B cells; activation of AID (activation-induced cytidine deaminase), which is a prerequisite for SHM and class switch (see below); downregulation of PAX-5 and upregulation of transcription factors promoting differentiation into long-lived plasma cells. How these fates are chosen is not yet fully understood (see Fig. 7.5).[19,26,27]

Germinal Center Reaction

T-cell-dependent activation of follicular B cells can induce the formation of GC, which provide the micro-environment in which affinity maturation of the humoral immune response takes place. Within these GC, B cells undergo multiple rounds of SHM and affinity selection after which cells that express BCR of high affinity may differentiate either into memory B cells or long-living plasma cells.[19,28]

During primary immune response, it takes about a week for the complex GC structure to develop in the spleen. A few days after activation of antigen-specific B cells and T cells, small clusters of proliferating B cells are observed at the border between the T-cell zone and the primary B-cell follicle. The rapidly expanding B-cell clone seems to push the naïve B cells toward the edge of the primary follicle so that the naïve B cells form a mantle zone around the newly developing GC. The primary follicle changes into a GC, also referred to as a secondary follicle. Subsequently, the FDC network becomes filled with proliferating, antigen-activated B cells. An influx of antigen activated Tfh follows. Tfh express the chemokine receptor CXCR5, which enables them to enter the B-cell follicle. During this GC reaction, expression of the chemokine CXCL13 by the FDC attracts both antigen activated B and Tfh cells.[25]

In the second week after immunization, the GC matures into a classical structure that contains a dark zone and a light zone (see Fig. 7.4). At this stage of GC development, proliferation is restricted to the dark zone. Within the network of FDC, B cells differentiate into plasma cells and memory cells. In a fully developed GC, dividing cells are termed centroblasts, whereas differentiating cells within the FDC network are termed centrocytes.[28]

In the dark zone, proliferating B cells activate SHM (Chapter 4).[29] This is a highly specific process that targets the gene segments that encode the V domains of the antibody molecule. SHM introduces single nucleotide changes into the sequences of the V domains. This enables production of clones of B cells expressing mutated antigen receptors with varying antigen affinity from a single B cell progenitor. By chance, a few of these mutations result in a receptor with higher affinity for antigen.

A highly efficient antigen-based affinity selection process takes place in the light zone of the GC where FDC present antigen to B cells. Only those B cells with high affinity receptors are able to internalize the antigen via their BCR. Processing of the internalized antigen and presentation of peptides to cognate Tfh cells are prerequisites for B-cell differentiation into memory and plasma cells.[25] The result is that only those B cells with high affinity receptors get adequate help. The cytokine IL-21 (Chapter 14) provided by the Tfh cells is crucial in this differentiation phase (see Fig. 7.5).[26] Upregulation of Bcl6 may support re-entry into the dark zone of the GC and hence further expansion of high-affinity B cell clones. Downregulation of PAX5 and concomitant upregulation of the TF BLIMP-1 will support plasma cell differentiation. Thus, IL-21 together with additional signals provided by Tfh and FDCs, controls the fate of the GC B cells.

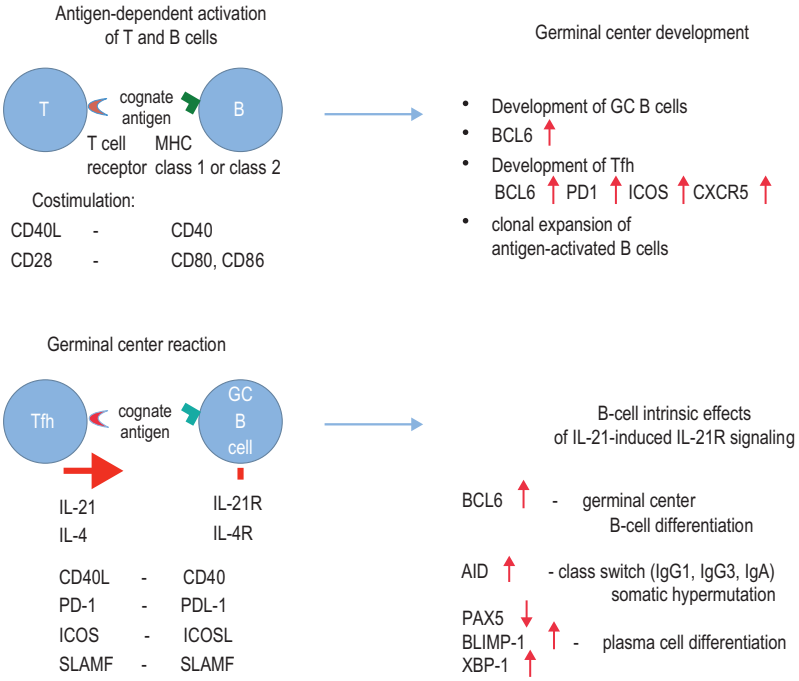


FIG. 7.5 T-Dependent B-Cell Activation and Germinal Centers Formation. Mutual activation of T and B cells requires cognate antigen recognition and co-stimulation. Both antigen activated B cells and T cells upregulate BCL6. Subsequently, the B cells differentiate into germinal center *(GC)* B cells and the T cells develop into T follicular helper cells *(Tfh)* cells. Proliferation of B cells leads to the formation of GC. Under the control of cytokines (e.g. interleukins *IL-21* and *IL-4*), GC B cells differentiate into long-lived memory B cells or into long-lived plasma cells. The differential expression of transcription factors, co-stimulatory factors, and the enzyme activation-induced cytidine deaminase is indicated.

Primary immune deficiencies, such as the hyperIgM syndrome (deficiency in CD40L/CD40 signaling, deficiency in expression of functional AID, failure to signal through the IL-21/IL-21R/STAT3 axis), demonstrated the pivotal role of the GC reaction, for protective humoral immune responses to vaccination or natural infections.

MOLECULAR MECHANISM OF SOMATIC HYPERMUTATION AND CLASS SWITCH RECOMBINATION

Ig SHM and CSR are essential mechanisms for the generation of a high-affinity, adaptive humoral immune response. They allow the generation of effector plasma cells secreting high-affinity IgG, IgA, and IgE antibodies. Both CSR and SHM are dependent on the enzyme AID.[29]

Somatic Hypermutation

Hypermutation occurs only during a narrow window in B-cell development. The mechanism is induced during B-cell proliferation within the microenvironment of the GC. Single nucleotide exchanges are introduced at a rate of about 10^{-3}/base pair/cell generation, into the rearranged V-region and its 3′ and 5′ flanking sequences. Mutations are randomly introduced, although there is a preference for transitions (cytidine <-> thymidine or adenosine <-> guanine) over transversions. Analysis of the pattern of somatic mutations has revealed that the sequence of the six (CDRs; Chapter 4), the loops that form the antigen-binding site, have been selected as mutation hot spots.

Effective hypermutation requires V-gene promoter and transcriptional enhancer sequences. The position of the V-gene promoter defines the start of the hypermutation domain, which spans about 2000 nucleotides. Any heterologous sequence that is introduced into the V-gene segment locus will become a target of the hypermutation machinery. For this reason, SHM can sometimes play a role in the development of lymphomas and leukemia in cases where oncogenes have been linked to Ig promoters and enhancers.

Class Switch Recombination

Upon leaving the bone marrow, the maturing B cell starts to express IgD as well as IgM. Both IgM and IgD antibodies use the same V_HDJ_H-exon and promoter. The molecular basis of co-expression of IgM and IgD by the same B cell is due to differential termination of transcription and splicing of the primary transcripts. Although sequences have been identified that are required for the control of termination and splicing of the Cμ and Cδ transcripts, none of the proteins involved are known. The role of IgD remains unclear, although IgM and IgD appear to create different types of signal transduction structures on the cell surface. In gene-targeted IgD-deficient mice, B cells show a slightly reduced capacity for affinity maturation, although B-cell activation and differentiation appears minimally affected.

Unlike IgD, the other antibody classes are not stably expressed together with IgM. B cells can switch from expression of their V_HDJ_H-exon with Cμ to expression of the same V_HDJ_H-exon with any of the downstream C_H genes (e.g., Cα1,2; Cγ1,2,3,4; or Cε) (Chapter 4) (Fig. 7.6).

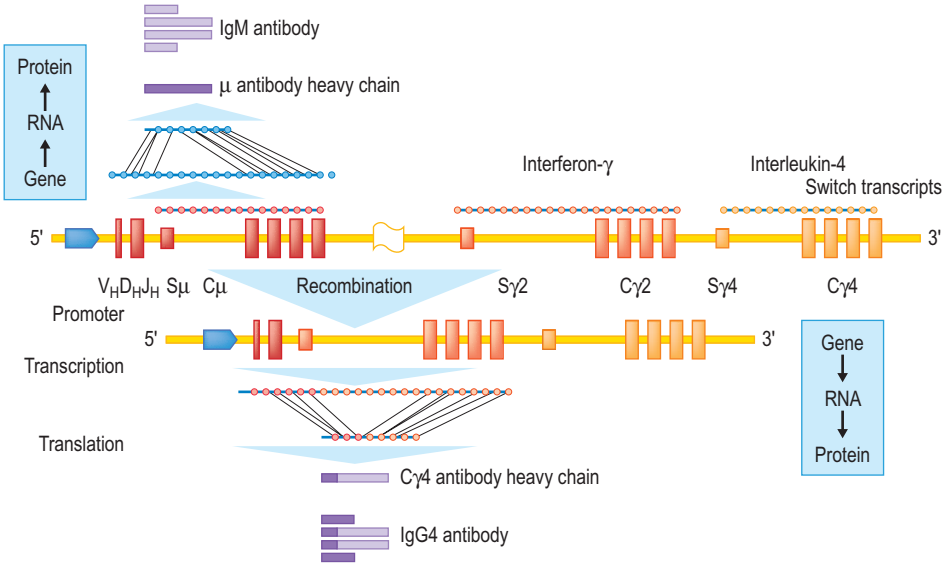


FIG. 7.6 Antibody Class Switch Recombination. Recombination between switch regions (Sμ and Sε) is preceded by transcription of these switch regions. Transcription is targeted by cytokines to distinct switch regions.

TATA-less promoters located in front of the switch regions initiate transcription just upstream of a small I-exon located between the promoter and the switch region. This transcription is essential in targeting CSR, as AID can bind only to single-stranded DNA.

The choice of C_H gene targeted for CSR in a particular B cell appears dependent on external cytokine signals recruiting exactly those classes of antibodies that provide the most useful functions for their respective branches of the immune system (Chapter 14). IL-21 favors class switch to IgG1 and IgG3 in humans, IL-4 to IgG4 and IgE in humans, and IgG1 and IgE in mice. IFN-γ targets CSR to IgG2 in humans and IgG2a in mice, and transforming growth factor-β (TGF-β) to IgA in both humans and mice.

Both Somatic Hypermutation and Class Switch Recombination Require Activation-Induced Cytidine Deaminase

Hypermutation occurs in two steps.[29] The mechanism is induced by AID-catalyzed deamination of deoxycytidine (C) to deoxyuridine (U). The mispairing of U and deoxyguanosine (G) is then processed by uracil DNA glycosylase (UNG) and targeted by repair pathways. This introduces mutations at C–G base pairs. In the second step, mutations at adjacent A–T pairs are induced, probably during a mutagenic patch repair of U–G mismatches introduced by AID. A number of proteins, such as MSH2 and MHS6 (homologues 2 and 6 of the *Escherichia coli* MutS), polymerase η, or exonuclease-1 seem to be involved.

In CSR, AID is targeted to the switch regions located upstream (5′) of each C_H gene.[29] These switch regions are composed of 1 to 6 kilobase-long GC-rich repetitive sequence motifs. Deamination of C and processing by UNG creates an abasic site that facilitates the introduction of double-strand DNA breaks. Joining and repair requires the presence of the DNA-phosphokinases Ku70, Ku80 and probably other members of the general double-strand repair mechanism (Chapter 4).

Both mechanisms, SHM and CSR, need to be tightly controlled, since the introduction of double-strand breaks into the DNA permit translocations involving and activating oncogenes.[30] For example, for Burkitt lymphoma and for plasma cell-derived myeloma, the translocation and ectopic expression of *c-MYC* is an apparent consequence of abnormal SHM and CSR.

B-CELL MEMORY

One of the key features of the immune system is immunological memory for previously encountered antigens. There are two layers of B-cell memory: long-lived memory B cells and protective humoral memory provided by long-lived plasma cells.[19,27,31] Memory B cells provide reactive memory, whereas long-lived plasma cells provide active protective memory. For efficient long-term humoral protection, both are required. Repeat exposure to a high concentration of original antigen may activate memory B cells and induce their differentiation, allowing them to rapidly generate new plasma cells that can secrete antibodies of high quality (Fig. 7.7).

Only plasma cells generated in a GC reaction can migrate to the bone marrow and be housed for long periods in highly specialized niches provided by underlying reticular stromal cells without the need for further activation and proliferation. Their maintenance is dependent on interaction with stromal cells and the survival factor APRIL. Eosinophils have been shown to be the main source of this cytokine and, when they are depleted, plasma cells rapidly go into apoptosis.[32]

The contribution of memory B cells and long-lived plasma cells varies. Some individuals are primarily protected by memory B cells and others by plasma cells.[33] This can be of vital importance in special situations such as transplantation where activation of the immune system should be avoided. For example, treatment of transplant recipients with rituximab, a monoclonal antibody specific for CD20, depletes memory B cells, but has no effect on long-living plasma cells, which may be secreting transplant-specific antibodies.

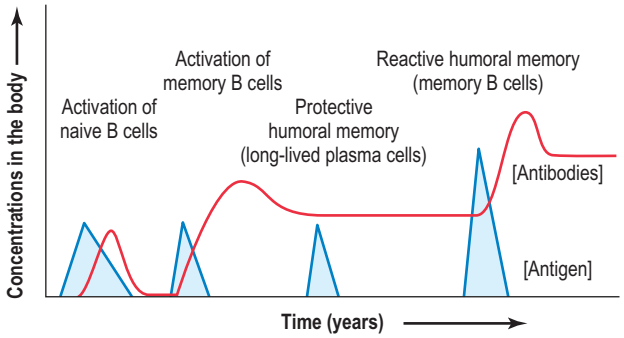


FIG. 7.7 Active and Reactive B-Cell Memory. Memory B cells provide reactive memory, whereas long-lived plasma cells provide active protective memory. The relative concentrations of antibody and antigen over time are indicated.

CLINICAL RELEVANCE

Abnormal B-Cell Development and Diseases of Immune Function

- Failure to generate B cells or a normal repertoire of antibodies leads to humoral immune deficiency, which is commonly marked by recurrent sinopulmonary infections.
- Failure to mount a germinal center reaction may also yield humoral immune deficiency.
- Failure to prevent the formation of antibodies with high avidity or high affinity to self-antigens can lead to autoimmune diseases.
- The process of antibody repertoire diversification lends itself to the creation of mutations that can activate and modify oncogenes as well, promoting leukemia or lymphoma. Mechanisms include:
  - RAG1/2-catalyzed juxtaposition of an oncogene to an immunoglobulin promoter or enhancer, activating the oncogene.
  - AID-induced DNA double-strand breaks and chromosome alterations.
  - AID-induced somatic hypermutation of oncogene, altering its function.

ECTOPIC LYMPHOID TISSUE AND B-CELL DEVELOPMENT

Ectopic lymphoid tissue may develop in the affected tissue or organ in autoimmune diseases, in chronic infection, and in tumors (cancer). Inflammatory cytokines and the presence of B cells support the development of these additional "tertiary" lymphoid tissues.[34]

The growth of ectopic lymphoid tissue in rheumatoid synovium (Chapter 53) offers an excellent example of this disease-related phenomenon. In healthy individuals, the synovium is made up of a thin lining layer of synoviocytes. In contrast, in patients with rheumatoid arthritis the diseased joint is highly infiltrated with varying numbers of T cells, B cells, plasma cells, macrophages, and dendritic cells. In the majority of patients, these mononuclear cells are dispersed loosely throughout the synovium. In some patients, large, well-organized, lymphoid structures can develop that are similar in appearance to the lymphoid follicles seen in the secondary lymphoid organs.[34] At the center of these cell clusters one finds a network of FDCs. Antigen presented by the FDC appears to activate B cells, which induces proliferation. The central B cells are surrounded by a layer of T cells, which may support local B-cell differentiation. A central question concerns the antigens that drive these immune responses and select B cells to differentiate into memory and plasma cells. Ectopic GCs may support an autoimmune response.

ON THE HORIZON

- A better understanding of the regulation of B-cell development by epigenetic modifications and by noncoding RNA as well as the mechanisms and antigens that support the development of ectopic germinal centers could lead to the definition of previously unknown mechanisms of immune deficiency or autoimmunity.
- Elucidation of the mechanisms used to control the antibody repertoire and shape B-cell epitope recognition offer the promise of being able to direct immunity toward production of broadly neutralizing or antitumorigenic antibodies, and away from pathogenic autoantibodies.
- Elucidation of the mechanisms that prevent the development of self-reactivity during affinity maturation could yield new insights into autoimmunity, as well as vaccination.
- A better understanding of the mechanisms that control the production of memory B cells or lead to long-lived plasma cells will pave the way to vaccination strategies that ensure long-term protection.

B-CELL FUNCTIONS IN ADDITION TO ANTIBODY PRODUCTION

B cells play an important function in the activation of T cells, because—like dendritic cells (Chapter 6)—B cells can internalize, process, and present MHC-bound antigen to the TCR on T cells. In cancer, B cells can secrete tumor-associated autoantibodies, inflammatory cytokines, and alter patterns of antigen presentation to T cells. By doing so they may modulate T-cell and innate immune responses to the tumor. By forming antigen–antibody complexes, B cells have the potential to influence immune cells such as granulocytes and natural killer cells that express Fc receptors. In autoimmune diseases, and also in response to inflammation, B regulatory cells (Breg) can have an immune suppressive function.[35] These Breg appear to exert their immunosuppressive effects on Th1 cells via the release of IL-10 and cell–cell contact.

REFERENCES

1. Petkau G, Turner M. Signalling circuits that direct early B-cell development. *Biochem J*. 2019;476(5):769–778.
2. Kunisawa J. Metabolic changes during B cell differentiation for the production of intestinal IgA antibody. *Cel Mol Life Sci: CMLS*. 2017;74(8):1503–1509.
3. Elsaid R, Yang J, Cumano A. The influence of space and time on the establishment of B cell identity. *Biomed J*. 2019;42(4):209–217.
4. Franks SE, Cambier JC. Putting on the brakes: regulatory kinases and phosphatases maintaining B cell anergy. *Front Immunol*. 2018;9:665.
5. Griffin DO, Holodick NE, Rothstein TL. Human B1 cells are CD3⁻: A reply to "A human equivalent of mouse B-1 cells?" and "The nature of circulating CD27⁺CD43⁺ B cells." *J Exp Med*. 2011;208(13):2566–2569.
6. Mei HE, Wirries I, Frolich D, et al. A unique population of IgG-expressing plasma cells lacking CD19 is enriched in human bone marrow. *Blood*. 2015;125(11):1739–1748.
7. Rajewsky K. Clonal selection and learning in the antibody system. *Nature*. 1996;381(6585):751–758.
8. Melchers F. Checkpoints that control B cell development. *J Clin Invest*. 2015;125(6):2203–2210.
9. Khass M, Blackburn T, Burrows PD, et al. VpreB serves as an invariant surrogate antigen for selecting immunoglobulin antigen-binding sites. *Sci Immunol*. 2016;1(1):aaf6628.
10. Lucas F, Woyach JA. Inhibiting Bruton's tyrosine kinase in CLL and other B-cell malignancies. *Target Oncol*. 2019;14(2):125–138.
11. Li X, Ding Y, Zi M, et al. CD19, from bench to bedside. *Immunol Lett*. 2017;183:86–95.

12. Martin-Subero JI, Oakes CC. Charting the dynamic epigenome during B-cell development. *Semin Cancer Biol*. 2018;51:139–148.
13. Aalaei-Andabili SH, Rezaei N. MicroRNAs (MiRs) precisely regulate immune system development and function in immunosenescence process. *Int Rev Immunol*. 2016;35(1):57–66.
14. Winkle M, Kluiver JL, Diepstra A, van den Berg A. Emerging roles for long noncoding RNAs in B-cell development and malignancy. *Crit Rev Oncol Hematol*. 2017;120:77–85.
15. Nagasawa T, Hirota S, Tachibana K, et al. Defects of B-cell lymphopoiesis and bone-marrow myelopoiesis in mice lacking the CXC chemokine PBSF/SDF-1. *Nature*. 1996;382(6592):635–638.
16. Wang J, Lin Q, Langston H, Cooper MD. Resident bone marrow macrophages produce type 1 interferons that can selectively inhibit interleukin-7-driven growth of B lineage cells. *Immunity*. 1995;3(4):475–484.
17. Grimaldi CM, Hill L, Xu X, et al. Hormonal modulation of B cell development and repertoire selection. *Mol Immunol*. 2005;42(7):811–820.
18. Schulz O, Hammerschmidt SI, Moschovakis GL, Forster R. Chemokines and chemokine receptors in lymphoid tissue dynamics. *Ann Rev Immunol*. 2016;34:203–242.
19. Cyster JG, Allen CDC. B cell responses: cell interaction dynamics and decisions. *Cell*. 2019;177(3):524–540.
20. Mackay F, Schneider P. Cracking the BAFF code. *Nat Rev Immunol*. 2009;9(7):491–502.
21. Rawlings DJ, Metzler G, Wray-Dutra M, Jackson SW. Altered B cell signalling in autoimmunity. *Nat Rev Immunol*. 2017;17(7):421–436.
22. Cerutti A, Cols M, Puga I. Marginal zone B cells: virtues of innate-like antibody-producing lymphocytes. *Nat Rev Immunol*. 2013;13(2):118–132.
23. Carsetti R, Rosado MM, Wardmann H. Peripheral development of B cells in mouse and man. *Immunol Rev*. 2004;197:179–191.
24. Baumgarth N. A hard(y) look at B-1 cell development and function. *J Immunol*. 2017;199(10):3387–3394.
25. Vinuesa CG, Linterman MA, Yu D, MacLennan IC. Follicular helper T cells. *Ann Rev Immunol*. 2016;34:335–368.
26. Tangye SG, Ma CS. Regulation of the germinal center and humoral immunity by interleukin-21. *J Exp Med*. 2020;217(1):e20191638.
27. Nutt SL, Hodgkin PD, Tarlinton DM, Corcoran LM. The generation of antibody-secreting plasma cells. *Nat Rev Immunol*. 2015;15(3):160–171.
28. Mesin L, Ersching J, Victora GD. Germinal center B cell dynamics. *Immunity*. 2016;45(3):471–482.
29. Methot SP, Di Noia JM. Molecular mechanisms of somatic hypermutation and class switch recombination. *Adv Immunol*. 2017;133:37–87.
30. Shaffer AL, Rosenwald A, Staudt LM. Lymphoid malignancies: the dark side of B-cell differentiation. *Nat Rev Immunol*. 2002;2(12):920–932.
31. Manz RA, Hauser AE, Hiepe F, Radbruch A. Maintenance of serum antibody levels. *Annu Rev Immunol*. 2005;23:367–386.
32. Chu VT, Berek C. The establishment of the plasma cell survival niche in the bone marrow. *Immunol Rev*. 2013;251(1):177–188.
33. Mamani-Matsuda M, Cosma A, Weller S, et al. The human spleen is a major reservoir for long-lived vaccinia virus-specific memory B cells. *Blood*. 2008;111(9):4653–4639.
34. Pitzalis C, Jones GW, Bombardieri M, Jones SA. Ectopic lymphoid-like structures in infection, cancer and autoimmunity. *Nat Rev Immunol*. 2014;14(7):447–462.
35. Mauri C, Menon M. Human regulatory B cells in health and disease: therapeutic potential. *J Clin Investig*. 2017;127(3):772–729.
36. Manjarrez-Orduno N, Quach TD, Sanz I. B cells and immunological tolerance. *J Investig Dermatol*. 2009;129(2):278–288.

8

Immunoglobulin Function

Neil S. Greenspan

Antibody-mediated immunity generally requires noncovalent contact between an antibody and the antigen. The ability of an antigen to bind noncovalently to an antibody, termed *antigenicity*, is a physical-chemical property evaluated with respect to a given antibody population. In contrast, *immunogenicity*, the ability to induce the biosynthesis and secretion of soluble antibody molecules, is a biological property requiring in vivo studies.

While antigenicity is necessary for immunogenicity (as defined by the production of antibodies), it is not sufficient. Moreover, the immunogenicity of a given molecule or molecular complex is influenced by host genetic variation.

When an antibody binds to a macromolecular antigen, it directly contacts only a portion of the molecular surface of that antigen. Similarly, only a portion of the antibody molecule makes direct contact with the antigen. By convention, the portion of an antibody or T-cell receptor that makes physical contact with an antigen is referred to as the *paratope* or combining site. Conversely, the region of the antigen in physical contact with the paratope, the antigenic determinant, is termed the *epitope*. Most of the amino acids in an antibody-variable domain that contact a given antigen are located in the hypervariable regions (also termed complementarity-determining regions or CDRs), although some contact residues reside in the framework regions (Chapter 4).[1] Although an epitope (paratope, etc.) is usually defined in terms of intermolecular contact, the region of a molecule involved in physical contact with another molecule may not correspond exactly to the structural correlates for energetics or specificity.[2]

ANTIGEN BINDING AND MOLECULAR IDENTITY

Physical Aspects of Binding

Antibody–antigen interactions are, with some exceptions, noncovalent. This fact is significant in that these interactions are, in principle, spontaneously reversible under the conditions of temperature, pressure, pH, and ionic strength that generally prevail in living organisms.

Several types of noncovalent bonds have been shown to contribute to antibody–antigen binding. These include van der Waals forces, hydrogen bonds, ionic bonds, and hydrophobic interactions. Individually, the strength of these bonds is in the range of one to a few kcal/mole, versus 50 to 100 kcal/mole for covalent bonds. Since the potential to engage in these types of bonds is shared by many of the components of biological macromolecules, individual weak bonds do not usually confer a high degree of specificity, so it is only through the simultaneous action of many such bonds that molecular specificity arises. Hence, the importance of a close fit, often referred to as *complementarity*, between the epitope and the paratope.

Complementarity can be maximized by matching the physical-chemical properties of the epitope and paratope. For example, binding can be facilitated when one molecule is concave and the other is convex, when one molecule is positively charged and the other is negatively charged, or when one molecule is a hydrogen bond donor whereas the other offers a hydrogen bond acceptor. It is expected that the greater the complementarity between receptor and ligand the stronger the interaction (greater affinity) between the two molecules. Complementarity (see below) is also expected to influence specificity.[2] In understanding the strength of interactions between antibodies and antigens, it is relevant that the antibody competes with solvent for binding to antigen. Thus, the thermodynamics of the interaction between these two structures reflects the influence of the interaction on the solvent and other solutes. Moreover, bound water molecules may make important, even crucial, contributions to an interaction between two biomolecules.

Antibody recognition of antigen serves as a paradigm for understanding molecular recognition in the immune system and in biology in general. This fact, coupled with the inducibility of antibodies, permits antibodies to be used as surrogate ligands for almost any receptor (or vice versa).

Affinity is the concept used to convey how strongly two molecules bind to each other. Reflecting the different types of antibody–antigen interaction, two categories of affinity merit consideration: *intrinsic affinity* and *functional affinity*. It should be noted that some immunologists use the term *avidity* in place of *functional affinity*.

Intrinsic affinity is a measure of the strength of the monovalent interaction between a particular paratope and a particular epitope under defined conditions of temperature, pressure, ionic strength, and pH (Fig. 8.1). By convention, the intrinsic affinity is taken to be the equilibrium association constant characterizing the paratope–epitope pair. It is the reciprocal of the concentration of monovalent antigen at which half of the paratopes will be occupied. It is not an intrinsic property of either the paratope or the epitope but rather characterizes the relationship between two molecules under defined conditions.

The intrinsic affinity of an antibody for a small molecule, such as a drug (e.g., digoxin) or a hormone (e.g., insulin), can be clinically important both in vivo and in vitro.[3] For example, the in vivo effectiveness of antibody F(ab) fragments in removing toxic levels of the drug digoxin from patients being treated for congestive heart failure likely depends on the intrinsic affinity of the F(ab) fragments for the drug. Alternatively, antibody intrinsic affinity can limit the analytical sensitivity of an in vitro immunoassay designed to determine the concentration of an

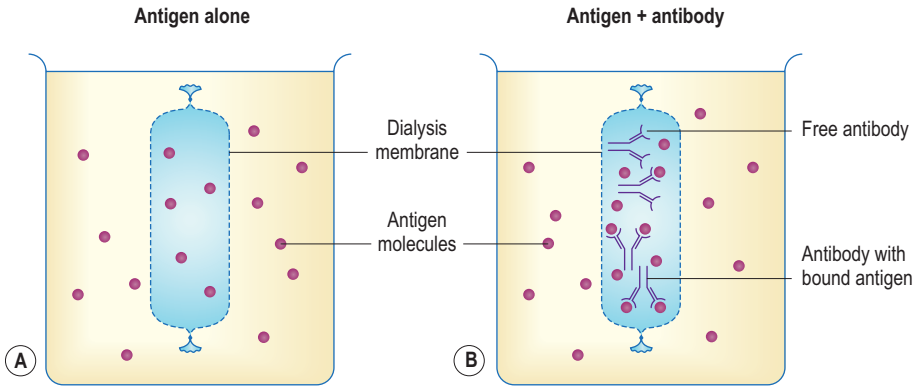


FIG. 8.1 Measurement of the Intrinsic Affinity Characterizing an Interaction Between Antibody and Antigen (Hapten) by Equilibrium Dialysis. At equilibrium (A), the amount of diffusible free hapten inside the dialysis bag will be equal to the amount of free hapten outside of the dialysis bag. However, in the presence of hapten-specific antibody (B) the total hapten concentration will be greater inside of the dialysis bag (free + antibody-bound) than outside of the bag (free). The extent of this difference can be used to determine the intrinsic affinity of the antibody for the hapten. (With permission from Abbas AK, Lichtman AH, Pober JS. *Cellular and Molecular Immunology*. W. B. Saunders Company, Philadelphia, PA; 1991.)

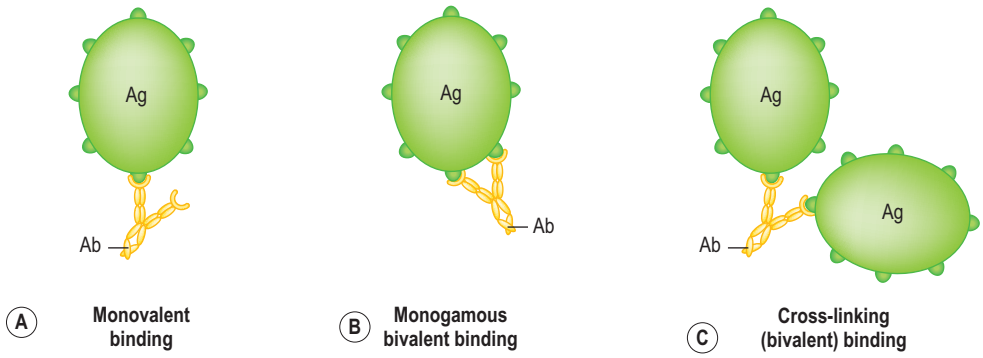


FIG. 8.2 Interaction of a bivalent antibody *(Ab)*, such as immunoglobulin G (IgG), with multivalent antigens *(Ag)* can result in monovalent binding (A), monogamous bivalent binding (B) or cross-linking (C). The complexes in (B) are referred to as cyclic antibody–antigen complexes. *Ag*; antigen; *Ab*; antibody. (With permission from Eisen HN. *General Immunology*. J. B. Lippincott Company, Philadelphia, PA; 1990.)

analyte, such as a hormone (e.g., insulin, parathyroid hormone), or a drug (e.g., digoxin).

In contrast, functional affinity is defined as the equilibrium association constant characterizing the interaction between an *intact antibody* and an *intact antigen*. For a monovalent immunoglobulin G (IgG) antibody–antigen interaction, the intrinsic affinity and the functional affinity will be the same. However, if two paratopes interact simultaneously with two epitopes on the same antigen, referred to as *monogamous* bivalency (Fig. 8.2), the functional affinity of the antibody for the *multivalent* antigen may be substantially greater (as much as 10,000-fold for IgG) than the intrinsic affinity of that antibody for the relevant epitope on that same antigen.[2]

Functional affinity is also influenced by the degree to which the geometric relationships among the epitopes are optimal for the paratopes, which will depend on the quaternary structure and segmental flexibility of the antibody molecule. In the presence of nonoptimal geometry, the average number of engaged sites may be less than maximal, and energy may be expended in achieving some epitope–paratope contacts. Therefore, the functional affinity for a multivalent interaction does not necessarily increase in direct relationship to the maximal number of binding sites that can be engaged simultaneously by an antibody molecule. For example, the effective valency of pentameric IgM with 10 paratopes can be less than 10.

Both concepts of affinity are valuable. Maximization of intrinsic affinity may be of prime importance for antibody-mediated inactivation of toxins or enzymes, which frequently involve monovalent interactions. However, in cases where antibodies bind to repeated epitopes on the surfaces of pathogens or mammalian cells, the functional affinity may play a larger role influencing the biological consequences of the interaction.

Bivalent (IgG, IgE) or multivalent (IgA, IgM) antibodies carry with them the potential to bind simultaneously to two or more epitopes on different antigenic particles, cross-linking them rather than engaging in monogamous bivalency or monogamous multivalency (see Fig. 8.2). This phenomenon has played an important historical role in immunology, such as for typing erythrocyte antigens (e.g., ABO and Rh antigens), which still relies on agglutination of red cells by antibodies (or lectins), and can contribute to host defense by clumping pathogens.

IMMUNOLOGICAL SPECIFICITY

The concept of specificity is fundamental to an understanding of the nature and consequences of interactions between

immunological receptors and antigens. However, in the immunological context, the term specificity encompasses multiple different senses, as discussed below.[2]

One aspect of specificity focuses on the goodness of fit between the paratope and the epitope. Intrinsic affinity is regarded as a reasonable measure of this goodness of fit. However, substantial conformational adjustments of either the paratope or the epitope may be necessary for formation of the complex.[4] Such conformational changes will generally incur energetic costs. Consequently, intrinsic affinity and final complementarity may not be perfectly correlated.

A second aspect of specificity focuses on the ability of a paratope to distinguish among different epitopes. Such specificity is most readily studied when the epitope is in monovalent form and evaluated relative to a specified set of ligands. Thus, one should be cautious about extrapolating claims that one antibody is more or less specific than another antibody without any reference to the relevant universe of ligands. However, there are practical cases where it is justifiable to speak globally of more-or less-specific antibodies. B-1 cells (Chapter 7) often produce polyspecific natural antibodies (NAb).[5] NAb appear to be globally less discriminating than antibodies typical of the B-2 immune repertoire (secondary or later response) when tested on large panels of antigens.

Nevertheless, it is important to note that even antibodies derived from secondary (or later) responses are not,[6] and cannot be, absolutely specific for both thermodynamic and structural reasons.[2] Recent results also indicate that at least some antibodies can adopt two or more different unbound conformations, each of which exhibits a different ligand-binding profile. Such paratopes may undergo further structural adjustment in the process of binding to an epitope.[7] This property can be advantageous to the function of an antibody. Antibodies that react with multiple conformations of a viral surface antigen may be much more likely to interfere with viral infection because they can bind more rapidly to the virus than the virus can bind to its receptor, as demonstrated for human immunodeficiency virus (HIV).[8]

Whereas the first two aspects of specificity focus on the epitope, a third relates to the ability of an antibody to discriminate among antigens that display multiple copies of one or more distinct epitopes. An antigen expressing many copies of one epitope is termed *multivalent*, and an antigen that expresses two or more structurally distinguishable epitopes is referred to as *multideterminant* (Table 8.1). Because two different cells, bacteria, viruses, and so on, may both express multiple copies of the same or nearly the same epitope, an antibody that is highly specific (in the first aspect above) for such a shared epitope may be a poor discriminator between such multivalent particles.[2] Yet, an antibody with a relatively poor degree of complementarity and intrinsic affinity for an epitope found on only one of two or more multivalent targets may be superior at discriminating among these antigens. Furthermore, antibodies (or other molecules) expressing two or more binding sites of identical structure may not discriminate identically among antigens displaying the same epitope in different two- or three-dimensional distributions.[2]

We offer some final points regarding specificity. First, the interactions between molecules such as CD4 (Chapter 9) and major histocompatibility complex (MHC) class II (Chapter 5), which are not clonally distributed, are often described as nonspecific, meaning "not specific for an antigen under consideration." Second, for many purposes immunological specificity has an ultimately biological, not a physical, definition. Third, the enormous utility of antibodies is crucially dependent on the discriminatory abilities of these molecules with respect to other molecules or molecular aggregates, so the usefulness of a particular antibody may depend on the context: for example, what particular antigens or potential antigens, in addition to the preferred target, are available for binding to the antibody. Fourth, apparent antibody specificity may vary with the methods used for analysis, as these methods may differ in sensitivity and in the conditions (pH, ionic strength, temperature) of application, such that the relevant intrinsic affinities may vary among the different assays.

TABLE 8.1 Antigens and Valence

Number of Types of Epitope	Epitope Copy Number	Examples
Monodeterminant	Monovalent	Hapten: DNP, digoxin
Monodeterminant	Multivalent	Polysaccharide: dextran[a]
Multideterminant[b]	Monovalent	Monomeric protein: myoglobin
Multideterminant	Multivalent	Virion: influenza virus

[a]Even a polysaccharide composed of one type of hexasaccharide can have two or more different kinds of epitope: terminal versus internal residues, for instance. However, a given antiserum may preferentially contain antibodies specific for only one such epitope.
[b]Typically, multideterminant recognition occurs with respect to a polyclonal antibody.
Adapted from Benjamini E, Leskowitz S. *Immunology: A Short Course.* 2nd ed. New York: Wiley-Liss; 1991, with permission from Wiley-Liss, Inc., a subsidiary of John Wiley & Sons, Inc.

PROTEIN EPITOPES

Based on the proximity of the relevant amino acids in the primary structure of the protein, several categories of epitopes have been defined for protein antigens (Fig. 8.3). The simplest is the *linear epitope*, where all of the relevant amino acids are derived from a contiguous, or nearly contiguous, stretch of the polypeptide chain. However, many epitopes on globular proteins involve amino acids from two or more stretches of polypeptide that are distant from one another in the primary structure but juxtaposed in the secondary or tertiary structure. These are referred to as a *conformational*, or *discontinuous*, *epitopes*. A conformational epitope may contain amino acids that are derived from separate but proximate polypeptide chains, as might be the case on nonenveloped viruses.

Another category of protein epitope, the *neo-epitope*, is reserved for those antigenic sites that become recognizable only after a post-translational event, such as phosphorylation or proteolytic cleavage. For example, neo-epitopes have been defined on cleavage products of human C1q, C3, and C9, which are components of the complement pathway (Chapter 40).[9] Antibodies recognizing such neo-epitopes can be used to monitor the extent of activation of the complement pathway.[9] A chimeric antibody ensituximab, which targets colorectal and pancreatic carcinoma–associated neo-epitopes, represents the therapeutic potential of targeting cancer neo-antigens (Chapter 80). In rheumatoid arthritis, autoantibodies to citrullinated epitopes on such antigens as filaggrin can be of diagnostic value (Chapter 53).

Studies in the 1970s on the sizes of epitopes associated with synthetic peptide antigens yielded results suggesting that protein epitopes would maximally involve six or seven amino acids. However, the first x-ray crystallographic structure of an

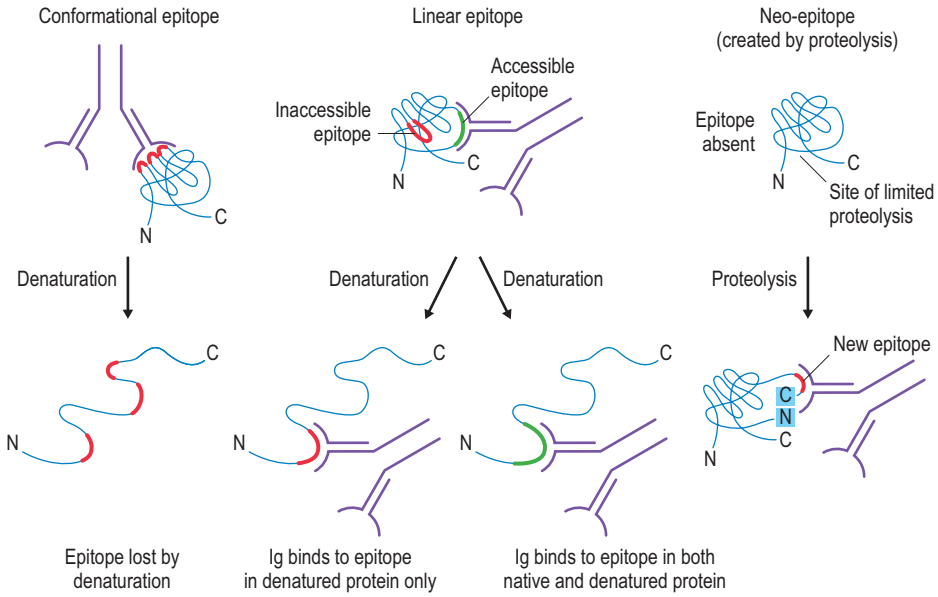


FIG. 8.3 Types of Protein Epitopes. Some antibodies recognize structural features of proteins that arise from the folding of the polypeptide backbone (conformational epitope). Others recognize groups of amino acid residues that are contiguous, or nearly so, in the primary (covalent) structure of the protein (linear epitope). If such a linear determinant is inaccessible in the native structure of the protein, the corresponding antibodies may only be elicited by the denatured form of the protein. Neo-epitopes are created by covalent post-translational modifications, such as phosphorylation or proteolytic cleavage. (With permission from Abbas AK, Lichtman AH, Pober JS. *Cellular and Molecular Immunology*. 3rd ed. W. B. Saunders Company; 1997.)

antibody variable module in complex with a globular protein antigen[1] indicated that protein epitopes, defined on the basis of intermolecular contact, could be as large as 15 to 20 amino acids. A similar number of amino acids in the antibody V domains constituted the paratope. Even peptide antigen–antibody interaction can involve at least 12 peptide amino acids in contact with the antibody. Still, it is possible that there are smaller epitopes on globular proteins, particularly for regions of proteins that protrude or have a high radius of curvature.

Antibodies specific for both linear and conformational epitopes have important practical applications. For example, a synthetic peptide corresponding in amino acid sequence to a segment of the polypeptide chain predicted from the nucleotide sequence can be used to elicit antibodies useful in identifying the protein following expression, electrophoresis, and blotting under denaturing conditions. Antibodies raised by challenge with synthetic peptides that bind to linear epitopes and that recognize a protein in denatured form often do not bind to or alter the function of the native protein.

Antibodies with the ability to neutralize protein function generally recognize conformations accessible to the native protein, usually at discontinuous epitopes. However, antibodies specific for peptides (that correspond in amino acid sequence to a portion of a native protein) or denatured protein, which can cross-react with the protein in a native (folded, functional) state can be extremely valuable. Such cross-reactivity is more likely to occur when the region being recognized is relatively disordered in the native structure.[10]

CARBOHYDRATE EPITOPES

The classical studies of Kabat on the binding of antibodies to dextran led to the concept that epitopes on carbohydrate antigens could be as large as six or seven monosaccharides. However, minimal carbohydrate epitopes can probably be as small as one or two monosaccharides. Even in the case of larger epitopes, it is typical for the terminal groups to play a dominant role in determining antibody specificity for carbohydrate antigens. Recent studies suggest that polysaccharide epitopes can sometimes also result from the conformational properties of the polysaccharides.

In comparison to antibody–protein interactions, interactions between antibodies and polysaccharides have typically been characterized by relatively low intrinsic affinities.[2] Relatively weak antibody–carbohydrate binding can result from biological constraints related to protection against self-recognition and consequent tissue damage, or from physical-chemical constraints related to the conformational freedom and high degree of solvation of unbound carbohydrates. On the other hand, antibodies produced in response to pathogens, such as HIV, may be much more effective at interacting with carbohydrate antigens.[11]

Another important feature of polysaccharide antigens is that they are generally multivalent. Bacterial and perhaps viral polysaccharide epitope densities can approach values in the millions per square micrometer, which is probably one to several orders of magnitude greater than the epitope densities for protein determinants on mammalian cells. Therefore, multipoint attachment and functional affinity are likely to be critical factors in the mediation of immunity by anti-polysaccharide antibodies or other carbohydrate-specific proteins.

IMMUNE COMPLEXES IN VIVO

Interactions between antibodies and antigens in vivo can result in the formation of molecular aggregates, referred to as *immune complexes*. Deposition of immune complexes in tissues, such as blood vessels, renal glomeruli, renal tubules, the thyroid gland,

and the choroid plexus, can result in pathological conditions.[12] Immune complexes can form in the circulation prior to deposition in a given tissue, or they can form directly in the affected tissue. A clinical situation in which immune complex formation can be a cause for concern is the production of anti-antibodies in response to therapy with pooled intravenous IgG (IVIg) (Chapter 82) or therapeutic monoclonal antibodies.

Variables such as concentration, composition, size, charge, and antibody isotype will influence the magnitude and sites of tissue deposition for these complexes. In conjunction with the sites and extent of tissue deposition, the magnitude of complement activation and the extent of interaction with Fc and complement receptors determine the biological properties of the complexes. Antigen–antibody lattice size is determined by antigen valence, epitope geometry, antibody valence, the intrinsic affinity of paratope for epitope, antibody and antigen flexibility, the ratio of antibody to antigen, and the absolute concentrations of antibody and antigen. The potential diversity of immune complex morphologies is illustrated in Fig. 8.4. These complexes, between a monoclonal antibody specific for a bacterial polysaccharide and various anti-idiotypic or anti-isotypic monoclonal antibodies, are visualized by electron microscopy.

Immune complexes have also been found to have immunoregulatory effects,[13] particularly with respect to antibody responses. Immune complexes can bind simultaneously to B-cell surfaces through antigen (to B-cell-surface immunoglobulin), antibody (to Fc receptors [FcR]), and associated complement components (to complement receptors). The interaction with FcγRIIB on the B-lymphocyte membrane has the effect of diminishing the B-cell response (Chapter 4). The molecular events underlying these immunoregulatory effects are being studied, and they have been clinically exploited for many years. Antibody to the erythrocyte Rh antigens is used to prevent immunization of an Rh^- mother by an Rh^+ fetus, thereby avoiding hemolytic disease of the newborn in a subsequent Rh^+ fetus.

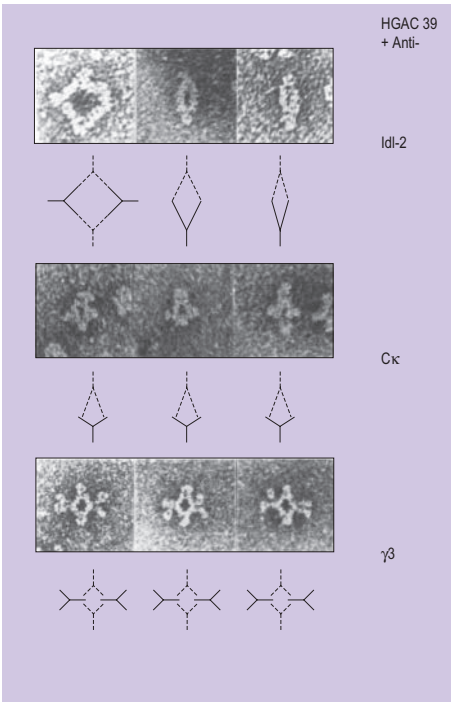


FIG. 8.4 Electron micrographs (*above*; × 350,000) and interpretive diagrams (*below*) of a mouse IgG3 mAb (HGAC 39; specific for a bacterial polysaccharide) in complex with mAbs specific for, respectively, an idiotope (*Idl-2*; *top*), a light-chain isotypic determinant (*Cκ*; *middle*), and a heavy-chain isotypic determinant (γ_3; *bottom*). The different antibodies are not intrinsically distinguishable in the electron micrographs, but the interpretations are based on information in addition to that provided directly by the electron microscopic images. In the top series of micrographs, the choice of which molecules to represent as *solid* or *dotted figures* is arbitrary. *mAb*, monoclonal antibody. (With permission from Greenspan NS. Analyzing immunoglobulin functional anatomy with monoclonal anti-immunoglobulin antibodies. *BioTechniques*. 1989;7:1086.)

CORRELATIONS BETWEEN C_H REGION STRUCTURE AND ANTIBODY FUNCTION

Antibodies are heterodimeric proteins that can be functionally divided into V domains, which bind antigen, and C domains, which define the effector function(s) of the immunoglobulin (Chapter 4). This division of labor allows an antibody to physically link a specific antigen to separate antigen-nonspecific effector molecules. Many features of C_H domain structure exhibited by the immunoglobulin isotypes can be understood in the context of this requirement for physical and/or functional linkage between antigens and antigen-nonspecific effector molecules.

One property of prime significance for antibody function is intramolecular mobility, often referred to as *segmental flexibility*. Hydrodynamic methods, electron microscopy, x-ray crystallography, and fluorescence polarization, have all been used to evaluate the degree of flexibility exhibited by immunoglobulin molecules. In the case of the best-studied isotype, IgG, it is clear that the structural feature most associated with relative motion of one subunit relative to another is the hinge, which connects the C_H1 domain to the C_H2 domain and is encoded by a separate exon(s). The human IgG3 subclass has an extended hinge region that can impart increased flexibility. In the case of IgA, the IgA1 hinge is flexible, such that the F(ab) arms can range from the typical "Y" configuration to a "T" configuration, whereas IgA2 molecules are relatively constrained.

Immunoglobulin flexibility has important functional consequences for the antibody. First, inter-F(ab) movements can play an important role in permitting antibodies to bind in monogamous bivalent (multivalent) fashion to antigenic surfaces that display repetitive epitopes. Second, efficiency in precipitation of multivalent antigen molecules or agglutination of multivalent antigen particles can be correlated with inter-F(ab) flexibility. Third, optimal interactions of effector molecules with IgG antibody Fc regions has been postulated to depend on the ability of the Fc region to bend out of the plane of the F(ab) arms (*dislocation*) (but see discussion on complement activation below).

FUNCTIONS MEDIATED BY ANTIBODY ALONE

While it is clear that in many in vivo situations antibodies mediate their effects with the aid of other molecules and, in some

cases, cells (see next section); there are circumstances where the antibody can influence antigenic targets directly, at least in vitro. The very name "antibody" implies the negation of some activity, and antibodies were first defined as factors that could inactivate or neutralize toxins. Subsequent studies have shown inactivation of viruses, parasites, and enzymes, as well.

Virus Neutralization and Immunity

KEY CONCEPTS

Immunity to Viruses

- Antibodies can neutralize (decrease the replication of) viruses by blocking attachment to the host cell, preventing penetration of the host cell membrane, or interfering with uncoating of the virus within the cell.
- Neutralizing antibodies use their V domains to bind proteins or glycoproteins on the virion surface but may depend on the interaction between their C domains and host Fc receptors to inhibit virus replication in vivo.
- Non-neutralizing antibodies that bind to virion surface proteins or glycoproteins may or may not contribute to immunity, and in some cases may actually enhance infection.
- The magnitude of neutralization mediated by a given antibody may vary with the host cell used for the measurement.
- Neutralization in vitro may not always correlate with protection in vivo.
- Non-neutralizing antibodies can contribute to protection for some viruses based on FcR-dependent mechanisms or via signal transduction.

A phenomenon of fundamental medical and biological importance is the *neutralization* of viruses by antibodies.[14] Although neutralization is defined as the elimination or reduction of the virus's ability to replicate, it does not imply a particular mechanism of interference with the process of replication. Moreover, the measurement of neutralization can depend on the choice of host cell. Thus, the neutralizing activity of a given antibody for a given virus is not an intrinsic property of the antibody but is a property of the relationship between antibody and virus under defined conditions. Consequently, neutralization titers in serum do not always correlate perfectly with protection from infection or disease in vivo.

There are several mechanisms by which antibodies can inactivate viruses. A virus infects a cell via multiple steps: (1) attachment to one or more membrane components, (2) penetration of or fusion with the membrane, (3) uncoating, and (4) genome expression. Although the most obvious mechanism of neutralization is prevention of viral attachment to the host cell surface, some antibodies can block other steps. Neutralizing antibodies for enveloped viruses, such as influenza virus, have been shown to prevent fusion between the virion and cell membranes, and neutralizing antibodies for poliovirus have been shown to interfere with viral uncoating in the host cell.

There is no one-to-one correspondence between isotypes and neutralization mechanisms; different isotypes of antibodies may employ different neutralization mechanisms to varying degrees. In some cases, IgG or IgM antibodies in the blood can mediate protection against a virus directly, or with the assistance of complement components. In mucosal secretions, where complement is less plentiful than in the blood, virus-specific IgA is more likely to utilize virus-inactivating mechanisms that do not require complement, such as prevention of attachment.

Traditional thinking maintains that antibody mediates any protective effects extracellularly. However, it has been reported that IgA antibodies, when transported by the polymeric immunoglobulin receptor (pIgR), can mediate protection against intracellular influenza virus.[15] Similar phenomena have been reported for rotavirus and HIV.

There are several other notable features of antibody–virus interactions. Not all antibodies that bind to molecules on the virion surface will neutralize the virus in all conditions. For the influenza virus hemagglutinin, binding of antibodies to some sites, but not others, will effect neutralization. Some gene products on the virion surface may fail to routinely support viral neutralization (e.g., influenza neuraminidase). However, antibody to influenza neuraminidase, while non-neutralizing, is thought to slow the spread of infection by interfering with the escape of progeny virions from infected cells. Non-neutralizing antibodies, or neutralizing antibodies at suboptimal concentrations, have been found in some instances to enhance the infection of host cells by virus (e.g., HIV-1 or dengue virus). It should be noted, however, that the clinical relevance of this enhancement, at least in regard to HIV, remains to be determined. Finally, some non-neutralizing antibodies, or those antibodies that fail to directly neutralize virus in an in vitro assay, can mediate protective effects in vivo, presumably by engaging antigen-nonspecific effector mechanisms (i.e., complement or FcR-bearing cells), or perhaps through cellular signal transduction.[16]

Neutralization of Toxins and Enzymes

KEY CONCEPTS

Immunity to Bacteria

- By neutralizing exotoxins, antibodies may prevent disease mediated by bacterial pathogens (e.g., *Corynebacterium diphtheriae* and *Clostridium tetani*).
- Antibodies can also bind and inhibit bacterial proteins that perform critical metabolic or virulence-related functions.
- Antibodies alone or with complement-derived split products can opsonize pyogenic pathogens such as *Streptococcus pneumoniae*.
- Antibodies can mediate destruction of some bacteria (e.g., *Neisseria meningitidis* and *Neisseria gonorrhoeae*) through activation of the classical pathway of complement leading to assembly of the membrane attack complex.
- Antibodies can bind to bacterial adhesins and thus interfere with pathogen attachment to mucosal epithelial cells.

In many bacterial infections, the clinical consequences of infection result from toxic molecules liberated by the bacterial cells rather than from the presence of the microorganisms themselves. Antibodies to such toxins can provide life-saving protection from disease, while not directly eliminating the bacteria producing the toxins. A classic example is infection with *Corynebacterium diphtheriae*, which secretes a potentially lethal exotoxin. A more recent example is the emergence of *Clostridium difficile*, which secretes both an enterotoxin (toxin A) and cytotoxin (toxin B). Not only is there a correlate between antibody titers to toxin A and B and prevention of relapse but also passive immunotherapy with antibody prevented relapse.[17]

Bacteria can also produce additional virulence factors, such as enzymes that facilitate pathogen spread through tissues. Host antibodies that inactivate such enzymes can be beneficial. Inactivation of toxins or enzymes is presumed to result from direct

competition between antibody and the target molecule or substrate of the toxin or enzyme, or from the stabilization or induction of conformations incompatible, to some degree, with the normal function(s) of the molecule. However, recent evidence in mice suggests that the protection afforded by exotoxin-neutralizing antibodies can depend on the presence of Fcγ receptors.[18]

FUNCTIONS MEDIATED BY ANTIBODY AND ADDITIONAL MOLECULES OR CELLS

Complement Activation

Antibody bound in vivo can activate antigen-nonspecific effector mechanisms. The exact mechanisms activated, if any, will depend on the isotype of the antibody as well as on other factors. One critical set of these effector mechanisms is encompassed by the classical pathway of complement activation (Chapter 40). Human antibody isotypes vary considerably in their intrinsic ability to activate this pathway. The consensus view is that IgM, IgG1, and IgG3 isotypes are effective activators. While some sources state that IgG2, IgG4, and IgA are weak or nonactivators of the classical complement pathway, evidence suggests that IgG2 can activate the classical pathway effectively when epitope density is high.[19] Of course, complement-fixing ability may not be determined solely by the subclass of an IgG antibody.

One obvious source for the isotype-related variation in complement-activating ability is variation in affinity for C1q (IgG3 > IgG1 > IgG2 > IgG4), the portion of the first component in the classical pathway that physically contacts the C_H2 domains of antibodies. The intrinsic affinity of the C1q globular heads for Fc regions of any isotype is relatively low, which may account in part for the observation that two or more IgG molecules in proximity are required for activation of the classical pathway beginning with C1. Thus, in the activation of the classical pathway, the functional affinity of C1q for antibody Fc regions is a crucial parameter.

IgG subclass–associated differences in some measures of complement activation have been found, under some experimental conditions, to depend on quantitative differences in steps of the cascade subsequent to the binding of C1q to antibody. Regarding the role of segmental flexibility in complement activation, there is no simple correlation between this physical property and activity in fixing the classical complement pathway.[20] Recent studies using a variety of imaging and biophysical methods demonstrate that hexameric assemblies of human IgG1, mediated by noncovalent Fc–Fc interactions, maximally activate complement.[21]

While it is generally agreed that IgA does not activate the classical pathway, its ability to activate the alternative complement pathway has been controversial. Studies with recombinant IgA molecules have suggested that neither IgA1 nor IgA2 activates either complement pathway. However, aberrantly glycosylated IgA and polymeric IgA may activate the lectin or alternative pathway, and this activation has been postulated to be associated with IgA nephropathy.

Antibody-mediated activation of the classical complement pathway has a variety of potential consequences. These include creation of additional sites for attachment to a foreign particle, thereby facilitating ingestion (opsonization), elaboration of substances that mediate leukocyte chemotaxis, additional metabolic changes in leukocytes involved in the destruction of pathogens, and changes in vascular permeability. In this process, it is the antibody that provides the specificity, whereas the other molecules function without specificity for the epitopes involved.

Receptors for Fc Regions

The other major system by which antibodies mediate effector functions is cellular. The specific molecules with which cells recognize antibodies are called FcR.[22] In humans, there are several FcR for IgG (FcγRI, FcγRIIa, FcγRIIb, FcγRIIIa, FcγRIIIb), as well as other FcR for IgA, IgE (FcεRI, FcεRII), and IgM (Fig. 8.5). We describe selected features of FcR that help to illuminate the principles by which they function.

Some receptors (FcγRI, FcεRI) have relatively high intrinsic affinities for antibody molecules and can thus bind significant fractions of monomeric Ig at physiological concentrations. For example, the high-affinity receptor for IgE (FcεRI) binds IgE with an intrinsic affinity of approximately 1×10^{10} M^{-1}. Single IgE molecules can bind to mast cells or basophils through cell-surface FcεRI prior to interacting with allergen (antigen). In contrast, FcγRII and FcγRIII have relatively weak intrinsic affinities for IgG Fc regions. Consequently, multivalent forms of IgG, such as are found in complexes of antibody and multivalent antigens (immune complexes), are much more readily bound to these FcR. Thus, for both the complement-dependent and the FcR-dependent effector function pathways, multivalency of Fc regions (functional affinity) can play a critical role.

Several types of functional consequences can follow ligation of FcR by antibody–antigen complexes. These include activation and metabolic alteration of the FcR^+ cells, phagocytosis of antibody-coated particulate antigens, antibody-dependent cellular cytotoxicity (ADCC), and release of mediators that promote inflammation. The end result of Fc binding depends not only on the receptor but also on the cell on which it is expressed and on concurrent stimulation, if any, of additional receptor types on that cell. As an example, the most studied FcR are those that bind IgG, and these receptors are expressed on many hematopoietic and even non-hematopoietic cells. Within the three classes of receptors (I, II, and III), the latter two FcR exist in two isoforms (A and B). Of interest to the regulation of the immune response, the B isoform for FcγRII transmits an inhibitory signal while the A isoform transmits an activating signal (Chapter 4).

CD89 has been identified in humans as a receptor for IgA and it is expressed on myeloid cells including polymorphonuclear neutrophils (PMNs), monocytes, and a population of dendritic cells.[23] Signaling through CD89 involves an ancillary chain that transmits an activating signal. However, not all CD89 molecules associate with this chain, in which case bound IgA is endocytosed and recycled back to the surface of the cell.[23] Interestingly, Fc binding to CD89 may be more potent at mediating ADCC than Fc binding to one of the FcγR.

Recent data suggest another possible function that depends on the interaction between antibody (IgA) and a cell-surface receptor able to bind to pIgR. Transport of IgA–antigen complexes across epithelial surfaces by pIgR may represent a form of antibody-facilitated antigen excretion.[24]

ANTIBODIES AS SURROGATE LIGANDS

The notion that one molecule can mimic a second molecule, in one respect or another, is of extraordinarily broad applicability and profound biological significance. At least three types of mimicry can be distinguished, and each type is best regarded as a continuous variable[24]: (1) limited structural mimicry of

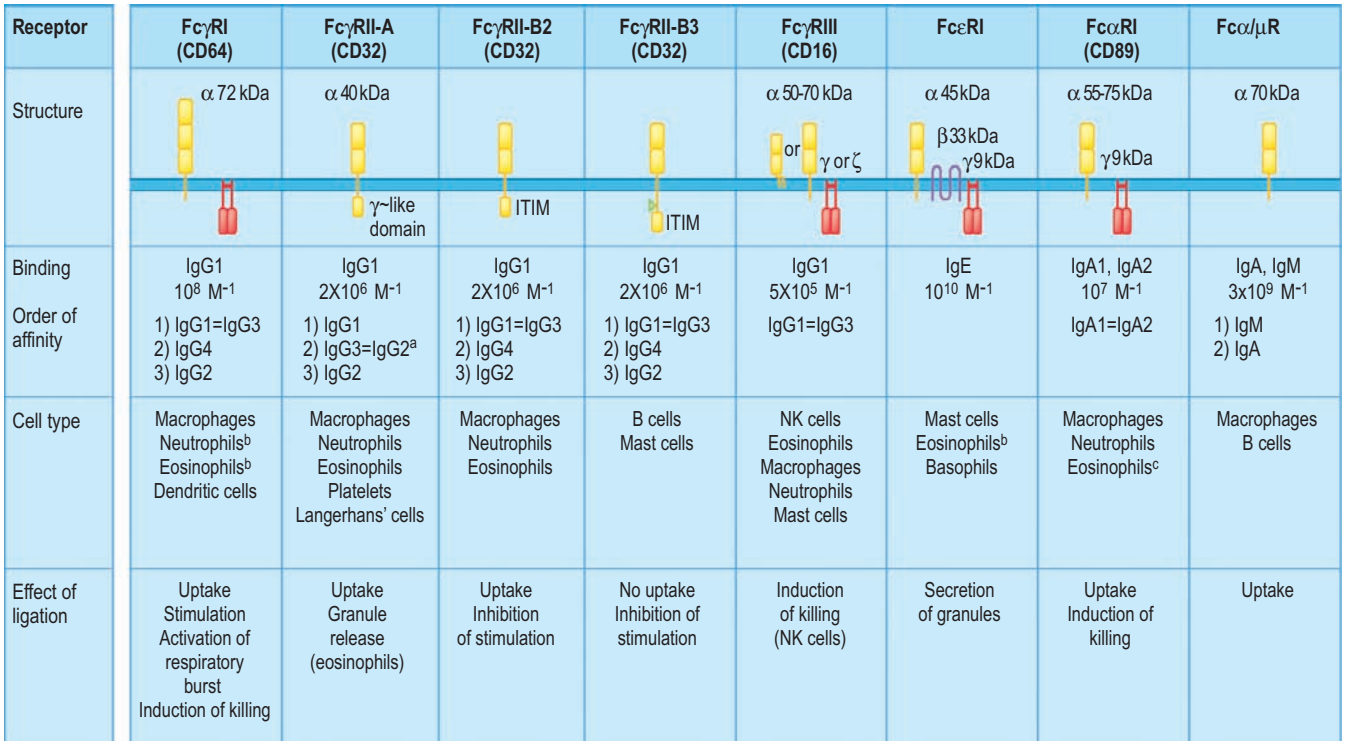


FIG. 8.5 Domain Structures, Binding Properties, Cellular Expression Patterns, and Functional Effects of Human Fc Receptors. A given Fc receptor (FcR) may exhibit differences in composition depending on the cell type expressing it. For example, FcγRIII is expressed on neutrophil plasma membranes bearing a glycosylphosphatidylinositol anchor, without FcR γ chains, while it is expressed on NK-cell plasma membranes as a conventional transmembrane protein in association with FcR γ chains. Similarly, FcγRII-B1 contains an additional stretch of polypeptide encoded by an exon whose product is not represented in the intracellular domain of FcγRIIB2. This additional portion of the polypeptide is believed to prevent the internalization of FcγRII-B1 subsequent to cross-linking. [a]A subset of FcγRII-A allotypes bind to human IgG2. [b]For these cells, FcR expression is inducible, not constitutive. [c]The molecular weight of CD89α chain is 70 to 100 kDa in eosinophils. *ITIM*, immunoreceptor tyrosine-based inhibitory motif; *NK*, natural killer cell. (With permission from Janeway CA Jr, Travers P, Walport M, Shlomchik M. *Immunobiology: The Immune System in Health and Disease*. 6th ed. New York: Garland Science; 2004.)

one molecule by another; (2) mimicry at the level of noncovalent interaction, that is, whether the model (object of mimicry) and the mimic bind the same receptor sites and with the same affinities; and (3) mimicry of more complex biological functions, such as cellular or enzymatic inactivation. It is important to make these distinctions because the extent of mimicry of one type is not a perfect predictor of the extent of mimicry of another. We have already noted that slight changes in structure sometimes have slight effects on binding affinity or specificity, yet in other cases they have dramatic effects on binding affinity or specificity. Thus, structural similarity (mimicry), as we perceive it, is not perfectly correlated with the extent of binding or the elicitation of higher biological function.

There are two aspects of receptor–ligand interaction that antibodies can potentially mimic. First, the inducibility of a vast repertoire of antibody specificities suggests the potential for identifying antibodies that can bind any given target molecule at (near) a given site. Thus, there should be a reasonable probability of obtaining antibodies that bind to a particular receptor at a site bound by some other, perhaps physiological, ligand or co-receptor. Evidence that antibodies can mimic the functional effects of other molecules is provided by many investigations of anti-idiotypic antibodies and conventional anti-receptor antibodies.[25]

Second, the triggering event for many cellular and effector processes in the immune system is the aggregation of receptor molecules by clustered ligands. Therefore, the ability of antibodies, which naturally have a maximal valence of two or greater, to cross-link cell-surface molecules and initiate signal transduction contributes to the abilities of antibodies to serve as surrogate co-receptors for cell-surface molecules. This has greatly facilitated the identification and functional characterization of many of these molecules and is also being exploited for therapeutic uses.[26]

FUNCTIONAL PROPERTIES OF ENGINEERED ANTIBODY MOLECULES

Monoclonal Antibodies

Many modern applications of antibodies in research, medicine, veterinary medicine, and other fields rely heavily on monoclonal antibody technology. By definition, a monoclonal antibody preparation is derived from a clonal population of B-lineage cells. All of the antibodies express identical V domains with identical antigen specificities. It is the homogeneity of the monoclonal V domain structures that most crucially distinguishes a monoclonal from a polyclonal, serum-derived antibody preparation. Both sets of antibodies bind the same antigen, but only the monoclonal preparation will bind the same epitope in the same way. Thus, homogeneous antibodies give more reproducible, and more easily interpreted, results for many kinds of assays.

ON THE HORIZON

Monoclonal Antibodies: Recent Advances

- Production of human monoclonal antibodies from patients who have recovered from infectious diseases could provide new therapeutics for the causative pathogens.
- Genetic engineering of human B cells that make human antibodies able to bind pathogens, pathogen-derived molecules, or other antigens will potentially provide a new pathway for developing cellular therapy for infectious and other diseases.[27]
- If key obstacles can be overcome, vector-mediated expression of antibodies (vectored immunoprophylaxis) will become an option for providing protection against infectious diseases in patients who are not likely to respond effectively to active immunization with standard vaccines due to varying types of immunodeficiency.[28]
- Advances in manipulating glycan structures on antibodies could enhance the efficacy of therapeutic monoclonal antibodies.[29]
- Improved ability to determine the precise geometry of the interaction of antibody with antigen can enable prediction and engineering of mechanisms of therapeutic effect.[30]

Monoclonal antibodies of selected specificity were first produced by cells referred to as *hybridomas*,[31] which are hybrid transformed cells that are created by the fusion of two types of cells. One parent of a hybridoma (the fusion partner) is typically a transformed cell, usually a myeloma cell line, which contributes a metabolism that supports unlimited growth in tissue culture and high rates of immunoglobulin synthesis and secretion, no longer synthesizes an immunoglobulin molecule, and can be selected against in special culture media. The second parental cell is a B lymphocyte that provides the genetic information for the production of a particular antibody. The choice of specificity on the part of the investigator is influenced by both the choice of immunogen and the nature of the screening assay, which permits identification of the minority of cells that secrete a monoclonal antibody of desired specificity, or in some cases, function.

Monoclonal antibodies are useful for the identification and quantitation of diverse molecules of biological or synthetic origin, including human immunoglobulins (e.g., paraproteins), antigens from infectious agents (e.g., HIV p24), hormones, drugs, and toxins. They have also been exploited for therapeutic purposes, such as reversing allograft rejection, killing tumor cells, or preventing cytokine activity contributing to autoimmune disease.

Recombinant Antibodies

ON THE HORIZON

Recombinant Antibodies—Antibody Engineering

- Enhance effector function potency and half-life through Fc modification.
- Reduce or eliminate bystander effects.
- Isolate and express antibody genes using single B-cell cloning or combinatorial libraries.
- Allow expansion of the antibody repertoire and targeting possibilities.
- Facilitate agents for molecular recognition of minimal size (i.e., nanobodies).
- Develop drugs or toxins that can be covalently attached to novel antibodies for therapy.
- Create multispecific or multimeric constructs.
- Guide vaccine design.
- Evaluate the roles of isotypes in protection or other functions.
- Define how differences in glycosylation can alter effector function.

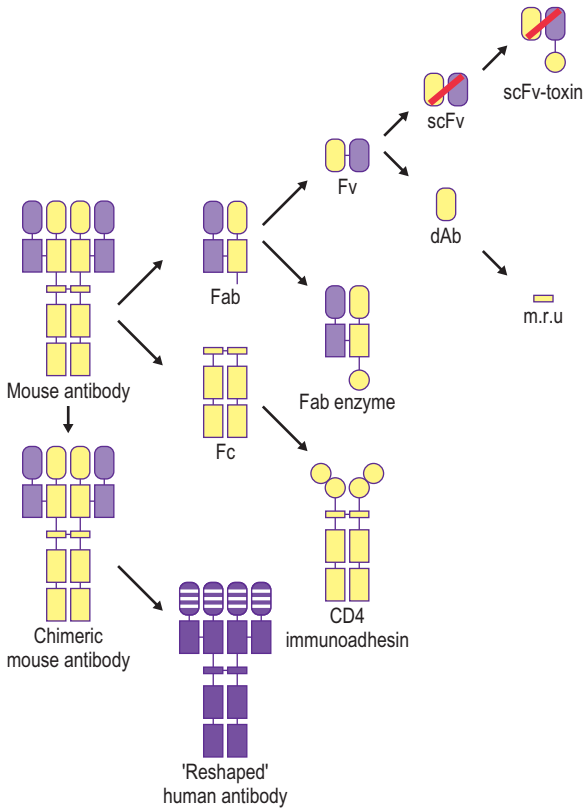


FIG. 8.6 Examples of Engineered Antibodies and Antibody-Derived Fragments That Can Be Created Through the Manipulation of Antibody Genes. Each *closed rectangular (constant)* or *rounded (variable) box* represents a domain. The molecule at the bottom of the figure represents a humanized antibody, where the constant domains and variable domain framework regions correspond to human amino acid sequences. Only the hypervariable regions, and in some cases a small number of framework residues, correspond to mouse or rat antibody amino acid sequences. Other structures depicted include an *Fab* fragment; an *Fv* fragment; a single-chain Fv fragment *(scFv)* in which the C-terminus of the V_H domain is linked covalently by a linker peptide to the N-terminus of the V_L domain; an Fab-enzyme fusion protein; an scFv-toxin fusion protein; an immunoadhesin in which extracellular domains from CD4 have been covalently attached to human heavy-chain constant domains; a single V_H domain *(dAb)*; and a peptide derived from a hypervariable region (minimal recognition unit, or *m.r.u.*). (With permission from Winter G, Milstein C. Man-made antibodies. *Nature*. 1991:349:293.)

The ability to manipulate the genes that encode antibodies, and thereby manipulate the structures of antibodies, has opened a new era in the study and application of antibodies (Fig. 8.6).

Progress includes expression of recombinant intact IgG molecules,[32] expression of Ig fragments [F(ab), Fv] in eukaryotic and prokaryotic host cells, proteomic mining of combinatorial libraries of antibody fragments displayed on the surfaces of filamentous phage[33] or yeast and bispecific or multispecific antibodies. Most recently, the ability to clone immunoglobulin genes from single B cells has revolutionized the production of human monoclonal antibodies. This technology has made possible the identification of broadly neutralizing antibodies for

rapidly evolving pathogens, such as HIV-1 and influenza A viruses. These antibodies may prove to be clinically useful.[34]

Recombinant antibodies are also being designed to improve distribution and half-life of administered antibodies. Antibody engineering is likely to contribute to the design of novel therapeutic agents. And use of these agents is likely to continue to yield new fundamental information. For example, advances in understanding the role of specific effector functions and the influence of the geometry of antibody–antigen interaction in tumor cell destruction have come from using mutant and engineered recombinant antibodies such as rituximab (anti-CD20).

REFERENCES

1. Amit AG, Mariuzza RA, Phillips SEV, et al. Three-dimensional structure of an antigen-antibody complex at 2.8 Å resolution. *Science*. 1986;233:747–753.
2. Greenspan NS. Dimensions of antigen recognition and levels of immunological specificity. *Adv Cancer Res*. 2001;80:147–187.
3. Steward MW. The biological significance of antibody affinity. *Immunol Today*. 1981;2:134–139.
4. Rini JM, Schulze-Gahmen U, Wilson IA. Structural evidence for induced fit as a mechanism for antibody–antigen recognition. *Science*. 1992;255:959–965.
5. Fereidan-Esfahani M, Nayfeh T, Warrington A, et al. IgM natural autoantibodies in physiology and the treatment of disease. *Methods Mol Biol*. 2019;1904:53–81.
6. Kramer A, Keitel T, Winkler K, et al. Molecular basis for the binding promiscuity of an anti-p24 (HIV-1) monoclonal antibody. *Cell*. 1997;91:799–809.
7. James LC, Roversi P, Tawfik DS. Antibody multispecificity mediated by conformational diversity. *Science*. 2003;299:1362–1367.
8. Kwong P, Doyle M, Casper D, et al. HIV-1 evades antibody-mediated neutralization through conformational masking of receptor-binding sites. *Nature*. 2002;420:678–682.
9. Mollnes TE, Harboe M. Neoepitope expression during complement activation—a model for detecting antigenic changes in proteins and activation of cascades. *Immunologist*. 1993;1:43–49.
10. Berzofsky JA. Intrinsic and extrinsic factors in protein antigenic structure. *Science*. 1985;229:932–940.
11. Horiya S, MacPherson IS, Kruas IJ. Recent strategies targeting HIV glycans in vaccine design. *Nat Chem Biol*. 2014;10:990–999.
12. Mannik M. Physicochemical and functional relationships of immune complexes. *J Immunol*. 1980;74:333–338.
13. Heyman B. The immune complex: possible ways of regulating the antibody response. *Immunol Today*. 1990;11:310–313.
14. Dimmock NJ. Neutralization of animal viruses. *Curr Top Microbiol Immunol*. 1993;183:1–146.
15. Mazanec MB, Kaetzel CS, Lamm ME, et al. Intracellular neutralization of virus by immunoglobulin A antibodies. *Proc Natl Acad Sci U S A*. 1992;89:6901–6905.
16. Binder GK, Griffin DE. Immune-mediated clearance of virus from the central nervous system. *Microbes Infect*. 2003;5(5):439–448.
17. Lowy I, Molrine DC, Leav BA, et al. Treatment with monoclonal antibodies against *Clostridium difficile* toxins. *N Engl J Med*. 2010;362:197–205.
18. Abboud N, Chow SK, Saylor C, et al. A requirement for FcgR in antibody-mediated bacterial toxin neutralization. *J Exp Med*. 2010;207(11):2395–2405.
19. Garred P, Michaelsen TE, Aase A. The IgG subclass pattern of complement activation depends on epitope density and antibody and complement concentration. *Scand J Immunol*. 1989;30:379–382.
20. Tan LK, Shopes RJ, Oi VT, et al. Influence of the hinge region on complement activation, C1q binding, and segmental flexibility in chimeric human immunoglobulins. *Proc Natl Acad Sci U S A*. 1990;87:162–166.
21. Diebolder CA, Beurskens FJ, de Jong RN, et al. Complement is activated by IgG hexamers assembled at the cell surface. *Science*. 2014;343(6176):1260–1263.
22. Nimmerjahn F, Ravetch JV. Fcgamma receptors as regulators of immune responses. *Nat Rev Immunol*. 2008;8(1):34–47.
23. Monteiro RC. Role of IgA and IgA Fc receptors in inflammation. *J Clin Immunol*. 2010;30(1):1–9.
24. Kaetzel CS, Robinson JK, Chintalacharuvu KR, et al. The polymeric immunoglobulin receptor (secretory component) mediates transport of immune complexes across epithelial cells: a local defense function for IgA. *Proc Natl Acad Sci U S A*. 1991;88:8796–8800.
25. Greenspan NS. Relections on internal images. *Nat Biotechnol*. 1997;15:123–124.
26. Cragg MS, French RR, Glennie MJ. Signaling antibodies in cancer therapy. *Curr Opin Immunol*. 1999;11:541–547.
27. Hartweger H, McGuire AT, Horning M, et al. HIV-specific humoral immune responses by CRISPR/Cas9-edited B cells. *J Exp Med*. 2019;216(6):1301–1310.
28. Sanders JW, Ponzio TA. Vectored immunoprophylaxis: an emerging adjunct to traditional vaccination. *Trop Dis Travel Med Vaccines*. 2017;3:3.
29. Wang M, Wang Y, Liu K, et al. Engineering a bacterial sialyltransferase for di-sialylation of a therapeutic antibody. *Org Biomol Chem*. 2020;18(15):2886–2892.
30. Rouge L, Chiang N, Steffek M, et al. Structure of CD20 in complex with the therapeutic monoclonal antibody rituximab. *Science*. 2020;367(6483):1224–1230.
31. Kohler G, Milstein C. Continuous cultures of fused cells secreting antibody of predefined specificity. *Nature*. 1975;256:495–497.
32. Morrison SL. In vitro antibodies: strategies for production and application. *Annu Rev Immunol*. 1992;10:239–265.
33. Hudson PJ. Recombinant antibody constructs in cancer therapy. *Curr Opin Immunol*. 1999;11:548–557.
34. Corti D, Lanzavecchia A. Broadly neutralizing antiviral antibodies. *Annu Rev Immunol*. 2013;31:705–742.

9

T-Cell Development

S. Thera Lee and Kristin Hogquist

T-CELL DEVELOPMENT

For an effective immune response, antigens from disease-causing agents such as bacteria and viruses are recognized by the adaptive immune system. Two major components of this arm of the immune response are B and T cells, both of which have antigen receptors that respond to a vast range of antigens. These lymphocyte populations, however, develop in different tissues: for T cells, the primary lymphoid organ is the thymus. Like B cells, T cells develop from hematopoietic stem cells (HSCs) in the bone marrow, and during fetal life, in the liver. Lymphoid progenitors access circulation and migrate into the thymus. There, these cells, called thymocytes, differentiate into mature T cells. This process requires passage through a series of developmental checkpoints designed to test the affinity of the thymocyte's T-cell receptor (TCR) to antigens expressed on the surface of thymic antigen-presenting cells (APCs) in the context of a major histocompatibility complex (MHC) (Chapter 6). This process generates a mature repertoire of T cells that is functional—able to protect the organism from pathogens it may encounter, but also self-tolerant.

In the laboratory, this process can be manipulated by exchanging a native T-cell receptor with a synthetic chimeric antigen receptor (CAR).[1] Allogeneic or autologous T cells can thus be genetically modified to express a CAR that may combine an extracellular binding domain, often an antibody-derived single chain variable fragment (scFv), with activating signaling domains from the T-cell-receptor complex, such as CD3ζ, CD28, and 4-1BB (Chapter 4). By so doing, one can take a cytotoxic T-cell clone that has been educated in the thymus to be self-tolerant, and redirect its activity to a protein on the surface of a cancer cell. In this way, one can provide a patient with a personalized immunotherapy.

THYMUS: THE SITE OF T-CELL DEVELOPMENT

In all species with T cells, development occurs in the thymus. If the thymus is surgically removed (thymectomized) early in life or if patients are born with mutations that impact the development of the thymus, there is striking immunodeficiency, leading to increased susceptibility to infection (Chapter 34).[2] Anatomically, the thymus is located behind the sternum. Its distinct lobes are divided into the outer cortex and inner medulla (Fig. 9.1). Although the thymus is primarily composed of developing thymocytes, thymic stromal cells and hematopoietic antigen-presenting cells (Chapter 6) are also present. In the cortex and medulla, the stromal cells and antigen-presenting cells create unique microenvironments. Compartmentalization of thymocytes into these microenvironments provides for the distinct cues that are needed to support T-cell development through the progressive developmental stages.

Initially, thymic precursors enter into the thymic parenchyma near the corticomedullary junction (CMJ). As they develop, immature thymocytes migrate into and through the densely packed cortex, where they are guided by cues from surrounding cells. These cells include large, branched cortical thymic epithelial cells (cTECs), dendritic cells (DCs), and macrophages. Here, thymocytes are committed to the T-cell lineage and progress through sequential differentiation steps. The thymic environment provides chemokines that later guide the developing thymocytes to leave the cortex and enter the thymic medulla.[3] The thymic medulla has distinct thymic epithelial cells, aptly named medullary thymic epithelial cells (mTECs), as well as other antigen-presenting cells. More mature thymocytes reside in the medullary region where αβ thymocytes become MHC class I-restricted CD8 thymocytes or MHC class II-restricted CD4 thymocytes. The thymocytes that survive this process then emigrate from the thymus into the periphery as mature T cells.

Thymic stromal cells, including cTECs and mTECs, form the supporting meshwork needed for thymic structure and proper thymocyte development.[4] Foxn1 is a transcription factor essential for TEC development, and *Foxn1* deficiency leads to a thymic rudiment (athymia) that cannot sustain lymphopoiesis. Defects in *Foxn1* have been described in both mice and humans. The resulting phenotype includes hair loss and T-cell immunodeficiency. In humans, this disorder is referred to as the Pignata Guarino syndrome, and in mice, the *nude* mutation. Functional T cells develop when bone marrow from *nude* mice is transplanted into mice with normal stromal cells, but not when bone marrow from normal mice is transplanted into mice with the *nude* mutation. Thus, the study of *nude* mice has shed particular light on a specific need for thymic stromal cells in normal T-cell development. Patients with DiGeorge syndrome (Chapter 34) exhibit a similar phenotype of athymia and T-cell deficiency. However, these patients have deletions in chromosome 22q11, which encodes *Tbx1*. These experiments of nature have shown that the thymus is the major site of T-cell maturation and that this organ is necessary for the development and production of functional peripheral T cells.

EARLY T-CELL DEVELOPMENT

Early Thymic Progenitors Encounter Notch Ligands

The thymus does not contain a pool of resident pluripotent stem cells. Instead, progenitor cells from the bone marrow enter the blood and then seed the thymus continuously

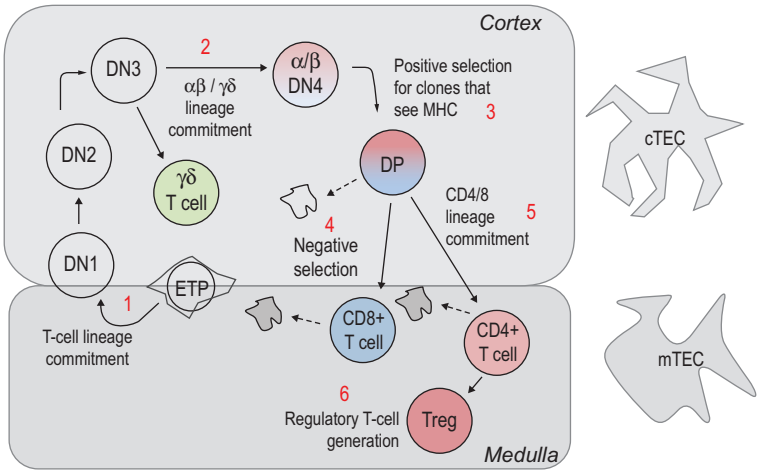


FIG. 9.1 Checkpoints in T-Cell Development Occur in Distinct Locations in the Thymus. The thymus is composed of several lobules, each with outer cortical and inner medullary regions. Thymic antigen-presenting cells localized within a specific anatomical location, such as cortical and medullary thymic epithelial cells *(cTEC and mTEC) (right)*, allow for developing thymocytes to pass through six major checkpoints through the thymus (see Table 9.1). The earliest thymic precursors enter via blood vessels at the cortico-medullary junction and commit to the T-cell lineage (1) before migrating to the outer cortex as immature $CD4^-$ $CD8^-$ double negative *(DN)* cells. As these thymocytes mature, they begin to rearrange their TCRs and diverge into αβ or γδ T cells (2). Cells destined to be αβ T cells upregulate CD4 and CD8 to become DP thymocytes that migrate back through the cortex. At this point, they express T-cell receptors on their surface that engage ligands in the thymus, allowing for T-cell positive (3) and negative (4) selection to occur. Surviving cells will downregulate CD4 or CD8 and commit to either the cytotoxic CD8 or helper CD4 lineage (5) while migrating into the medulla. The medulla contains only mature single positive (either CD4 or CD8) thymocytes, which will continue to undergo negative selection. A fraction of the single positive CD4 cells will become regulatory T cells (6). All fully mature T cells then emigrate to the periphery via blood vessels. *DP*, Double-positive; *ETP*, early thymic progenitors; *Treg*, T regulatory cell.

TABLE 9.1 Critical Checkpoints and Factors in T Cell

	Checkpoint	Role	Critical Factors
1	T-lineage commitment	Progenitors lose potential to generate other cell types	**Notch** signaling
2	γδ/αβ lineage commitment	Progenitor becomes either a γδ or αβ T cell	Depends on which receptor is assembled first
3	Positive selection	Selects for clones that can bind antigens in the context of that individual's MHC alleles	MHC/self-peptide in cTEC full TCR signaling capability
4	CD4/8 lineage commitment	Ensures MHC class specificity is linked to helper or killer function	**Runx** for cytotoxic CD8 T cells, **ThPOK** for helper CD4 T cells
5	Negative selection (clonal deletion)	Ensures "self-tolerance" to prevent autoimmunity	**AIRE** in mTEC **CD80/86+** APCs
6	Regulatory T-cell generation	Ensures "self-tolerance" to prevent autoimmunity	**FoxP3** in Treg
7	Selection of other rare specialized cells	IEL cells NKT cells MAIT cells M3 cells	**PLZF** in NKT cells
8	Emigration	Allows fully mature, selected T cells to leave the thymus and seed peripheral tissues	**KLF2** in all mature T-cell types

AIRE, Autoimmune regulator protein; *APCs*, antigen-presenting cells; *cTEC*, cortical thymic epithelial cell; *IEL*, intraepithelial lymphocyte; *MAIT*, mucosal-associated invariant T; *MHC*, major histocompatibility complex; *mTEC*, medullary thymic epithelial cell; *NKT*, natural killer T; *TCR*, T-cell receptor; *Treg*, T regulatory cell; *APC*, antigen-presenting cell.

throughout life.[3,5] HSC in the bone marrow can give rise to all lineages of blood cells. As stepwise developmental programs progress, downstream progenitors lose the potential to enter non-T lineages.[5,6] T-cell progenitors immigrate into the thymus at the cortico-medullary junction (see Fig. 9.1, step 1). This step involves the chemokine receptors CCR7 and CCR9, and the selectin PSGL-1. At this point, the cell is considered a multipotent early thymic progenitor (ETP).

ETPs experience Notch signaling in the thymus, which is essential for their commitment to the T-cell line (Table 9.1).[5–8] Notch is a heterodimeric receptor that binds to the Delta family and the Serrate family of ligands. In particular, Delta-like 4 is expressed by TEC and is required for the specification of ETP into the T-cell lineage pathway, with a concomitant repression of the B-cell lineage. In the absence of Notch, the thymus fills with B cells. Conversely, if Notch activation is forced in the bone marrow, T-cell development will initiate there. Thus, Notch ligands are an important part of what defines the thymus as a special environment for T-cell development. In the clinic, activating mutations in Notch are a common cause of human T-cell acute lymphoblastic leukemia (T-ALL).

Other Key Transcriptional Regulators

Notch signaling activates several additional transcription factors that are important for T-cell lineage fate and later αβ TCR selection. These include GATA3 (*Gata3*), TCF-1 (*Tcf7*), and

HES1 (*Hes1*). GATA3 is a zinc-finger transcription factor that is expressed as early as the ETP stage. Expression increases until the thymocytes reach the DN3 stage of T-cell development. TCF-1 and HES1 are also highly expressed by the ETP population. HES1 has been shown to be important for restraining the myeloid cell developmental pathway, as ETP cells maintain the potential to develop into natural killer (NK) and myeloid cell populations. Activation of Bcl11b, a transcriptional repressor that is also highly upregulated from the early ETP stage onward, restrains NK cell development.

Fate Commitment Occurs at the Double-Negative Stage

ETP undergo 1000-fold expansion over a period of 10 days.[3] Developing thymocytes eventually go through a series of well-defined stages marked by changes in the co-receptors CD4 and CD8 (see Fig. 9.1). Because ETP and their immature thymocyte progeny express neither CD4 nor CD8, they are called double-negative (DN) thymocytes. As DN thymocytes proceed through developmental checkpoints, they migrate outward from the CMJ into the cortical subcapsular zone, transitioning through so-called DN1, DN2, DN3, and DN4 stages in sequential order. These different stages are marked by differential surface expression of other cell-surface markers: CD44 and CD25.

In mice, DN1 cells express CD44 but not CD25 ($CD44^+CD25^-$). They can give rise to either αβ or γδ T cells but retain the potential to differentiate into NK cells and myeloid cell populations.[5,6] In human, DN cells are distinguished by other cell-surface molecules: $CD34^+CD1a^-$.[7] As DN1 progress, they begin to express CD25. These DN2 cells (mice: $CD44^+CD25^+$ and humans: $CD34^+CD1a^-$) migrate into the thymic cortex. As they mature, they lose their NK and myeloid potential.

T-CELL RECEPTOR GENE REARRANGEMENT BEGINS AT DN2

During the DN2 to DN3 transition, thymocytes are firmly committed to the T-cell lineage. However, they still have the potential to be αβ or γδ T cells. During this stage, the TCR β-chain, γ-chain, and δ-chain loci begin to rearrange (Chapter 4). The TCR α/δ, β, and γ gene loci are on three different chromosomes. The TCR δ-chain locus is embedded within the TCR α locus, which ensures that cells committed to the αβ T-cell lineage cannot express γδ TCR as well—the δ-chain genes are excised in the process of α-chain gene rearrangement as an extrachromosomal circle. The β, δ, and γ (but not α) loci adopt an open conformation at this stage, which allows recombination to initiate simultaneously. Commitment to the αβ or the γδ T-cell fates is dependent on which receptor is expressed first. This race occurs at the DN3 stage (see Fig. 9.1, step 2), where cells are $CD44^-CD25^+$ in mice and $CD34^+CD1a^+$ in humans.

γδ T Cells Diverge at DN3

If a developing thymocyte has productive, in-frame rearrangements at both the TCR γ and δ loci before a productive β-chain rearrangement, the cell is fated to become a γδ T cell. Because of the necessity for successful rearrangements at two loci rather than one, the thymus produces fewer γδ T cells than αβ T cells. Furthermore, the selection requirements for γδ T cells are different from αβ T cells and generally less well understood. After they leave the thymus, γδ T cells are disproportionally represented in nonlymphoid tissues, particularly epithelial tissues such as the skin, lung, and intestine.[9] In addition to T-cell commitment at the ETP stage, Notch signaling regulates the αβ versus γδ cell fate. αβ T cells have a higher requirement for Notch signaling, whereas γδ T cells are less sensitive. ID3, a negative regulator of E2A, plays a key role in integrating Notch and TCR signals to determine lineage commitment.[9] Notch signaling is required for β-selection, and γδ T cells induce higher levels of ID3 compared to β-selected thymocytes.

Beta Selection Generates αβ T Cells

If a productive rearrangement at the TCR β locus occurs at the DN3 stage, the β chain pairs with a nonrearranging pre-T α chain (also known as a surrogate α chain). This β-preTα heterodimer is assembled with CD3 molecules to form a productive pre-TCR complex that is transported to the cell surface. Signaling from this receptor is called β selection, and represents a key checkpoint in committing cells to the αβ lineage (see Fig. 9.1, step 2). At this point, the cell proceeds to downregulate CD25 to become DN4 cells ($CD44^-CD25^-$), where gene rearrangement is halted and rapid proliferation is induced. DN4 is a transient stage, and these cells rapidly begin to upregulate CD4 and CD8 on their cell surface, becoming positive for both (e.g., double-positive or DP) as they proliferate. These DP thymocytes make up the majority of thymocytes (approximately 80%). During this time, there is a second wave of *Rag* gene expression, which allows rearrangements at the TCR α locus. Productive rearrangements at the TCR α locus lead to expression of αβ TCR on the cell surface. The TCR can then recognize self-peptide:MHC (Chapters 5 and 6) to undergo further selection processes.

KEY CONCEPTS

The Thymus is the Anatomic Site Where T Cells Develop

- T-cell development occurs in and requires the thymus.
  - Patients without a thymus due to *FOXN1* mutations or DiGeorge syndrome lack circulating T cells.
- T cells are the progeny of HSC that circulate through blood and seed the thymus.
- ETPs are the first cells to seed the thymus, and their development into T cells requires Notch signals.
- Cells progress through the DN, then DP, stages as they recombine the TCR genes in order to express a surface T-cell receptor.
- Both αβ and γδ T cells are produced in the thymus, and their distinction occurs at the DN3 stage.

POSITIVE AND NEGATIVE SELECTION

The rearrangement and pairing of αβ TCR genes is random, and the TCRs on DP thymocytes have specificities that react to a wide range of antigens (both self or foreign) or don't react at all. Additionally, some receptors, while potentially capable of binding to a peptide antigen, may not bind to the peptide when presented by that individual's MHC molecules. For any individual, a useful T cell is one that can recognize an antigen presented by a self-MHC molecule. Thus, a positive selection step is needed to enrich the T-cell repertoire "with MHC-restricted" TCRs that are specific for antigens only in the context of the host's particular set of MHC molecules (Chapters 5 and 6).[10]

TCRs that bind strongly to self-peptides also present a risk, as these T cells may trigger autoimmunity. Therefore, a negative selection step is required to prune autoreactive clones from the T-cell repertoire. Together, these selection checkpoints permit the production of a diverse repertoire of mature T cells that are MHC-restricted, self-tolerant, and can be exported from the thymus into the periphery.

Positive Selection Generates an "MHC-Restricted" T-Cell Pool

In the thymic cortex, DP thymocytes audition for selection.[10] These immature cells only have a life span of 3 to 4 days and, without engagement of the TCR, their default pathway is one of apoptosis—a form of programmed cell death called "death by neglect" (Chapter 17). It is estimated at anywhere between 85% to over 90% of these precursors are unable to be selected and are eliminated in this manner.[10,11] Cells that avoid this fate undergo positive selection, in which the αβ TCR binds with low to intermediate avidity to self-peptide:MHC complexes presented by cTEC, promoting cell survival (see Fig. 9.1, step 3). As the self-peptides presented in the thymus are displayed on the host's own MHC molecules, positive selection ensures that only self-MHC-restricted DP thymocytes mature into CD4 or CD8 single positive (SP) thymocytes. After undergoing positive selection in the cortex, DP thymocytes up-regulate chemokine receptors such as CCR7 and migrate towards the medulla.[3] As the thymocyte crawls through the CMJ into the medullary region, it interacts with other thymic APCs, which can drive negative selection.

Negative Selection Generates a "Self-Tolerant" T-Cell Pool

Among the TCR expressed by DP thymocytes, some bind strongly to self-peptide:MHC. During the selection processes in the thymus, these autoreactive T-cell clones are removed from the repertoire in a process called *negative selection* (see Fig. 9.1, step 4). Apoptosis by these cells "deletes" them from the T-cell repertoire. The elimination of autoreactive TCRs via cell death is also known as clonal deletion. While the majority of negative selection occurs in the cortex at the DP stage, there is also a second wave of clonal deletion in the more mature thymocytes that are located in the medulla.[10,12] The deletion of self-reactive T-cell clones plays a key role in the establishment of central tolerance. This process is mediated by the different APCs found throughout the thymus (Chapter 6). While self-antigens presented by cTECs promote positive selection, negative selection can be mediated by mTEC as well as hematopoietic APCs such as DC and B cells.[13]

Role of Peptide:MHC in Negative and Positive Selection

The affinity model of selection holds that the strength of signal that results from the interaction between the self-peptide:MHC and TCR underlies positive and negative selection (Fig. 9.2).[10] According to this model, there is a spectrum of strength of signal across the large array of developing thymocytes. At one end of the spectrum are T-cell receptors that do not bind, and thus do not respond, to self-peptide:MHC. These cells will automatically undergo cell death after a few days—which is termed *death by neglect*. At the other end of the spectrum are cells that express T-cell receptors that bind with high affinity and generate a strong signal. These cells undergo cell death by clonal deletion. In the middle is a window of positively selected conventional T cells with receptors that bind to self-peptide:MHC, but with a weaker, more acceptable, affinity.

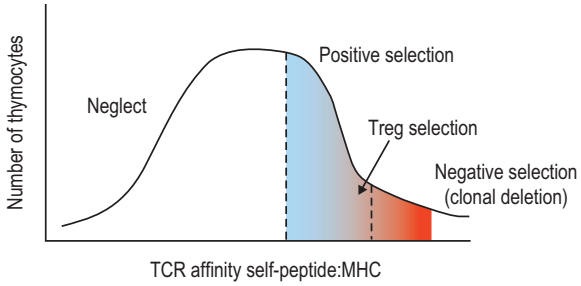


FIG. 9.2 The Fate of Thymocytes is Determined by How Strongly They Bind to Major Histocompatibility Complex (MHC)/Self-Peptide. In the affinity model of selection, the strength of the interaction between the self-peptide:MHC and T-cell receptor (*TCR*) determines whether cells are selected or undergo apoptosis. Cells with TCR that cannot form a stable interaction with self-peptide:MHC will either undergo another round of TCR gene rearrangement in a process termed *receptor revision* or *receptor editing*, thereby expressing another TCR, or they will die by neglect after a few days. T cells with TCRs that interact weakly with self-peptide:MHC will avoid death by neglect and undergo positive selection. These cells can potentially become conventional CD4 or CD8 T cells. T cells with TCR that interact strongly, and thus are highly self-reactive, will undergo negative selection by clonal deletion. Some autoreactive cells, particularly those that recognize the very rare self-peptide:MHC, are able to avoid apoptosis and instead become regulatory T cells, although this is not thought to have a strict affinity threshold. *Treg*, T regulatory cell.

Under this model, thymocytes that have high-affinity interactions with self-peptides will generally undergo apoptosis and be deleted. However, some cell types, such as regulatory T cells and other specialized lymphocytes, are able to avoid this cell fate in spite of being "more" self-reactive (see below). These agonist selected cells depend on other molecular factors, such as CD1d and costimulatory molecules, for their survival. Although the choice of survival or death appears binary, there is some stochastic overlap between agonist selected cells and clonal deletion.[14] This enables the survival of selected T cells with some potentially hazardous T cells.

The different epithelial cells in the cortex and medulla of the thymus are specialized for positive and negative selection. cTECs are essential for positive selection. The crucial role that cTECs have in selection is mediated in part by their peptide processing machinery, which gives them the ability to present a largely unique peptide:MHC repertoire (also known as the *peptidome*).

In APC, peptide fragments are loaded onto MHC molecules after cytosolic proteins are degraded by proteasomes. These proteasomes have a catalytic core that consists of three subunits: β1, β2, and β5. cTECs have the unique proteasome subunit β5t, while other APCs can only express β5 or β5i. This endows cTECs with the ability to present specialized peptides produced by the thymoproteosome (specifically composed of subunits β1i, β2i, and β5t), allowing for a peptide:MHC peptidome that promotes positive selection and shapes the T-cell repertoire.

mTECs, on the other hand, play a crucial role in negative selection. These cells are specialized to express tissue-restricted antigens (TRAs), proteins normally produced only in one or two tissues in the body. An example of a TRA is insulin—a protein otherwise only produced in the pancreas. The thymic expression of TRA is essential to achieve self-tolerance to all proteins in the body. TRA expression in mTECs is largely dependent on a transcriptional regulator called *autoimmune regulator* protein (AIRE) (see Table 9.1).

CD4 AND CD8 T-CELL CHARACTERISTICS ARE SET IN THE THYMUS

Conventional αβ T cells in the periphery are either CD4 helper or CD8 cytotoxic T cells. This divergence of co-receptor expression and effector cell type is determined in the thymus. This decision is made after surface TCR expression: only positively selected DP thymocytes can transition into CD4 or CD8 SP thymocytes (see Fig. 9.1, step 5). A DP thymocyte with an MHC class I-restricted TCR loses CD4 expression and becomes a cytotoxic CD8 SP thymocyte, while a DP thymocyte with an MHC class II-restricted TCR loses CD8 expression and becomes a helper CD4 SP thymocyte.

The molecular pathways by which CD4/CD8 lineage commitment is established and reinforced include two critical transcription factors: ThPOK and Runx3 (see Table 9.1).[6,15,16] ThPOK is essential for differentiation to the CD4 helper lineage, while Runx3 is essential for differentiation to the CD8 cytotoxic lineage. These factors are mutually antagonistic; for example, Runx3 promotes *Cd8* expression while repressing *Cd4* and ThPOK expression.

Specialized Lymphocytes

Conventional αβ and γδ T cells are not the only cells that develop and emerge from the thymus. There are also specialized lymphocyte populations that are numerically fewer but still play important roles in host immune responses. These include lipid-reactive natural killer T (NKT) cells, CD8αα$^+$ intraepithelial lymphocytes (IELs), and mucosal-associated invariant T (MAIT) cells (see Table 9.1). Compared to conventional αβ T cells, NKT cells and IELs express TCRs that are more self-reactive. Thus, their development is referred to as agonist selection in the thymus.[14] These specialized lymphocytes require distinct molecular factors for their development.

NKT cells express an αβ TCR. However, rather than being peptide specific and MHC class I- or II-restricted, they recognize lipid antigens in the context of CD1d molecules.[14,17] The TCR repertoire expressed by NKT cells is highly restricted, consisting of either TCR Vα14-Jα18 (mice) or Vα24-Jα18 (humans) chains paired with limited TCR Vβ chains. Due to this "invariant" TCR, these cells are named invariant NKT (iNKT) cells. Selection of iNKT cells happens in the thymic cortex at the DP stage. However, rather than being positively selected by thymic cTEC like conventional T cells, iNKT cells are selected by other DP thymocytes presenting lipid antigens by CD1d, a nonclassical MHCI-like molecule. Strong TCR signaling at this stage, along with interactions between signaling lymphocyte activation molecule family (SLAMF) receptors. Drives agonist selection of iNKT cells. iNKT cells are specified by the expression of the transcription factor PLZF and can further differentiate into different iNKT effector subsets within the thymus.

Like NKT cells, MAIT cells have a limited TCR repertoire due to restricted TCRα chain usage. These TCR generally consist of either Vα19-Jα33 (mice) or Vα7.2-Jα33 (humans) paired with limited TCR Vβ chains.[18] MAIT cells recognize metabolites of vitamin B in the context of the nonclassical MHCI-like molecule MR1 and are selected for during the DP stage of thymocyte development by other DP thymocytes.

IEL are T cells that reside within the gut epithelium and include a population that express αβ TCR and CD8αα homodimers. These CD8αα$^+$ IEL derive from agonist-selected thymic precursors (IELp) and have a small repertoire size.[14,19] During thymocyte selection, strongly self-reactive DP thymocytes undergo clonal deletion; but without CD28 costimulation, more cells are diverted into the IELp fate. These TCR αβ$^+$ IELp can then be found within DN thymocyte population, and most are localized within the cortex.

KEY CONCEPTS

The Thymus Selects for Clones That Are Useful and Safe

- A positive selection step selects for progenitors with an αβ TCR that can bind to MHC molecules, leading to an "MHC restricted," or useful, pool of T cells in each individual.
  - These MHC molecules must be present on epithelial cells of the thymus.
- CD4 helper and CD8 killer lineage is determined by which class of MHC molecule is recognized—a process called *lineage commitment*.
  - MHC class II recognition by the TCR and CD4 co-receptor generates helper T cells, through the transcription factor ThPOK.
  - MHC class I recognition by the TCR and CD8 co-receptor generates killer T cells, through the transcription factor Runx3.
- A negative selection step eliminates T cells with the most strongly self-reactive TCRs, leading to a "tolerant" or safe pool of T cells.
- The thymus also selects and provides instruction for smaller populations of specialized T cells, such as those that reside at barrier surface (skin and gut), recognize metabolites (MAIT cells), or recognize lipids (NKT cells).

TOLERANCE

An effective immune system is one that can respond robustly to foreign pathogens without having unwanted responses against the host's own cells and tissues (Chapter 10). This immunological tolerance to self-antigens is enforced early during T-cell development in the thymus by the elimination of strongly self-reactive thymocytes by means of clonal deletion. This central tolerance mechanism is not complete. Self-reactive T cells are able to escape into the periphery. However, there is an additional tolerance mechanism, the development of regulatory T cells (Treg), that helps control immune responses.[20]

Regulatory T cells are self-reactive CD4 T cells that express CD25 and the transcription factor FoxP3 (see Table 9.1). They generally suppress immune responses, as opposed to initiating them. They are only a small proportion of both developing CD4 SP thymocytes (approximately 1%) and CD4 T cells in secondary lymphoid organs (approximately 10% to 15%), but they play a key role on immune homeostasis.[21]

Despite being self-reactive, some cells survive and undergo "agonist selection" instead of being deleted.[10,14] This step for Treg development is thought to happen in the thymus at the

CD4 SP stage in the medulla (see Fig. 9.1, step 6). As TCR that have high affinity for self-peptide:MHC can also be clonally deleted, there is some overlap between conventional T cells that undergo apoptosis and Treg selection (see Fig. 9.2). Treg progenitors are able to avoid apoptosis due to pro-survival cytokine signals, such as interleukin-2 (IL-2). Cytokines also play a role in inducing FoxP3 expression, converting progenitors into mature $CD25^+$ $FoxP3^+$ Treg.

Autoimmunity Results From the Lack of Tolerance (AIRE, FoxP3)

Treg can be specific for tissue-restricted antigens. AIRE, which mediates the expression of these TRA by thymic antigen-presenting cells like mTEC, is thus critical for both Treg development and clonal deletion of self-reactive thymocytes.[21] Mice that lack *Aire* have increased numbers of autoreactive T cells in peripheral tissues, leading to multi-organ destruction and autoantibody production. In humans, mutations in *AIRE* lead to multi-organ autoimmune disease, particularly in endocrine organ. This is referred to as *autoimmune polyendocrine syndrome 1* (APS-1) *or autoimmune polyendocrinopathy candidiasis ectodermal dystrophy* (APECED).[20,22] The importance of Treg in immune tolerance is demonstrated by patients that have mutations in *FOXP3*. Such patients develop a fatal lymphoproliferative disease called *immune dysregulation, polyendocrinopathy, enteropathy, X-linked* (IPEX) *syndrome* (Chapter 13).

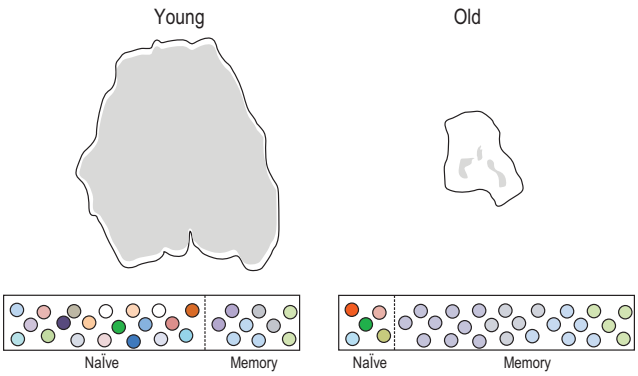


FIG. 9.3 Thymic Atrophy Results in a Less-Diverse Repertoire With Age. The thymus expands early in life, reaching peak size at about puberty. The aging thymus undergoes involution, where reduced tissue mass and altered thymic architecture leads to a decrease in T-cell output. While a young thymus can support robust thymopoiesis, leading to a diverse T-cell repertoire (distinct clones depicted as different colors), an aged thymus is characterized by reduced T-cell output. Due to this change in thymic function with age, the naïve T-cell pool has reduced size and diversity, and, as a consequence, the T-cell pool of an older individual is skewed more towards memory populations.

KEY CONCEPTS

An Important Function of the Thymus is to Make T Cells Tolerant of Self

- Some strongly reactive thymocytes undergo apoptosis, eliminating the most dangerous clones from the repertoire—this process is called *clonal deletion*.
- Other strongly reactive thymocytes upregulate the transcription factor FoxP3 instead, and become regulatory T cells, which also contribute to tolerance.
  - Patients with mutations in the *FOXP3* gene experience lethal lymphoproliferative disease.
- The transcription factor AIRE causes tissue-specific antigens to be expressed in medullary epithelial cells of the thymus, leading to clonal deletion and Treg induction, so that T cells are generally tolerant of tissues.
  - Patients with mutations in the *AIRE* gene experience multi-organ autoimmune disease.

MIGRATION OF THYMOCYTES INTO THE PERIPHERY

Following positive selection, SP thymocytes spend a few days in the thymic medulla perusing APCs there for tolerance purposes and completing their differentiation to become functional T cells, before emigrating via blood vessels at the CMJ.[3,23] The competence to emigrate is signaled by the upregulation of a zinc-finger transcription factor Kruppel-like factor 2 (KLF2) (see Table 9.1). KLF2 upregulates both sphingosine-1-phosphate receptor 1 (S1PR1), a lipid receptor that allows for thymic egress, and CD62L (also known as L-selectin), which allows for entry into peripheral lymphoid tissue. Recent thymic emigrants (RTEs) and other naïve T cells express IL-7 receptor, and IL-7 signaling, along with low or "tonic" levels of TCR signaling, allow for survival and homeostatic cycling.[24] These T cells can then function in adaptive immune responses.

THYMIC ATROPHY

In the young, approximately 10^6 thymocytes in mice and 10^7 to 10^8 thymocytes in humans are exported from the thymus into the circulation every day. These populations account for about 1% of all thymocytes.[3] However, there is a decline in T-cell output by the thymus with age. The decline begins around puberty[25] and accelerates in older adults (Chapter 21). This decrease in T-cell output is due to age-related regression of the thymus, or *thymic involution* (Fig. 9.3).

During thymic involution, there is both a reduction in tissue mass and an alteration in the thymic architecture, with an accumulation of adipose tissue. This leads to reduced thymopoiesis and a loss of diversity among the thymic emigrants that join the naïve T-cell pool. These RTEs are not only fewer in number but also of different quality: aged RTEs have lower proliferative capacity and defective TCR-induced calcium signaling. Taken together, this results in elderly individuals having reduced immune responses, which leads to higher rates of infection, autoimmune disease, and cancer.

Age-related disorganization of thymic stromal cells is associated with accumulation of adipose tissue, and obesity can enhance involution. Furthermore, age and obesity are not the only factors that can drive thymic atrophy. Acute stressors—examples being infections (bacterial, viral, fungal) and medical treatments (such as chemotherapy or stem cell transplantation)—can lead to rapid, stress-induced thymic atrophy.[26] When this stressor is removed, the tissue will often recover. However, reduced thymic output can lead to lymphopenia in patients, and this delay in thymic function during recovery can leave patients vulnerable to opportunistic infections.

KEY CONCEPTS

T-Cell Homeostasis

- Once thymocytes complete maturation in the thymus, they emigrate and become established in lymphoid organs throughout the body.
  - This process requires the sphingosine phosphate receptor, S1PR1, and L-selectin.
- Naïve T cells require the cytokine IL-7 and "basal" TCR signals from continued MHC recognition for long-term survival.
- The thymus atrophies starting after puberty, producing fewer new T cells as individuals age.
  - This contributes to poor immune responses in the elderly.
- T cells are depleted, termed lymphopenia, during some acute and chronic viral infections and some medical treatments.

ON THE HORIZON

- Deducing how the T-cell receptor senses MHC class I versus MHC class II and coordinates the expression of the appropriate CD8 or CD4 co-receptor, respectively.
- Understanding what thymic processes malfunction in common autoimmune diseases, such as type I diabetes.
- Better tools and markers to study αβ T-cell development and the acquisition of innate-like functions prior to thymic export.
- Developing strategies to combat thymic involution with age and increase T-cell development, especially following bone marrow transplantation.

REFERENCES

1. Brandt LJB, Barnkob MB, Michaels YS, et al. Emerging approaches for regulation and control of CAR T cells: a mini review. *Front Immunol*. 2020;11:326.
2. Miller JFAP, Osoba D. Current concepts of the immunological function of the thymus. *Physiol Rev*. 1967;47:437–520.
3. Dzhagalov I, Phee H. How to find your way through the thymus: a practical guide for aspiring T cells. *Cell Mol Life Sci*. 2012;69:663–682.
4. Abramson J, Anderson G. Thymic epithelial cells. *Annu Rev Immunol*. 2017;35:85–118.
5. Rothenberg EV. T cell lineage commitment: identity and renunciation. *J Immunol*. 2011;186:6649–6655.
6. Shah DK, Zuniga-Pflucker JC. An overview of the intrathymic intricacies of T cell development. *J Immunol*. 2014;192:4017–4023.
7. Taghon T, Rothenberg EV. Molecular mechanisms that control mouse and human TCR-alphabeta and TCR-gammadelta T cell development. *Semin Immunopathol*. 2008;30:383–398.
8. Yui MA, Rothenberg EV. Developmental gene networks: a triathlon on the course to T cell identity. *Nat Rev Immunol*. 2014;14:529–545.
9. Ciofani M, Zuniga-Pflucker JC. Determining gamma-delta versus alpha-beta T cell development. *Nat Rev Immunol*. 2010;10:657–663.
10. Klein L, Kyewski B, Allen PM, et al. Positive and negative selection of the T cell repertoire: what thymocytes see (and don't see). *Nat Rev Immunol*. 2014;14:377–391.
11. McDonald BD, Bunker JJ, Erickson SA, et al. Crossreactive alphabeta T cell receptors are the predominant targets of thymocyte negative selection. *Immunity*. 2015;43:859–869.
12. Breed ER, Watanabe M, Hogquist KA. Measuring thymic clonal deletion at the population level. *J Immunol*. 2019;202:3226–3233.
13. Breed ER, Lee ST, Hogquist KA. Directing T cell fate: how thymic antigen presenting cells coordinate thymocyte selection. *Semin Cell Dev Biol*. 2018;84:2–10.
14. Stritesky GL, Jameson SC, Hogquist KA. Selection of self-reactive T cells in the thymus. *Annu Rev Immunol*. 2012;30:95–114.
15. Singer A, Adoro S, Park JH. Lineage fate and intense debate: myths, models and mechanisms of CD4- versus CD8-lineage choice. *Nat Rev Immunol*. 2008;8:788–801.
16. Wang L, Bosselut R. CD4-CD8 lineage differentiation: Thpok-ing into the nucleus. *J Immunol*. 2009;183:2903–2910.
17. Wang H, Hogquist KA. How lipid-specific T cells become effectors: the differentiation of iNKT subsets. *Front Immunol*. 2018;9:1450.
18. Godfrey DI, Koay HF, McCluskey J, et al. The biology and functional importance of MAIT cells. *Nat Immunol*. 2019;20:1110–1128.
19. Ruscher R, Hogquist KA. Development, ontogeny, and maintenance of TCRalphabeta(+) CD8alphaalpha IEL. *Curr Opin Immunol*. 2019;58:83–88.
20. Gregersen PK, Behrens TW. Genetics of autoimmune diseases—disorders of immune homeostasis. *Nat Rev Genet*. 2006;7:917–928.
21. Owen DL, Sjaastad LE, Farrar MA. Regulatory T cell development in the thymus. *J Immunol*. 2019;203:2031–2041.
22. Caramalho I, Nunes-Cabaco H, Foxall RB, Sousa AE. Regulatory T-cell development in the human thymus. *Front Immunol*. 2015;6:395.
23. James KD, Jenkinson WE, Anderson G. T-cell egress from the thymus: should I stay or should I go? *J Leukoc Biol*. 2018;104:275–284.
24. Fry TJ, Mackall CL. The many faces of IL-7: from lymphopoiesis to peripheral T cell maintenance. *J Immunol*. 2005;174:6571–6576.
25. Palmer DB. The effect of age on thymic function. *Front Immunol*. 2013;4:316.
26. Gruver AL, Sempowski GD. Cytokines, leptin, and stress-induced thymic atrophy. *J Leukoc Biol*. 2008;84:915–923.

T-Cell Activation and Tolerance

Erik J. Peterson

Activation of T lymphocytes during immune responses triggers a series of gene transcriptional regulatory, proliferation, differentiation, and effector functions. T-cell functions coordinate with other leukocytes to permit the immune system to react against foreign antigens without initiating self-reactivity or autoimmunity. Each of these functions is fully dependent on environmental cues that are recognized by cell surface receptors and are then translated through biochemical alterations within the cell. This chapter discusses signal transduction through one of the best studied of these receptors, the antigen-specific T-cell receptor (TCR) complex. It addresses the mechanisms whereby signals propagated through the TCR combine with those from costimulatory receptors to yield either productive activation or immune tolerance. It also discusses how abnormal TCR signaling, and imbalanced signaling through costimulatory or coinhibitor molecules, can contribute to T-cell dysfunction and disease (Table 10.1). Targeting these molecular pathways has resulted in several clinically relevant drugs currently used to treat autoimmunity, transplant rejection, and cancer.

CLINICAL RELEVANCE

- Dysfunction or deficiency of T-cell signaling proteins (induced or spontaneous) has been causally linked to disease states in animals or humans
- TCR signaling molecules wherein mutations may lead to immune deficiency and/or T-cell hypofunction include:
  - CD45
  - LCK
  - ZAP-70
  - SLP-76
  - LAT
  - Mst1
  - PI3Kδ
  - RasGRP1
- Molecules wherein mutations may lead to exaggerated lymphocyte proliferation include:
  - CTLA-4
  - PD-1
  - SHP-1
  - CD95/CD95 ligand
  - SAP
  - CBL
  - ZAP-70
  - LYP
  - DGK

THE T-CELL ANTIGEN RECEPTOR COMPLEX

The TCR complex consists of a ligand-binding TCR α/β or γ/δ heterodimer (Chapter 4) in association with the CD3/ζ chain complex, which provides transmembrane signal transduction capability.[1] Specificity of the TCR for antigen resides exclusively within the highly polymorphic, clonotypic, ligand-binding α/β or γ/δ heterodimers. Although many of the biochemical events leading to α/β and γ/δ T-cell activation are similar, α/β T cells exhibit a broader spectrum of antigen reactivity and are thought to participate in a wider range of specific immune responses. This chapter focuses on α/β T cells.

The α/β TCR specifically recognizes short (8 to 9 amino acid) polypeptide ligands bound to major histocompatibility complex (MHC) protein (Chapter 5) on the surface of antigen-presenting cells (APCs) (Chapter 6). Coreceptor molecules expressed on subsets of α/β T cells determine whether the TCR recognizes class I or class II MHC. CD4 T cells are stimulated by processed exogenous antigen presented by class II MHC molecules on the surface of professional APCs. CD8 T cells respond to peptides synthesized by APCs and presented by class I molecules. CD4 and CD8 associate with MHC class II and class I molecules, respectively, to stabilize the tripartite interaction between the TCR, antigen, and MHC, which increases the effectiveness of TCR engagement.

Although the α/β chains of the TCR contain all information necessary for antigen/MHC binding, these proteins are not sufficient to initiate the intracellular biochemical events that signal antigen recognition. Instead, signal transduction is accomplished by noncovalently associated CD3 and TCRζ polypeptides, which include several pairs of transmembrane hetero- or homodimers (Fig. 10.1). Each CD3 and ζ chain derives signaling capacity from the presence of one or more cytoplasmic domains known as *immunoreceptor tyrosine-based activation motifs* (ITAMs).[2]

KEY CONCEPTS

T-Cell Receptor Induces Serial Tyrosine Phosphorylation

TCR engagement activates several families of protein tyrosine kinases, which are required for propagation of second messenger-instigated intracellular signaling:

- Src family: LCK, FYN
- Syk family: ZAP-70
- Tec family: ITK, RLK

TABLE 10.1 Phenotypes Associated With Deficient Function of Selected T-Cell Signaling Molecules

		PHENOTYPE	
Molecule	**Affected Signaling Event**	**Mouse**	**Human**
TCR Signaling			
CD3 γ	TCR expression	$B^+T^+NK^+$ SCID	$B^+T^+NK^+$ SCID
CD3 ε	TCR expression	$B^+T^-NK^+$ SCID	$B^+T^{+/-}NK^+$ SCID
CD3 δ	TCR expression	$B^+T^-NK^+$ SCID	$B^+T^-NK^+$ SCID
CD3 ζ	TCR expression, TCR-mediated PTK activation	$B^+T^+NK^+$ SCID	$B^+T^+NK^+$ SCID
ZAP-70	TCR-mediated PTK activation	$B^+T^{+/-}NK^+$ SCID. TCR αβ T cells are absent, but TCR γδ T cells survive. Arthritis occurs in some inbred strains	$B^+T^{+/-}NK^+$ SCID. CD8 T-cell lymphopenia. Overexpressed in some hematological malignancies
LCK	TCR-mediated PTK activation	$B^+T^{+/-}NK^+$ SCID. Impaired thymopoiesis and proliferation	$B^+T^+NK^+$ SCID. CD4 lymphopenia, absent CD28 expression on CD8 T cells, and hypogammaglobulinemia
CD45	Maintenance of SRC family PTK in "open" conformation	$B^+T^{+/-}NK^+$ SCID. Impaired thymopoiesis	$B^+T^{+/-}NK^+$ SCID. Impaired thymopoiesis, decreased cytotoxic T-cell responses, progressive hypogammaglobulinemia, genetic polymorphisms may correlate with increased prevalence of autoimmune disease
SAP	SHP-2 binding to SLAM	Increased susceptibility to lymphocytic choriomeningitis virus, reduced IgE production, NKT-cell deficiency	X-linked lymphoproliferative disease (XLP) with B-cell hyperresponsiveness, NKT-cell deficiency
WASP	Actin polymerization	Decreased T-cell proliferation and interleukin-2 (IL-2) production	Wiscott-Aldrich syndrome (immunodeficiency, atopic dermatitis, thrombocytopenia, bloody diarrhea)
CBL/CBL-b[a]	E3 ubiquitin ligase. Recruitment of CrKL/C3G inhibitory complex	Hyperproliferative T cells[a]	Proto-oncogene for leukemia
LAT	Coupling PTK activation to downstream signals	$B^+T^-NK^+$ SCID. Absolute block in thymopoiesis	
SLP-76	Coupling PTK activation to downstream signals	$B^+T^-NK^+$ SCID. Absolute block in thymopoiesis. Defect in vascular/lymphatic development	
ITK/RLK	Amplification of proximal PTK signals. Activation of PLCγ1	Defective Th2 immune responses	
CTLA-4	Inhibition of CD28-mediated costimulation	Fatal lymphoproliferative disease with myocarditis, pancreatitis	Allelic variants associated with autoimmunity, including Hashimoto thyroiditis, Graves disease, and systemic lupus erythematosus
SHP-1	Downregulation of PTK activity	Autoimmunity, inflammatory lung disease. "Moth-eaten" mice	
LYP (Lymphoid phosphatase; *PTPn22* gene product)	Attenuation of LCK activity	Augmented TCR-stimulated IL-2 production and proliferation	Allelic variants are associated with increased risk of rheumatoid arthritis, systemic lupus erythematosus, type 1 diabetes mellitus
DGKζ	Downregulation of DAG-dependent Ras activation	Impaired T-cell anergy induction	
Mst1 (STK4)	Suppression of Akt, phosphorylation of FoxO transcription factor	Inefficient thymic egress; impaired positive selection; impaired Treg development and function	Recurrent infections; progressive loss of peripheral CD4 T cells; autoimmune hemolytic anemia
IL-2R Signaling			
γc	Coupling IL-2 binding to JAK activation	$B^+T^-NK^-$ SCID	$B^+T^-NK^-$ SCID, X-linked SCID
JAK3	Phosphorylation of STAT proteins	$B^+T^-NK^-$ SCID	$B^+T^-NK^-$ SCID

γc, Common γ-chain (IL2Rγ); *IgE*, immunoglobulin E; *NK*, natural killer; *SCID*, severe combined immunodeficiency; *TCR*, T-cell receptor.
[a]CBL and CBL-b are closely related; CBL-b-deficient mice develop autoimmune features and more severe lymphoproliferative disease compared with mice lacking CBL.

Activation of Protein Tyrosine Kinases by the T-Cell Receptors and the Role of the Immunoreceptor Tyrosine-Based Activation Motifs

The likelihood of a given TCR interaction with peptide/APC ligand to result in productive intracellular signaling is a function of multiple biophysical factors, including TCR-binding affinity ("off rate"), conformational states of TCR components driven by application of mechanical forces, as well as the activation states of membrane-associated enzymatic mediators.[1] Together these factors confer extraordinary sensitivity and selectivity of the TCR for ligand, and determine signal strength. An array of membrane-proximal molecular events translates signals conferred by ligand binding at the cell surface.

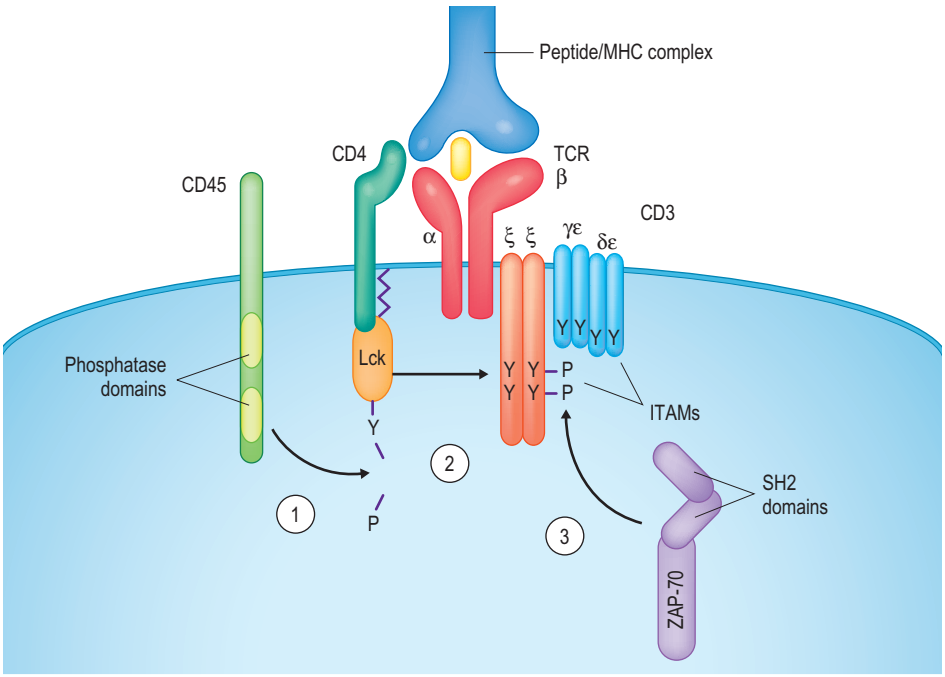


FIG. 10.1 Biochemical Events in Early T-cell Receptor *(TCR)* Signaling. (1) The tyrosine phosphatase CD45 dephosphorylates the negative regulatory tyrosine residue on the CD4-associated protein tyrosine kinase (PTK) LCK, maintaining LCK in an activatable conformation. (2) Engagement of the TCR α/β heterodimer and the CD4 (or CD8) coreceptors by major histocompatibility complex *(MHC)*–bound peptide antigen brings activated LCK into proximity with immunoreceptor tyrosine-based activation motif *(ITAM)*-bearing CD3 chains. LCK phosphorylates the CD3ζ chain within ITAMs. (3) The phosphorylated CD3ζ-chain ITAMs interact with the tandem SH2 domains of the cytoplasmic PTK ZAP-70, permitting activation of ZAP-70 and phosphorylation of downstream substrates.

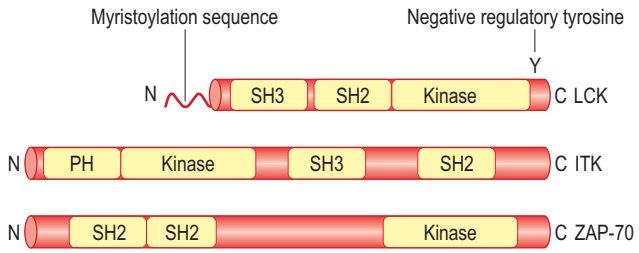


FIG. 10.2 Domain Organization of T-cell Receptor (TCR)-Stimulated Protein Tyrosine Kinases (PTKs). Comparative schematic representation of members of three families of PTKs required for T-cell-activating signals. In addition to catalytic domains, LCK (SRC family), ITK (TEC family), and ZAP-70 (SYK family) each contain regions responsible for mediating protein–protein interactions, including SH3 and SH2 domains. SH3, Src homology 3; SH2, Src homology 2; PH, pleckstrin homology.

Among the earliest biochemical events following engagement of the TCR is the activation of LCK and FYN, two members of the SRC family of protein tyrosine kinases (PTKs).[2] Shared LCK and FYN functional motifs (Fig. 10.2) include an amino-terminal myristoylation sequence that directs membrane localization, a SRC homology 3 (SH3) domain that permits associations with other proteins containing regions rich in proline residues, a SRC homology 2 (SH2) domain that dictates interactions with proteins phosphorylated on tyrosine residues, a catalytic region, and a carboxyl-terminal tyrosine residue. The precise mechanism whereby LCK and FYN are stimulated by the TCR is not clear, but both have been shown to associate physically with TCR CD3 components and/or coreceptors CD4 and CD8.

SRC family protein tyrosine kinase (PTK) enzymatic function is regulated, in part, by the state of tyrosine phosphorylation of the kinase. When the conserved carboxyl-terminal tyrosine residue is phosphorylated, SRC family PTKs adopt a "closed" conformation that is the product of an intramolecular interaction between that phosphotyrosine and the SH2 domain (Fig. 10.3). This intramolecular interaction inhibits the enzymatic activity of the PTK, limiting subsequent tyrosine phosphorylation–dependent signaling events. Phosphorylation of the carboxyl-terminal tyrosine (Y505 in LCK and Y527 in FYN) is dynamically regulated.[3] Phosphate is transferred to the carboxyl-terminal tyrosine by the cytoplasmic negative regulator PTK CSK and is removed by the transmembrane protein tyrosine phosphatase CD45. Current models hold that TCR signal propagation depends upon both recruitment of coreceptor-associated catalytically pre-activated LCK and de novo TCR-induced LCK activation.

Phenotypes of CD45-deficient cells in mice and humans highlight the critical regulatory importance of SRC family PTK carboxyl-terminal tyrosine phosphorylation.[4] TCR signal transduction in cell lines lacking CD45 is blocked at the most proximal step, and CD45 "knock-out" mice exhibit profound defects in thymocyte development and subsequent T-cell activation. CD45 deficiency in humans results in a T^-, B^+, NK^+ severe

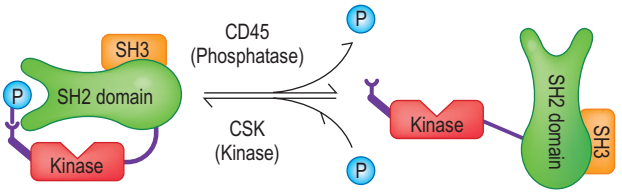


FIG. 10.3 Model for Dynamic Regulation of LCK by Intramolecular Interaction Between an SH2 Domain and Phosphotyrosine. The transmembrane phosphatase CD45 dephosphorylates tyrosine 505 in the carboxyl-terminus of the SRC family protein tyrosine kinase (PTK) LCK. CD45 activity maintains LCK in an "open" conformation, permitting LCK kinase domain access to intracellular substrates. CSK activity opposes that of CD45; phosphorylation of tyrosine 505 results in an intramolecular interaction between the SH2 domain and phosphotyrosine. Inhibition of LCK kinase activity correlates with the "closed" conformation *(left)*.

combined immunodeficiency (SCID) (Chapter 34). These outcomes correlate with markedly impaired LCK enzymatic activity and hyperphosphorylation of Y505, the regulatory tyrosine.

Following TCR engagement and above-described PTK activation, numerous cellular substrates become tyrosine-phosphorylated, including the CD3 and TCR ζ-chain immunoreceptor tyrosine-based activation motifs (ITAMs) (see Fig. 10.1). In resting T cells, key tyrosine residues within the ITAMs are embedded within the hydrophobic core of the plasma membrane lipid bilayer. Upon TCR triggering, conformational changes induced within the CD3 cytoplasmic tails result in enhanced tyrosine accessibility to the action of SRC family kinases.[2] ITAM phosphorylation creates a docking site for another cytosolic PTK, ζ-associated phosphoprotein of 70 kilodaltons (kDa) (ZAP-70). ZAP-70, a member of the SYK family PTKs, contains a catalytic domain that is located carboxyl-terminal to two tandem SH2 domains (see Fig. 10.2). The ZAP-70 SH2 domains have affinity for phosphotyrosine present within ITAMs. Thus inducible phosphorylation of the CD3 and ζ-chain ITAMs results in the formation of docking sites that mediate recruitment of ZAP-70. Upon recruitment to TCR, ZAP-70 enzymatic activity is increased as a result of phosphorylation by LCK as well as autophosphorylation. The net result of these phosphorylations is conversion of the TCR from an enzymatically inactive ligand-binding complex to a potent membrane-associated PTK.

LCK and ZAP-70 are critically important for both thymocyte development and mature T-cell activation. Mice deficient in Zap-70 or Lck exhibit a significant yet incomplete block in early T-cell development (Chapter 9). The pre-TCR (a complex present on immature thymocytes that includes signaling components thought functionally similar to the TCR on mature T cells) appears to require Src and Syk family PTKs to transduce signals. ZAP-70 deficiency and abnormal LCK function in humans create a T^-, B^+, NK^+ SCID.[5]

Downstream of SRC and SYK family PTKs, TCR engagement results in the activation of a third family of cytosolic PTKs, the Tec family, which includes TEC, ITK, and RLK. Tec PTKs contain SH2, SH3, and catalytic domains, as well as pleckstrin homology (PH) domains that mediate interactions with membrane-localized phospholipids (see Fig. 10.2). PH domains permit recruitment of Tec family kinases to the plasma membrane, where they can phosphorylate important substrates.[6] ITK positively regulates antigen-receptor signaling through recruitment and activation of the lipid modulator PLCγ to a "signalosome" nucleated by the cytoplasmic adaptor protein SLP-76, and commencement of signaling via PLCγ and Ras/MAPK cascades. Mice deficient in Tec kinases display variable partial defects in thymocyte development and peripheral T-cell maturation. The role of Itk in effector CD4 T-cell fate is shown by observations that loss of Itk leads to defective Th17 (Chapter 11) and regulatory T cell (Treg) 1 (Tr1) (Chapter 13) subset differentiation, but increased generation of Foxp3-positive Treg cells.

KEY CONCEPTS

T-Cell Receptor Signaling Pathways

TCR engagement leads to the activation of signaling cascades and transcription factors
- PLCγ1 activation →
  - IP3 → calcium flux → NFAT
  - Diacylcglycerol → PKC → NF-κB
  - Diacylglycerol → Ras/MAPK → AP-1

Second Messenger Cascades Downstream of the T-Cell Receptor–Stimulated Protein Tyrosine Kinases

TCR engagement incurs numerous biochemical changes that are dependent on the activation of membrane-proximal PTKs. One intermediate event in TCR signaling is activation of the membrane-associated enzyme phospholipase Cγ1 (PLCγ1) (Fig. 10.4).[7] PLCγ1 is phosphorylated by multiple TCR-dependent PTKs, including both ZAP-70 and members of the Tec family. TCR-stimulated tyrosine phosphorylation alone is not sufficient to activate PLCγ1; relocalization of the enzyme into adaptor-protein nucleated complexes probably plays a critical role.

Activated PLCγ1 catalyzes the hydrolysis of plasma membrane phospholipid phosphatidylinositol-4,5-bisphosphate (PIP_2), giving rise to two second messengers, inositol 1,4,5-trisphosphate (IP_3), a sugar, and the lipid diacylglycerol (DAG). IP_3 binds to a cognate receptor on endoplasmic reticulum (ER) and releases stored calcium from that organelle. Falling calcium ER concentrations are sensed by stromal interaction molecule 1 (STIM1), an EF-hand domain–containing protein localized in the ER membrane.[8] STIM1 aggregation activates membrane-localized store-operated calcium entry channels, including members of the transmembrane protein Orai family.

An Orai-dependent increase in intracellular calcium serves as a second messenger that activates calcineurin, a serine/threonine phosphatase. Activated calcineurin dephosphorylates members of the nuclear factor of activated T cells (NFAT) family, allowing NFAT to translocate from the cytosol to the nucleus in order to activate genes important for T-cell activation. Knowledge of the calcineurin pathway has been exploited clinically in disorders wherein suppression of T-cell activation is desired. Calcineurin inhibitors cyclosporine and tacrolimus have long been employed in the prevention of human solid-organ transplant rejection (Chapter 89) and in treatment of T-cell–driven autoimmune diseases (Chapter 51). Orai regulators include the transient receptor potential melastatin (TRPM) family of membrane ion channels. TRPM4 shapes the magnitude of

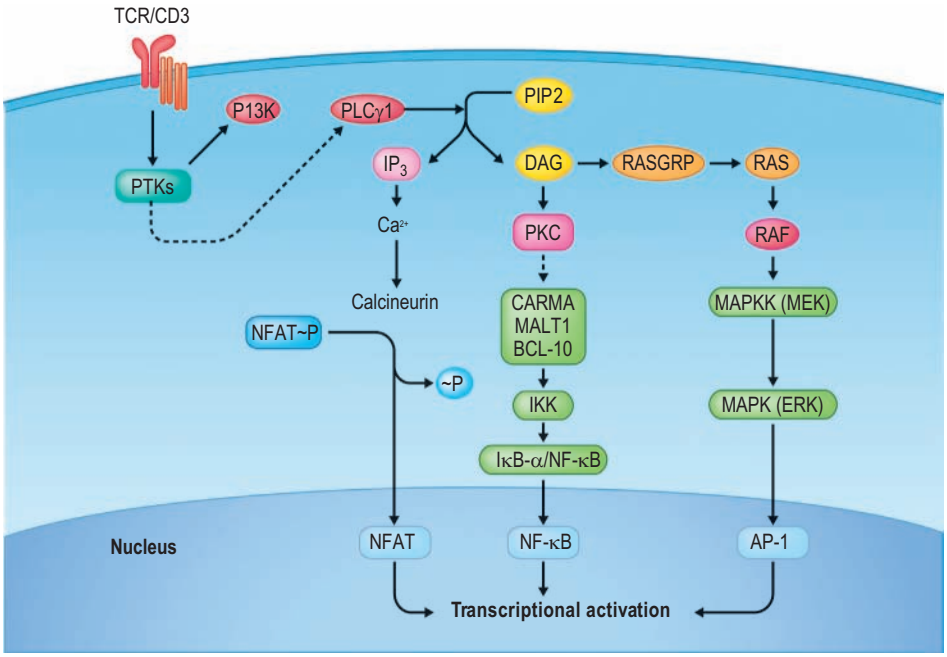


FIG. 10.4 Signaling Pathways Activated by T-cell Receptor *(TCR)* Engagement. TCR ligation results in activation of protein tyrosine kinases *(PTKs)*, such as LCK and ZAP-70. Phospholipase Cγ1 *(PLCγ1)* becomes phosphorylated and activated by PTKs, including ITK. Hydrolysis of phosphatidyl inositol bisphosphate *(PIP_2)* by PLCγ1 releases diacylglycerol *(DAG)* and inositol trisphosphate *(IP_3)*. IP_3 stimulates an increase in intracellular calcium concentration, which activates the phosphatase calcineurin. Calcineurin dephosphorylates nuclear factor of activated T cells *(NFAT)*, thereby signaling NFAT translocation to the nucleus. The formation of DAG leads to activation of RAS-GRP1 GEF activity and RAS activation. Active RAS binds and stimulates the kinase RAF1, initiating a cascade of serine/threonine kinases (MAPK cascade), leading to phosphorylation and nuclear translocation of the extracellular signal-regulated kinase *(ERK)* kinases. DAG formation also results in activation of the CARMA/BCL-1/MALT1 complex, leading to phosphorylation of IκB kinase *(IKK)*. Active IKK phosphorylates IκB-a, leading to IκB-α degradation and release of NF-κB to the nucleus. TCR-activated PI3K catalyzes formation of PIP_3 from membrane-associated PIP_2; the phosphatase PTEN antagonizes PIP_3 formation. PIP_3 binds to the Akt lipid-binding pleckstrin homology domain, a required element in Akt activation. Active Akt both promotes the PKC-CARMA-NFκB pathway and blocks FoxO-dependent transcriptional regulation. Transcription factors NFAT, NF-κB, and those activated by the MAPK pathway cooperate to upregulate transcription of genes, such as *IL-2*, critical for T-cell activation.

TCR Ca^{2+} signals, and TRPM7 plays a key role in TCR signaling termination and T-cell homeostasis. Purinergic ionotropic receptors (P2RX) are membrane channels that are activated by Ca^{2+} entry into mitochondria due to Orai/Stim function. ATP transit through P2RX promotes Th17 differentiation and suppresses Treg development.[8]

Another product of TCR-driven PIP_2 hydrolysis, DAG, functions as second messenger to a parallel cascade of TCR signaling intermediates, including protein kinase D (PKD), Ras guanyl nucleotide releasing proteins (RasGRPs), and members of the protein kinase C (PKC) family of serine/threonine kinases.[9] PKD cooperates with PKC-dependent signals to activate high-affinity binding capacity and clustering of integrins, a family of molecules that mediate TCR signal-augmented cell binding to adhesion molecules on APCs.[10] RasGRP mediates activation of the Ras cascade. PKC is essential for full activation of the Ras/extracellular signal-regulated kinase (ERK) and nuclear factor-κB (NF-κB) cascades that are required to mount a TCR-stimulated gene transcriptional program and drive cellular activation.

In T lymphocytes, RasGRP1 is a guanine exchange factor that mediates exchange of guanosine triphosphate (GTP) for guanosine diphosphate (GDP) bound to RAS.[11] RasGRP1-mediated activation can be envisioned as an analog rheostat, leading to varying intensities of downstream-activated Ras signaling, depending on the strength of upstream stimulus.[11] Genetic loss of RasGRP1 leads to severe defects in thymocyte development in mice, and inactivating RasGRP1 mutations leads to human immunodeficiency associated with decreased TCR-induced ERK activation, highlighting the importance of the RasGRP pathway in TCR signaling.

Active RAS activates and recruits the serine/threonine kinase RAF to the plasma membrane. Active RAF, in turn, phosphorylates MEK, which phosphorylates ERK. ERK translocates to the nucleus and phosphorylates and activates several transcription factors that are critical for TCR-induced transactivation of cytokines and other activation genes. Studies in cell lines and in genetically altered mice attest to the central importance of RAS activation for T-cell function. T-cell lines expressing activated RAS produce more IL-2 after TCR engagement, whereas cell lines carrying inhibitory RAS mutants produce minimal IL-2. Similarly, mice transgenic for activating Ras mutants show alterations in thymocyte development and demonstrate a partially stimulated state in the absence of antigen binding.

TCR stimulation also triggers a cascade headed by phosphatidylinositol 3′-hydroxyl kinase (PI3K) (see Fig. 10.4).[12] PI3Ks are composed of two noncovalently bound subunits. The p85 regulatory subunit activates the kinase activity of the p110 catalytic

subunit. PI3K phosphorylates phosphoinositides, which play an important role in the regulation of several downstream serine/threonine kinases, including protein kinase B (PKB) (Akt). Akt promotes the NF-κB pathway downstream of PLCγ1- and DAG-driven PKC. Although TCR engagement alone can stimulate some degree of PI3K function, full activity of the lipid kinase requires costimulation of the T cell through receptors, such as CD28 (see below). The outcomes of PI3K and Akt signaling play essential diverse roles in T-cell development, differentiation, and effector function, most of which maximize effector T-cell activation responses. These include suppressing the induction of Foxp3$^+$ Tregs and the expression of proapoptotic molecules, including Bim or Bad, and of cell cycle inhibitors. At the same time, PI3K signaling promotes T-cell survival via enhancing metabolic processes, including glucose uptake and glycolysis. The importance of this pathway is suggested by the efficacy of PI3K inhibitors in animal models of T-cell–driven autoimmunity, and in treatment of human malignancies, in which inhibition of PI3Kδ activity can lead to an acute impairment of Treg-mediated suppression at tumor and inflammatory sites.[13]

TCR cross-linking also leads to activation of the cytoplasmic serine/threonine kinase PKC-θ. PKC activation results in nuclear translocation of members of the NF-κB family of transcription factors (see Fig. 10.4).[14] In the basal state, NF-κB family members are sequestered in the T-cell cytoplasm through interaction with inhibitors of NF-κB (IκB). In the TCR-stimulated cell, IκB kinases (IKKs) phosphorylate and degrade IκB, leading to transient NF-κB freedom and translocation to the nucleus. Upstream of IKKs, PKC-θ activation results in the formation of a multimolecular activating complex composed of CARMA1, BCL-10, and MALT1 (CBM). Requirements of CBM assembly for optimal TCR signaling have been established through observation of defects in T-cell activation and survival resulting from deficiencies in CBM proteins.

INTEGRATION OF SECOND-MESSENGER PATHWAYS BY ADAPTOR PROTEINS

Adaptor proteins lack enzymatic or transcriptional regulatory activity. Instead, they possess modular domains responsible for subcellular relocalization and intermolecular interactions. Both constitutive and induced intermolecular interactions mediated by adaptor molecules can promote TCR signal transduction.[15]

Adaptor proteins commonly contain modular domains that exhibit affinity for phosphorylated tyrosine residues (Fig. 10.5A). Such regions include the SH2 and phosphotyrosine-binding (PTB) domains, which recognize phosphorylated tyrosine residues within particular sequence contexts. PTB domains obtain their specificity based on residues amino-terminal to the key phosphotyrosine, whereas SH2 domains recognize sequence motifs carboxyl-terminal to phosphotyrosine. Other adaptor domains that confer binding specificity include SH3 modules, which bind proline-rich regions, WW regions that are responsible for interactions with proline/tyrosine or proline/leucine motifs, and PH domains that have specificity for phospholipids.

Several hematopoietic-specific adaptors play essential roles in T-cell development, in coordinating the signals necessary for mature T-cell activation, and in the process of terminating T-cell responses. Two examples are linker of activated T cells (LAT) and SH2 domain-containing leukocyte phosphoprotein of 76 kDa (SLP-76). Both LAT and SLP-76 were identified during efforts to characterize the substrates of PTKs stimulated by TCR engagement.

LAT is an integral membrane protein that contains tyrosine residues within specific sequence motifs that interact with the SH2 domains of other T-cell signaling molecules (see Fig. 10.5A).[16] In TCR-stimulated T cells, LAT inducibly associates with the SH2 domains of GRB2, GADS (Grb2-related adaptor downstream of Shc), PLCγ1, and the regulatory (p85) subunit of PI3K. It is likely that these induced intermolecular interactions are critical for communicating TCR engagement to downstream second-messenger cascades. The importance of LAT for T-cell activation was suggested by complete loss of TCR signaling events downstream of ZAP-70 phosphorylation in LAT-deficient cell lines. Lat plays an essential role in T-cell development, as Lat-deficient mice have significantly decreased thymocyte numbers. Residual Lat-deficient thymocytes are arrested at an early stage of development; peripheral T cells do not develop.

The cytoplasmic adaptor SLP-76 is absolutely required for both T-cell development and signaling via the mature TCR.[7] By means of a proline-rich region, SLP-76 constitutively associates with the adaptor GADS. Through its SH2 domain, SLP-76 can inducibly interact with other tyrosine-phosphorylated adaptors, such as HPK-1 and ADAP. Following TCR-induced tyrosine phosphorylation, SLP-76 also inducibly binds other SH2-domain-containing proteins, including VAV, an exchange factor for the RAC GTP-binding protein NCK, an adaptor protein, and ITK, the Tec family kinase. In mutant T-cell lines lacking SLP-76, TCR-stimulated Zap-70 phosphorylation is normal, but PLCγ1 and RAS/MAPK signaling cascades are not activated, indicating an early functional niche for SLP-76 in TCR signaling. Mice deficient in SLP-76 show complete block in early thymocyte development and possess no peripheral T cells, indicating an absolute requirement for SLP-76 in T-cell development.

One model for how SLP-76 and LAT functionally interact holds that each contributes multiple molecular interactions to a larger "signalosome" complex at the plasma membrane (see Fig. 10.5B).[16] Following TCR engagement, the two adaptors associate with each other, bridged by GADS and PLCγ1. Inducible movement of SLP-76-bound Itk into complex with LAT brings the kinase into proximity with PLCγ1, resulting in its phosphorylation and activation and leading to the generation of IP_3 and DAG, as described above.

Some adaptor proteins play important roles in restraining TCR-mediated signal transduction events, frequently by dictating subcellular localization of regulatory enzymes. For example, in the absence of TCR stimulation, the transmembrane adaptor PAG binds to the cytoplasmic PTK Csk, bringing it to the plasma membrane.[17] Membrane-localized Csk phosphorylates the regulatory tyrosine of LCK, thus keeping LCK inactive in resting T cells.

Coreceptors Transduce Signals That Are Integrated With T-Cell Receptor Signals

TCR stimulation alone does not lead to full T-cell proliferation or initiation of effector function. Instead, the cell must receive complementary signals via TCR accessory molecules.[18] A requirement for multiple signals to engender full activation allows for an extremely sensitive sensor to detect the presence of specific TCR ligands while protecting against inappropriate

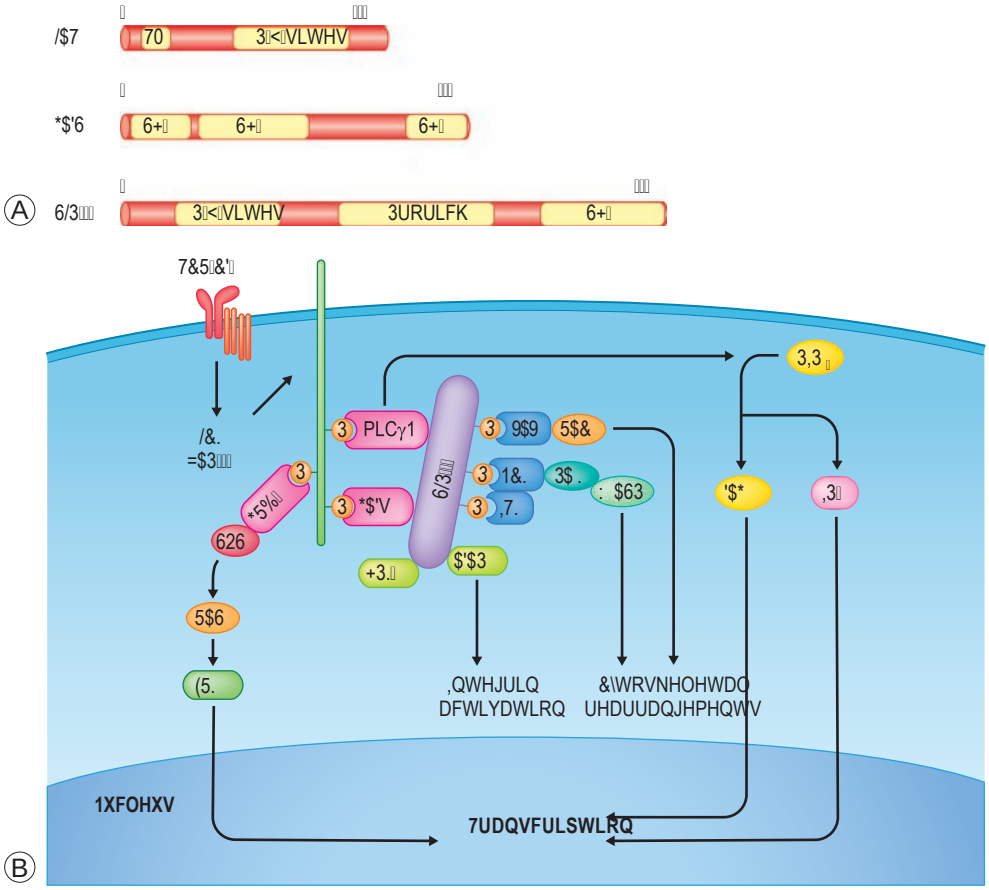


FIG. 10.5 Model for Adaptor Protein-Mediated Coupling of the T-cell Receptor (TCR) to Phospholipase Cγ1 (PLCγ1) Activation. (A) Structural schematics of three adaptors implicated in plasma membrane proximal biochemical events. SH3 domains mediate association with proline-rich regions; SH2 domains associate with phosphorylated tyrosine residues. (B) LAT and SLP-76 are among the substrates of the TCR-activated protein tyrosine kinases (PTKs). When tyrosine residues within the LAT cytoplasmic tail are phosphorylated, GADS binds to LAT through the GADS SH2 domain. Recruitment of GADS results in relocalization of SLP-76, as the proline-rich region of SLP-76 mediates constitutive association with the SH3 domain of GADS. Tyrosine-phosphorylated SLP-76, in turn, becomes associated with ITK via the ITK SH2 domain. ITK is thus brought into proximity with membrane-localized substrates, including PLCγ1. Activation of PLCγ1 leads to hydrolysis of phosphatidylinositol 4,5-bisphosphate *(PIP_2)* and activation of transcription factors, such as nuclear factor-κB (NF-κB), activating protein 1 (AP-1), and nuclear factor of activated T cells (NFAT). SH2 domain containing leukocyte phosphoprotein of 76 kDa *(SLP-76)* also recruits several other signaling molecules, such as VAV, NCK, HPK1, and ADAP, thereby regulating changes in the actin cytoskeleton and adhesion. Phosphorylation of linker of activated T cells *(LAT)* also leads to recruitment of Grb2/SOS and an additional pathway for RAS activation. *TM*, transmembrane domain; *P-Y*, sites for phosphorylation of tyrosine; *pro-rich*, proline-rich regions.

activation of potentially autoreactive—and therefore dangerous—effector T cells. Since T cells respond to antigens presented on APCs, stimulation under physiological conditions involves the potential engagement of multiple coreceptors on the T cell by cognate ligands on the APCs. Some coreceptors may function to increase the avidity of T cells for interacting APCs. However, many coreceptors exhibit intrinsic signal-transducing capacity, both independently of the TCR and via intersection with TCR-driven signaling machinery. Additionally, coreceptors may function to recruit cytoplasmic signaling molecules—for example, enzymes and adaptor proteins—as described above.

The most intensively studied coreceptors are CD4 and CD8 (Chapter 4). Mutually exclusive CD4 or CD8 expression on peripheral T cells defines subsets that respond to MHC class II- or class I-bound peptide antigens, respectively (Chapter 6). Surface-expressed CD4 or CD8 contribute to enhanced TCR signal strength not only because they associate with the extracellular domains of MHC molecules but also because their cytoplasmic domains constitutively associate with LCK, serving to localize a key signal propagator to the TCR complex.[19]

Counterbalanced Costimulatory and Coinhibitory Signals Determine T-Cell Response Thresholds

Productive T-cell stimulation results from a composite of TCR cross-linking and costimulatory signals. Non-TCR activating signals are transduced via a functional family of T-cell surface-expressed proteins termed *costimulatory molecules*. An additional grouping of coreceptors termed *coinhibitors* serve to dampen or restrict potential for TCR-induced cellular activation, and to balance the stimulation-promoting effects of costimulatory molecules.[20] Although functionally divergent,

many costimulatory and coinhibitory molecules share structural features, such as immunoglobulin (Ig)–like extracellular domains.[18] Key costimulatory receptors possessing Ig-like domains include CD28 and ICOS (inducible costimulator); important coinhibitory counterparts include CTLA-4 (cytotoxic T lymphocyte antigen-4, which binds the same ligands as CD28, but with higher affinity), PD-1 (programmed death 1), and BTLA (B- and T-lymphocyte attenuator).[20]

The best-characterized T-cell costimulatory molecule is CD28, a constitutively expressed homodimeric transmembrane glycoprotein.[21] CD28 binds to two ligands expressed on APCs: B7.1 (CD80) and B7.2 (CD86). While ligation of CD28 in isolation has little effect on T-cell activation, when engaged along with the TCR, many TCR signals are augmented. Indeed, concomitant CD28 and TCR engagement is required for activation of naïve T cells.

CD28 ligation engages several signal-transduction pathways implicated in T-cell activation.[21] CD28 contains no intrinsic enzymatic activity, but tyrosine residues within its cytoplasmic tail become inducibly phosphorylated during T-cell activation. These phosphorylated tyrosines recruit several signal-transducing molecules possessing SH2 domains, including GRB2 and the regulatory p85 subunit of PI3K. TCR- and CD28-dependent coactivation of PI3K leads to transactivation of the pro-survival genes *BCL-2* and *BCL-XL*. CD28-driven activation of the GTP-binding protein RAS-related C3 botulinum toxin substrate 1 (RAC1) is crucial for the protein kinase activity of c-JUN N-terminal kinase (JNK), which, in turn, plays an important role in CD28-dependent T-cell cytokine production and apoptosis resistance.

Costimulation with CD28 agonists dramatically augments TCR-driven IL-2 production, both by increasing transcription of the *IL-2* gene and by stabilizing its messenger RNA (mRNA).[21] CD28 is required for T-cell priming during infection and promotes protective secondary T-cell responses during microbial challenge.[22] CD28 is also required for Treg function. Animals engineered to lack CD28 specifically in Treg show splenomegaly, lung inflammation, and accumulation of activated CD4 T cells.[23] These findings suggest that the survival- and activation-promoting function of CD28 costimulation is important for homeostasis of Treg and effector CD4 T-cell subsets. On balance, inhibition of CD28 ligation by cognate ligands CD80/CD86 results in decreased immune responses. In the clinic, blockade of CD28 costimulation using CTLA4-Ig fusion proteins (abatacept; belatacept) can significantly reduce synovial inflammation in human rheumatoid arthritis (Chapter 53) and restrain lymphocyte-driven rejection responses in transplanted kidneys, respectively.

ICOS is an Ig domain-containing costimulatory molecule induced on the T-cell surface after combined TCR and CD28 stimulation.[18] ICOS interaction with APC-expressed ICOS-L is required for development of T-follicular helper (Tfh) cells, a subset required for germinal center formation and for B-cell antibody class switching (Chapter 11). In murine models of rheumatoid arthritis (Chapter 53) and multiple sclerosis (Chapter 66), ICOS antibody blockade results in reduced Tfh and germinal center formation, associated with suppression of autoimmune responses. ICOS deficiency associates with human common variable immune deficiency (CVID) (Chapter 33), suggesting that ICOS is important for protective T-cell-dependent humoral immunity.

Members of the tumor necrosis factor receptor (TNFR) family comprise another large group of costimulatory molecules.[24] OX40 (CD134), 4-1BB (CD137), herpesvirus entry mediator (HVEM), CD30, and glucocorticoid-induced TNF receptor (GITR) all have costimulatory potential. Cytoplasmic domains within these type I transmembrane proteins contain sequences that recruit a family of adaptor molecules known as TNF-receptor–associated factors (TRAFs). A well-characterized member of this group is OX40.[25] OX40 is upregulated on activated CD4 T cells after CD3/CD28 stimulation. Trimerization of OX40 induced by engagement of APC-expressed OX40L leads to recruitment of TRAFS and engagement of a survival-enhancing NF-κB pathway. OX40 deficiency leads to defects in CD4 T-cell proliferation, reduced survival of effector memory T cells, and impaired formation of effective responses to secondary T-cell stimulation with antigen. Antibody-mediated blockade results in reduced induction of experimental autoimmune encephalitis (EAE) and collagen-induced arthritis.

CTLA4 and PD-1 are coinhibitory molecules that can restrain TCR-induced activation. The CTLA4 coinhibitory molecule shares at least two features with CD28: membership in the Ig domain–containing superfamily and high-affinity binding capacity for APC-expressed ligands CD80 and CD86. Unlike CD28, CTLA4 expression is inducible in conventional CD4 T cells. However, CTLA4 is present constitutively on Treg cells. Ligation of CTLA4 induces a biochemical cascade, including the activation of PP2A and SHP-2 phosphatases.[20] CTLA4 functions to antagonize the TCR "stop signal," whereby T cells are induced to pause and maintain lengthy physical contact with antigen-bearing dendritic cells (DCs).[18] Treg can use CTLA4 to limit access of non-Treg conventional T cells to CD80/CD86 by downregulating these ligands on APCs.

CTLA4 represents one of a growing number of molecular "checkpoints" that can negatively regulate natural antitumor T-cell responses.[26] Monoclonal, antagonist anti-CTLA4 antibodies can enhance T-cell antitumor responses against many human malignancies. Approved by the US Food and Drug Administration (FDA) in 2011, such antibodies have resulted in tumor regression and durable remissions in patients with advanced metastatic melanoma and other formerly intractable malignancies.

The coinhibitor PD-1 is an Ig domain-containing surface receptor that functions as a negative regulator of T-cell activation.[26] PD-1 signals attenuation of TCR-driven PI3K and Akt. Ligation of PD-1 by cognate ligand PDL-1 results in recruitment of phosphatases, such as SHP2, to motifs within the PD-1 cytoplasmic tail. Phosphatase-associated PD-1 interacts with TCR microclusters, resulting in dephosphorylation of apical TCR signaling mediators, including ZAP-70 and CD3ζ, thus dampening TCR signal strength. At a cellular level, PD-1 enforces T-cell unresponsiveness by inhibiting TCR-mediated "stop signals" that promote stable, prolonged contact between T cells and APCs. Disruption of PD-1 in mice results in autoantibody formation and glomerulonephritis. In chronic viral infections, CD8 T cells (Chapter 25) display a hyporesponsive "exhausted" phenotype associated with elevated surface expression of PD-1.[27]

Antibody blockade of PD-1/PDL-1 interactions results in restoration of effector functions in exhausted T cells, suggesting that PD-1 manipulation could benefit patients with viral infections. The therapeutic potential of PD-1 has been most dramatically realized in the field of tumor immunotherapy.[26] In many malignancies, both CD4 and CD8 T cells accumulate within the tumor stroma. Tumor-infiltrating lymphocytes

(TILs) are characterized by high expression of PD-1 and other coinhibitory molecules. Tumor-expressed ligands for these molecules are thought to enforce a state of reduced cytolytic or cytokine-secreting functions. Monoclonal anti-PD-1 and anti-PDL-1 antagonist antibodies have been employed to reinvigorate TIL antitumor capacity. PD-1/PDL-1 blockade has sometimes resulted in dramatic tumor regressions of diverse tumors (e.g., metastatic melanoma, non–small cell lung carcinoma, and lymphoma). Other potential targets for checkpoint inhibitor therapy include the T-cell coiinhibitors Lag3 and BTLA.

Knowledge of the additive or synergistic effects of cosignaling by TCR and costimulatory molecules has been exploited for therapeutic purpose through the generation of chimeric antigen receptors (CARs).[28] These engineered molecules possess an antigen-binding extracellular region (often a single-chain variable fragment), a transmembrane domain, and cytosolic domains containing a portion of the CD3 ζ-chain fused to a costimulatory molecule (e.g., CD28, OX40, 4-1BB) cytosolic region. Cross-linking of the CARs thus may provide both "signal 1" (TCR) and "signal 2" (costimulation) for T-cell activation, leading to direct tumor killing. Cytotoxic T cells modified to express CARs that recognize CD19 have been used in successful treatment of B-cell malignancies.

Spatial and Temporal Distribution of T-Cell Receptor Signaling Proteins

Following TCR engagement, a highly organized interface termed the *immunological synapse* (IS) forms between the T cell and the APC (Chapter 6). Within the IS, many transmembrane and cytoplasmic proteins are coordinately polarized and form a structure called the *supramolecular activation cluster* (SMAC).[29] The SMAC is composed of concentric rings with a central region (cSMAC) enriched in TCR and CD3/ζ. The more peripheral ring (pSMAC) is enriched for the integrin leukocyte function–associated antigen-1 (LFA-1). Inhibitory signaling molecules, such as the transmembrane phosphatase CD45, are excluded from the SMAC. These observations initially suggested that formation of the SMAC is required for signal initiation. The subsequent finding of a paucity of active signaling molecules located in the SMAC has led to alternative models in which the SMAC is involved in directed cytokine secretion or in signal termination.

Efficient formation of the T cell IS is heavily dependent on nucleation-promoting factors, such as Wiskott-Aldrich syndrome protein (WASP), which promotes reorganization of the actin cytoskeleton.[29] Complexes composed of TCRs, active kinases, and adaptor molecules, termed *microclusters*, form at the immune synapse within seconds following TCR engagement.[30] TCR microclusters are dynamically regulated in space and time. Within 2 to 3 minutes of formation, microclusters migrate along polymerized actin from the periphery to the center of the IS, where dephosphorylation and dissociation of the components occur. Sustained TCR signaling depends on constant re-formation of microclusters containing SLP-76 and ZAP-70 at the periphery of the central zone. Costimulatory signals generated by CD28 or the integrin very late antigen-4 (VLA-4) can alter the dynamics of microcluster formation and movement and enhance T-cell activation. The functions of either the SMAC or the microclusters in TCR signal initiation and termination remain incompletely understood.

KEY CONCEPTS

Mechanisms of Tolerance: Central and Peripheral

- Central
  - Clonal deletion
  - AIRE-mediated expression of tissue-specific antigens
- Peripheral
  - Immune privilege
  - Anergy
  - Regulation

TOLERANCE

Tolerance is an inherent property of the immune system that governs the ability to respond against foreign antigens (nonself) without attacking the host (self). In normal hosts, tolerance protects against autoimmune tissue injury. The concept of immune tolerance predated understanding of the cellular and molecular bases for the phenomenon. Owen's experiments in cattle showed that a shared blood supply during early development led to lifelong immune tolerance.[31] Billingham et al. later showed that in utero inoculation with foreign tissue resulted in tolerance to foreign skin grafts applied long after birth.[32] Layered and complementary mechanisms of immune system tolerance have developed to calibrate immune responsiveness with maintenance of maximum capacity for protective activation. The remainder of this chapter discusses critical roles played by T cells in the establishment and maintenance of tolerance.

Establishment of immune tolerance to self may occur either during thymocyte development or after T-cell maturation in tissues.[33] *Central tolerance* refers to the phenomenon of induced apoptosis occurring during development of T cells in the thymus (Chapter 9). Despite the presence of mechanisms that promote expression of a comprehensive repertoire of self-peptides in the thymus, central tolerance alone is not sufficient to prevent tissue-damaging autoimmune responses. In healthy vertebrates, a small percentage of newly matured T cells exiting the thymus carry autoreactive TCRs. Control of such "escaped" autoreactive cells in secondary lymphoid organs and tissues is achieved through a series of extrathymic processes. In aggregate, these are referred to as *peripheral tolerance*.

Central Tolerance/Clonal Deletion

Techniques allowing study of clonal populations during T-cell development have opened windows into the molecular mechanisms of thymocyte clonal deletion.[34,35] Fate-mapping experiments that track specific Vβ17 TCR variable regions indicate that clonal deletion of autoreactive cells occurs principally during the transition between $CD4^+CD8^+$ double-positive "DP" to CD4 or CD8 single-positive "SP" transition during thymocyte development. TCR transgenic mice expressing a single specificity TCR potentiated additional investigation into central tolerance. Mice bearing transgenic TCRs reactive with a Y chromosome–encoded antigen (H-Y) are among the oldest models of clonal deletion. Massive deletion of developing thymocytes results in small thymi in self-antigen–expressing (male) mice; only a small number of DP thymocytes avoid the apoptotic cell death induced by TCR engagement. In contrast, transgenic T cells develop normally in littermate female mice that lack H-Y antigen. Additional TCR transgenic models of central tolerance

have shown that autoreactive thymocyte deletion can occur before, during, or after the DP stage.

Deletion due to self-reactivity implies that the extent of autoreactivity of the developing thymocytes is systematically calibrated. A long-held model is that proclivity for deletion due to self-reactivity is a function of the TCR strength of signal.[36] In this model, TCR signals that exceed an intensity threshold lead to clonal deletion of cells bearing that antigen receptor. This model is supported by studies of animals in which the signaling machinery has been genetically altered to increase or decrease TCR signal intensity. For example, an increase in the number of TCR-associated ITAMs (presumably leading to increases in downstream signals) enhances clonal deletion in a TCR transgenic system.[37]

For clonal deletion to establish comprehensive self-tolerance, T cells must come into contact with all potential self-antigens during thymocyte maturation. It is straightforward to imagine how deletion of cells reactive to MHC or to other ubiquitously expressed protein antigens can occur in the thymus. However, thymocytes bearing TCR reactive with self-antigens whose expression is restricted to a specific tissue or developmental time point (e.g., peptides expressed only in the pancreas or testes) would not be expected to encounter such antigens during thymic maturation.

In recent years, mechanisms whereby clonal deletion of developing thymocytes reactive with such tissue-restricted antigens (TRAs) have been described. Some evidence supports the possibility that TRAs are transported to the thymus by APCs, such as migratory DCs.[36] Another, non–mutually exclusive model holds that subsets of thymus-resident APCs may "ectopically" express TRAs. In this regard, the transcription factor AIRE (autoimmune regulator) controls TRA expression within medullary thymic epithelial cells (mTECs).[38] AIRE expression ectopically drives mTEC to express peptides from open reading frames representing TRA. Self-reactive developing T cells are thus exposed to TRA in the thymus and are deleted.

AIRE-deficient humans and mice develop tissue damaging autoimmunity in a syndrome termed polyendocrinopathy–candidiasis–ectodermal dystrophy (APS-1 or APECED in humans; Chapter 34), emphasizing the role of central tolerance in preventing T-cell–mediated autoimmune injury.

TRA presentation by mTEC is unlikely to fully account for successful clonal deletion, given the relative rarity of mTEC in thymic stroma relative to the abundance of autoreactive thymocytes. An additional mechanism for enhancing TRA-reactive thymocytes exposure to deleting TRA-driven signals is "shared" antigen presentation capacity among thymic stromal cells other than mTECS. Intercellular transfer of TRA between mTEC and thymic-resident or migrant DCs has been observed, supporting this model.[36]

Peripheral Mechanisms of Tolerance

Despite highly efficient negative selection in the thymus, significant numbers of autoreactive T cells escape to the periphery, and have potential to cause autoimmune tissue injury. Accordingly, additional mechanisms have evolved to control autoreactivity of T cells that have managed to escape the thymus. These include immune privilege, anergy, and regulation.

Immune Privilege

Medawar first proposed the concept of "immune privilege," whereby intrinsic factors in local tissue engender resistance to autoimmune tissue damage. Tissues that possess features of immune privilege include the anterior chamber of the eye, the brain, and the developing fetus in pregnant females.[39] The eye and the brain are critical organs for basic survival functions and yet have a limited capacity for regeneration. Thus uncontrolled immune responses in these organs could have a detrimental effect on survival. A fetus expresses MHC derived from both parents; thus the mother's immune system must develop tolerance of the paternal antigen-bearing fetus to prevent pregnancy loss.

Immune-privileged tissues evade or suppress immune effector functions through multiple mechanisms. Cells of the eye, brain, and fetal villous trophoblast display low level or absent surface expression of classic MHC class Ia proteins. This feature likely protects them from cytotoxic T-lymphocyte (CTL)–mediated lysis (Chapter 12). Ocular cells express proapoptotic cell surface molecules, such as CD95 ligand (CD95L) and TRAIL (TNF-related apoptosis inducing ligand) (Chapter 17). Since activated T cells or other inflammatory cells may bear high levels of cognate ligands (e.g., CD95, other "death" receptors) for the proapoptotic proteins, they can be induced to undergo apoptosis upon migration to immune-privileged sites, thus limiting their potential for tissue injury.

In mice, the presence of CD95L on ocular cells is critical for the acceptance of corneal allografts. Soluble factors likely also contribute to immune privilege. Ocular DCs can elaborate molecules including the cytokines transforming growth factor-β (TGF-β) and IL-10, which may lead to induction and/or recruitment of Tregs. In addition, ocular DCs may produce high levels of indoleamine oxidase, an enzyme that supports Treg differentiation. Finally, ocular cells can produce migration inhibitory factor (MIF) that suppresses natural killer (NK) cell–dependent cytolytic capacity.

T-Cell Anergy

Cellular proliferation and/or potentiation of T-cell effector function are not inevitable consequences of TCR engagement. Under some conditions, TCR engagement results in *anergy*, a cellular fate characterized by reduced proliferation potential and blunted cytokine production in response to subsequent TCR stimulation.[40] T-cell anergy can ensue either when the TCR is engaged without concomitant costimulation (e.g., CD28 ligands) or when the ligand for the TCR does not possess sufficient affinity to initiate the full spectrum of biochemical second messengers.

Anergy is observed in T cells stimulated with metabolically inactivated APCs incapable of providing costimulation. Conversely, CD28 costimulation prevents the induction of anergy. Treatment of cultured T cells with IL-2 can overcome the anergic state in vitro. In addition to TCR and costimulatory signals, environmental cues (e.g., nutrient and energy store availability) and the products of Tregs (described below) also control the anergy/activation fate choice.

T-cell anergy can be induced by stimuli that engage the calcineurin/NFAT pathway without causing concomitant increases in RAS/ERK pathway–dependent activating protein-1 (AP-1) transcription factor activity (Fig. 10.6A). Experimentally, relatively unopposed NFAT activity can be induced by treatment with calcium ionophore or by stimulation through the TCR while blocking CD28 costimulation. Concurrent treatment of TCR-stimulated cells with protein synthesis inhibitors or with NFAT pathway inhibitors, such as cyclosporine, abrogates the

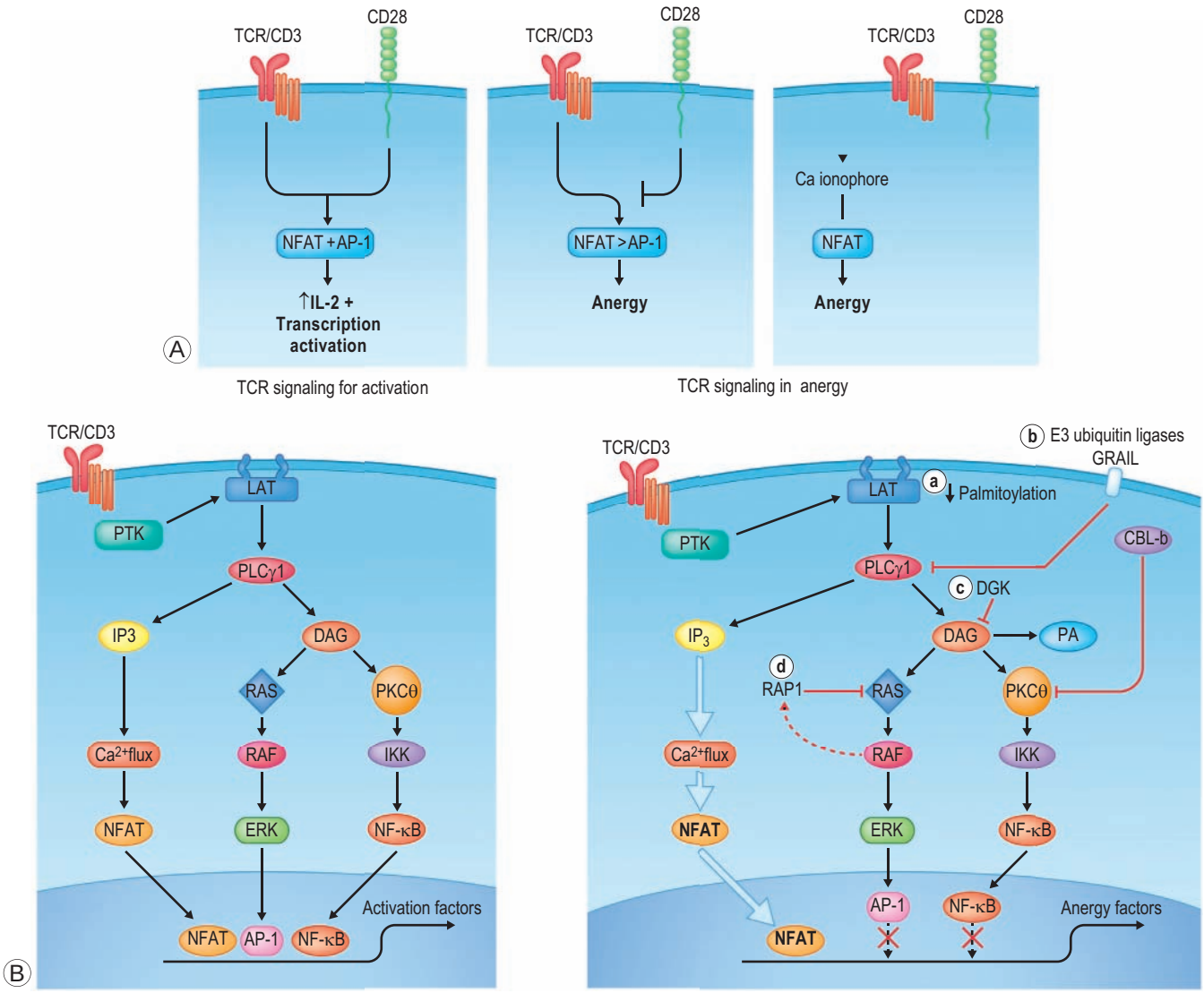


FIG. 10.6 T-cell Anergy Induction and Maintenance Correlates with Differential Activation of T-cell Receptor *(TCR)*–Dependent Second-Messenger Signaling Cascades. (A) Stimulation of T cells by cross-linking of both TCR/CD3 and CD28 lead to upregulation of both the nuclear factor of activated T cells *(NFAT)* and activating protein-1 *(AP-1)* transcription factors, leading to increased transcription of the interleukin-2 *(IL-2)* gene and activation. An imbalance of activated NFAT and AP-1 by blockade of CD28 signals *(middle panel)* or calcium ionophore *(right panel)* leads to an anergic phenotype. (B) TCR signaling required for full T-cell activation features calcium flux-, RAS-, and PKC-dependent biochemical events leading to cooperative transcriptional regulation by NFAT, AP-1, and NF-κB transcription factors *(left panel)*. In energized cells, TCR-dependent signaling is differentially impaired (DAG-dependent events more so than calcium-dependent events) through multiple mechanisms: (a) decreased palmitoylation of linker of activated T cells *(LAT)* results in diminished recruitment to the immunological synapse; (b) upregulated gene related to anergy in lymphocytes *(GRAIL)* and CBL-b degrade positive signaling regulators PLCg1 and PKCθ; (c) diacylglycerol kinases *(DGKs)* convert DAG (PKC and RAS activator) to phosphatidic acid *(PA)*; (d) active RAP1 recruits RAF, thus preventing RAS-mediated signaling to extracellular signal-regulated kinase *(ERK)*. Anergy mechanisms and mediators are highlighted in *red*.

development of later unresponsiveness, supporting the notion that anergy induction and maintenance depend on actively transcribed and translated anergy-associated factors.[41]

Increased NFAT pathway signaling coupled and reduced Ras pathway function in T-cell anergy correlates with a number of biochemical and gene regulatory events (see Fig. 10.6B).[42] First, preferential FYN kinase–dependent activation of c-CBL within anergized cells leads to recruitment of the RAF kinase to the nucleotide exchange factor RAP1. The RAF-RAP1 association prevents RAF recruitment to RAS and thus restricts ERK pathway activation. Second, upregulation of NFAT-dependent transcription factors EGR2 and EGR3 results in transactivation of proteins implicated in the restraint of T-cell activation, including GRAIL (gene related to anergy in lymphocytes), CBL-b, and ITCH.[41] E3 ubiquitin ligase activity associated with a number of the latter factors is responsible for ubiquitin-mediated proteolysis of TCR signal–promoting molecules, such as PKCθ and RAS-GRP. Third, anergic stimuli cause reduced palmitoylation and phosphorylation of the positive TCR signaling mediator LAT, which may reduce LAT membrane

localization in detergent-insoluble lipid rafts and participation in activating signaling.[43] Fourth, increased expression of members of the diacylglycerol kinase family (DGKα and DGKζ) is observed in T-cell anergy. Augmented DGK expression reflects increased capacity for phosphorylation-dependent conversion of DAG, the key lipid mediator upstream of RAS signaling, into phosphatidic acid (PA), an inert metabolite.[44]

Cellular sensing of adequate nutrient and energy stores required for optimal differentiation and proliferation also regulates T-cell fate decisions between anergy and activation. Antagonists of leucine or glucose can cause T-cell anergy if administered in the context of normally activating combined antigen receptor and costimulatory signals, suggesting the importance of amino acid- and energy-sensing pathways in cell fate determination.[40] Several lines of evidence suggest that the cytosolic molecule mechanistic target of rapamycin (mTOR) functions as a switch between activation and anergic states by integrating signals from antigen recognition (TCR), immune (CD28, IL-2 receptor), and metabolic (e.g., GLUT1) receptors and sensors.[45]

TCR and CD28 ligation induce T-cell anergy rather than activation when given in the presence of rapamycin, a selective mTOR inhibitor, suggesting that optimal T-cell activation and anergy avoidance requires activation of mTOR. mTOR is activated by signals that communicate abundant nutrients (e.g., leucine-stimulated RAG proteins). Conversely, mTOR is inhibited by AMP-activated protein kinase (AMPK), an enzyme that responds to low-energy stores reflected by increased ratios of adenosine monophosphate (AMP) to adenosine triphosphate (ATP). Downstream effectors of activated mTOR such as AKT promote cell cycle entry and prevent transcriptional activation of the anergy factors GRAIL and CBL-b.

KEY CONCEPTS

Immunomodulatory Agents That Affect T-Cell Signaling

- Immunosuppressants:
  - Cyclosporine, tacrolimus, and anti-TCR antibodies inhibit TCR-generated signals
  - CTLA4-Ig blocks CD28 signals
  - Rapamycin inhibits mechanistic target of rapamycin (mTOR) activation
- Immunostimulants work through inhibitory receptor blockade, enhancing T-cell antitumor function
  - Anti-CTLA4 antagonist antibodies permit CD28-CD80/86 interactions
  - Anti-PD1/anti-PDL1 antibodies prevent suppressive interactions between tumor-expressed PDL1 and PD1 on tumor-infiltrating Treg and effector T cells

Regulation

Subsets of T lymphocytes can enforce peripheral tolerance through regulation of autoreactive immune responses. Tregs can suppress effector cells of both myeloid and lymphoid lineages (Chapter 13). The most extensively studied Treg-sensitive immune functions—and arguably the most important for maintenance of T-cell tolerance—are proliferation and cytokine production by conventional T cells (Tconv). Tregs inhibit these processes through both cell contact–dependent and soluble molecule secretion mechanisms.[47] By regulating the activation and proliferation of antigen-specific effectors, Tregs promote immune self-tolerance and suppress autoimmunity in vivo. Treg manipulation for therapeutic purposes holds promise for treatment of human autoimmune and malignant disease.

Tregs were initially described on the basis of the correlation between high CD25 expression and the potent suppressive activity of a subset of CD4 T cells. Most Tregs also express GITR, CD103, CTLA-4, lymphocyte activation gene-3 (*LAG-3*), and low levels of CD45RB.[46] No single surface marker specifically identifies Treg. However, features that together correlate tightly with Treg suppressive capacity and a unique differentiation pathway include CD4 T-cell expression of the X chromosome-encoded transcription factor Foxp3, an epigenetic DNA methylation signature that prominently suppresses the *IL2* gene and drives high constitutive CTLA4 expression, and a TCR-driven, antigen-primed activation state.[47]

Absence of Foxp3 occurs either via spontaneous mutation (exemplified by the scurfy mouse, which develops fatal autoimmune disease) or through targeted disruption of the gene, which leads to the complete loss of T cells with regulatory activity. Conditional deletion of Foxp3 in peripheral T cells results in loss of the suppressive phenotype. Conversely, overexpression of Foxp3 leads to an excess of T cells with regulatory activity. Thus, Foxp3 appears both necessary and sufficient for Treg suppressive functions.

The bulk of Foxp3-positive Tregs develop in the thymus. Dynamic Foxp3 expression and development of suppressive capacity can be observed in naïve peripheral CD4 T cells after exposure to TGF-β or retinoic acid. Recent work strongly suggests that inducible Treg may have nonredundant functions in suppressing chronic inflammation.

In humans, mutations in FOXP3 account for a majority of cases of immune dysfunction/polyendocrinopathy/enteropathy/X-linked (IPEX) syndrome.[47] Affected human males develop an autoimmune syndrome consisting of lymphoproliferation, thyroiditis, insulin-dependent diabetes mellitus, enteropathy, and other immune disorders. The signs and symptoms of FOXP3 deficiency are similar in mice and humans.

Mechanisms whereby Treg inhibit conventional T-cell responses include secretion of suppressive cytokines, induction of T-cell apoptosis, and repression of APC function.[48] Key Treg-secreted cytokines include IL-10, TGF-β, and IL-35; each of these molecules has the capacity to induce cell cycle arrest. Tregs, via high expression of CD25 (IL-2Rα), may compete with neighboring T effector cells for limited supplies of IL-2. The resulting growth factor deprivation results in apoptosis of proliferating T cells in a B-cell lymphoma 2–interacting (Bcl2-interacting) mediator (BIM)-dependent manner. Tregs preferentially express galectin-1, a β-galactoside-binding protein that ligates CD45 and other lymphocyte surface molecules and that may suppress lymphocyte activation through induction of cell cycle arrest. Treg surface proteins that work to effectively decrease efficiency of antigen presentation by DCs include CTLA-4, which interferes with CD28-dependent T-cell costimulation; LAG3, which binds to MHC class II molecules and prevents DC maturation; and Nrp1, which mediates prolonged Treg/DC interactions that may restrict access of effector T cells to DCs.

Growing insight into the molecular bases of Treg development and of their suppressive capacities has been exploited in recent manipulations of Treg for therapeutic purpose.[47] Conceptually, strategies to boost Treg hold promise for treatment of human autoimmune and inflammatory diseases, whereas strategies that selectively deplete Treg or impair Treg suppressive capac-

ity may lead to improved efficacy of immunotherapy for malignancy. Supplemental interleukin-2 (IL-2) can selectively drive expansion of Treg in vivo in patients with graft-versus-host disease (GvHD), type 1 diabetes, and hepatitis C-induced vasculitis, with the expectation of ameliorating inflammatory tissue injury. Adoptive cell therapy is another Treg-based approach to autoimmune disease. Tregs isolated from circulation can be expanded in vitro and then transferred back to transplant recipients for treatment of rejection reactions or GvHD prevention.

Tregs can apparently hinder naturally occurring antitumor immunity. Tumor-infiltrating FOXP3$^+$ T cells are in a proliferative activated state, express more T-cell activation molecules compared with Tregs in noncancerous tissues, and confer worse prognosis when present in high numbers.

Treg depletion may be a major mechanism by which ipilimumab (anti-CTLA4 antibody) works, in addition to blocking inhibitory signaling in conventional T cells. Tumor-resident CTLA4-high Treg are opsonized by ipilimumab and rendered susceptible to selective antibody-dependent cellular toxicity. Resultant Treg depletion correlates with greater cytotoxic function of tumor-reactive CD8 T cells, and improved tumor control. Opposite to the effect of anti-CTLA4 treatment, anti-PD-1 therapy may promote Treg function, thus diminishing efficiency of the checkpoint inhibition therapy. Tumor-infiltrating Treg express PD-1 highly, and PD-1-blocking antibody could act to remove restraints on Treg proliferation and suppressive function.

Unwanted enhancement of Treg function may help explain a hyperaggressive tumor phenotype observed in a subset of cancer patients treated with anti-PD-1. Tumor infiltration by enhanced numbers of proliferating and active Tregs is observed in such patients. Current studies are testing whether Treg-depletion strategies, used in combination with or prior to anti-PD-1 therapy, could enhance effectiveness of the checkpoint inhibitor.

ON THE HORIZON

- Promising targets for manipulation of T-cell signaling to treat malignancy include:
  - SYK inhibition
  - PI3K/AKT inhibition
  - Tec kinase inhibition in cancer
- Targets for relief of immune checkpoints on T-cell activation for immune enhancement in cancer or infection include:
  - LAG3, TIGIT, TIM3, BTLA, VISTA
- Harnessing Treg for GvHD, transplant rejection, autoimmunity
  - Cytokine (e.g., IL-2)-driven Treg expansion in vivo
  - Adoptive transfer of ex vivo expanded Treg

SUMMARY AND FUTURE DIRECTIONS

Antigen receptor signaling controls key T-cell developmental checkpoints as well as the fate decisions and behavior of mature T cells. Early TCR signaling depends on tyrosine phosphorylation events coupled to "second messenger" cascades through adaptor proteins. TCR affinity for ligand is the major determinant for life-or-death decisions by immature thymocytes that result in establishment of a broadly self-tolerant mature T-cell repertoire that includes regulatory cells (Treg). In peripheral T cells, auxiliary signals from costimulatory molecules, coreceptors, growth factor receptors, and metabolic sensors all shape the transcriptional regulation programs that govern cell function. T cells proliferate and produce cytokines when members of the CD28 or TNFR families of receptors signal in tandem with the TCR. Immunoreceptor and costimulatory signals are also critical in maintaining immune tolerance through clonal deletion, anergy, and peripheral suppression by Tregs. Co-inhibitory signals, such as CTLA4 and PD1, that restrain T-cell antitumor functions may be antagonized for immunotherapy of cancer. More detailed understanding of T-cell intracellular signaling pathways will likely lead to more specific, targeted immunomodulatory therapies.

REFERENCES

1. Chakraborty AK, Weiss A. Insights into the initiation of TCR signaling. *Nat Immunol*. 2014;15(9):798–807.
2. Guy CS, Vignali DA. Organization of proximal signal initiation at the TCR:CD3 complex. *Immunol Rev*. 2009;232:7–21.
3. Gaud G, Lesourne R, Love PE. Regulatory mechanisms in T cell receptor signalling. *Nat Rev Immunol*. 2018;18(8):485–497.
4. Rhee I, Veillette A. Protein tyrosine phosphatases in lymphocyte activation and autoimmunity. *Nat Immunol*. 2012;13:439–447.
5. Notarangelo LD. Immunodeficiency and immune dysregulation associated with proximal defects of T cell receptor signaling. *Curr Opin Immunol*. 2014;31:97–101.
6. Wang X, Hills LB, Huang YH. Lipid and protein co-regulation of PI3K effectors Akt and Itk in lymphocytes. *Front Immunol*. 2015;6:117.
7. Smith-Garvin JE, Koretzky GA, Jordan MS. T cell activation. *Annu Rev Immunol*. 2009;27:591–619.
8. Trebak M, Kinet JP. Calcium signalling in T cells. *Nat Rev Immunol*. 2019;19(3):154–169.
9. Krishna S, Zhong XP. Regulation of lipid signaling by diacylglycerol kinases during T cell development and function. *Front Immunol*. 2013;4:178.
10. Burbach BJ, Medeiros RB, Mueller KL, et al. T cell receptor signaling to integrins. *Immunol Rev*. 2007;218:65–81.
11. Salzer E, Cagdas D, Hons M, et al. RASGRP1 deficiency causes immunodeficiency with impaired cytoskeletal dynamics. *Nat Immunol*. 2016;17(12):1352–1360.
12. So T, Croft M. Regulation of PI-3-Kinase and Akt signaling in T lymphocytes and other cells by TNFR family molecules. *Front Immunol*. 2013;4:139.
13. Lim EL, Okkenhaug K. Phosphoinositide 3-kinase δ is a regulatory T cell target in cancer immunotherapy. *Immunology*. 2019;157(3):210–218.
14. Shi JH, Sun SC. TCR signaling to NF-kappaB and mTORC1: expanding roles of the CARMA1 complex. *Mol Immunol*. 2015;68:546–557.
15. Jordan MS, Singer AL, Koretzky GA. Adaptors as central mediators of signal transduction in immune cells. *Nat Immunol*. 2003;4:110–116.
16. Balagopalan L, Kortum RL, Coussens NP, et al. The linker for activation of T cells (LAT) signaling hub: from signaling complexes to microclusters. *J Biol Chem*. 2015;290:26422–26429.
17. Simeoni L, Lindquist JA, Smida M, et al. Control of lymphocyte development and activation by negative regulatory transmembrane adapter proteins. *Immunol Rev*. 2008;224:215–228.
18. Sharpe AH. Mechanisms of costimulation. *Immunol Rev*. 2009;229(1):5–11.
19. Artyomov MN, Lis M, Devadas S, et al. CD4 and CD8 binding to MHC molecules primarily acts to enhance Lck delivery. *Proc Natl Acad Sci USA*. 2010;107:16916–16921.
20. Baumeister SH, Freeman GJ, Dranoff G, et al. Coinhibitory pathways in immunotherapy for cancer. *Annu Rev Immunol*. 2016;34:539–573.
21. Esensten JH, Helou YA, Chopra G, et al. CD28 costimulation: from mechanism to therapy. *Immunity*. 2016;44(5):973–988.
22. Linterman MA, Denton AE, Divekar DP, et al. CD28 expression is required after T cell priming for helper T cell responses and protective immunity to infection. *Elife*. 2014:3.
23. Zhang R, Huynh A, Whitcher G, et al. An obligate cell-intrinsic function for CD28 in Tregs. *J Clin Invest*. 2013;123:580–593.
24. Ward-Kavanagh LK, Lin WW, Šedý JR, Ware CF. The TNF receptor superfamily in co-stimulating and co-inhibitory responses. *Immunity*. 2016;44(5):1005–1019.

25. Croft M. Control of immunity by the TNFR-related molecule OX40 (CD134). *Annu Rev Immunol*. 2010;28:57–78.
26. Wei SC, Duffy CR, Allison JP. Fundamental mechanisms of immune checkpoint blockade therapy. *Cancer Discov*. 2018;8(9):1069–1086.
27. Wherry EJ, Kurachi M. Molecular and cellular insights into T cell exhaustion. *Nat Rev Immunol*. 2015;15:486–499.
28. June CH, O'Connor RS, Kawalekar OU, et al. CAR T cell immunotherapy for human cancer. *Science*. 2018;359(6382):1361–1365.
29. Kumari S, Curado S, Mayya V, Dustin ML. T cell antigen receptor activation and actin cytoskeleton remodeling. *Biochim Biophys Acta*. 2014;1838(2):546–556.
30. Hashimoto-Tane A, Saito T. Dynamic regulation of TCR-microclusters and the microsynapse for T cell activation. *Front Immunol*. 2016;7:255.
31. Owen RD. Immunogenetic consequences of vascular anastomoses between bovine twins. *Science*. 1945;102:400–401.
32. Billingham RE, Brent L, Medawar PB. Actively acquired tolerance of foreign cells. *Nature*. 1953;172:603–606.
33. Bluestone JA, Bour-Jordan H, Cheng M, et al. T cells in the control of organ-specific autoimmunity. *J Clin Invest*. 2015;125:2250–2260.
34. von Boehmer H, Kisielow P. Negative selection of the T cell repertoire: where and when does it occur? *Immunol Rev*. 2006;209:284–289.
35. Xing Y, Hogquist KA. T cell tolerance: central and peripheral. *Cold Spring Harb Perspect Biol*. 2012;4(6).
36. Klein L, Kyewski B, Allen PM, et al. Positive and negative selection of the T cell repertoire: what thymocytes see (and don't see). *Nat Rev Immunol*. 2014;14:377–391.
37. Love PE, Hayes SM. ITAM-mediated signaling by the T cell antigen receptor. *Cold Spring Harb Perspect Biol*. 2010;2:a002485.
38. Taniguchi RT, Anderson MS. The role of Aire in clonal selection. *Immunol Cell Biol*. 2011;89:40–44.
39. Taylor AW. Ocular immune privilege and transplantation. *Front Immunol*. 2016;7:37.
40. Chappert P, Schwartz RH. Induction of T cell anergy: integration of environmental cues and infectious tolerance. *Curr Opin Immunol*. 2010;22:552–559.
41. Mueller DL. Mechanisms maintaining peripheral tolerance. *Nat Immunol*. 2010;11:21–27.
42. Valdor R, Macian F. Induction and stability of the anergic phenotype in T cells. *Semin Immunol*. 2013;25:313–320.
43. Hundt M, Tabata H, Jeon MS, et al. Impaired activation and localization of LAT in anergic T cells as a consequence of a selective palmitoylation defect. *Immunity*. 2006;24:513–522.
44. Zhong XP, Guo R, Zhou H, et al. Diacylglycerol kinases in immune cell function and self-tolerance. *Immunol Rev*. 2008;224:249–264.
45. Myers DR, Wheeler B, Roose JP. mTOR and other effector kinase signals that impact T cell function and activity. *Immunol Rev*. 2019;291(1).
46. Shevach EM, Thornton AM. tTregs, pTregs, and iTregs: similarities and differences. *Immunol Rev*. 2014;259:88–102.
47. Sakaguchi S, Norihisa Mikami N, Wing JB, et al. Regulatory T cells and human disease. *Ann Rev of Immunol*. 2020;38(1):541–566.
48. Plitas G, Rudensky AY. Regulatory T cells: differentiation and function. *Cancer Immunol Res*. 2016;4(9):721–725.

11

Helper T-Cell Subsets and Control of the Inflammatory Response

Todd N. Eagar and Stephen D. Miller

CD4 T cells are an integral part of the immune system, and key regulators of many inflammatory processes involved in health and disease including responses against infectious organisms, anti-cancer immunity, and autoimmunity. Disruption of normal T-cell functions through genetic disease (Chapter 34), such as severe combined immunodeficiency (SCID), or immunosuppressive therapies often results in recurrent bacterial, fungal, and viral infections and certain cancers.[1] Conversely, dysregulated T-cell responses are associated with rampant inflammation and autoimmunity, as seen in patients with such conditions as autoimmune lymphoproliferative syndrome (ALPS), autoimmune polyendocrinopathy syndrome type 1 (APS-1) and immune dysregulation, polyendocrinopathy, enteropathy, X-linked (IPEX) syndrome.[2]

Thus a balance must be struck between allowing T-cell inflammation to protect the host from pathogens and ensuring control mechanisms to prevent injury from unchecked T-cell–mediated inflammation. The requisite control is established as T cells transit through distinct checkpoints of the immune response. These include activation, clonal expansion, migration, differentiation, and response termination.

T cells are generated in the thymus from lymphocyte progenitor cells (Chapter 9). Mature T cells enter the peripheral immune system as inexperienced or naïve cells that are incapable of carrying out effector immune responses. To participate in inflammation, T cells must go through an instructional process with defined checkpoints. This chapter discusses the functional heterogeneity within the CD4 T-cell compartment in the context of the checkpoints involved in the generation and regulation of T-cell biology.

KEY CONCEPTS

T-Cell–Mediated Inflammation Requires Activated or Memory T Cells

Naïve T Cells
- Low frequency
- Traffic through circulatory and lymphatic systems
- Require professional antigen-presenting cells (APCs) for activation
- Require strong costimulation
- Delayed expansion
- Delayed cytokine production
- Dependent on interleukin-7 (IL-7)

Activated T Cells
- Traffic through most tissues
- Respond to antigen presented by nonprofessional APCs
- Require less costimulation
- Rapidly expand following antigen encounter
- Rapidly produce effector cytokines
- Dependent on IL-2

Memory T Cells
- Traffic through most tissues
- Dependent on IL-7 and IL-15
- Respond to antigen presented by nonprofessional APCs
- Require less costimulation
- Rapidly expand following antigen encounter
- **Activation:** T-cell function requires antigen-specific T-cell receptor (TCR) signals and accessory signals. This provides fine specificity against discrete protein antigens and limits self-reactivity.
- **Clonal expansion:** Activated T cells rapidly expand in number to scale the immune response.
- **Migration:** T cells traffic through blood, lymphatics, and organs. This allows for surveillance and localized effector responses in diverse tissue types.
- **Effector responses:** T cells adapt their phenotypes to tailor the response to the organism and infected tissue. Long-term memory cells provide protection from re-exposure to the same pathogens.
- **Termination:** T-cell responses are limited through inhibitory receptors, cell death pathways, and suppression.

ACTIVATION

T-cell activation is the process by which resting T cells become functional. Due to the fact that activation is required at multiple stages of a T-cell response, activation is the most critical checkpoint controlling T-cell activity (Fig. 11.1). Full activation requires the T cell to integrate signals through distinct receptor types including the T-cell antigen receptor (TCR, signal 1), costimulatory receptors (signal 2), and cytokines (signal 3).[3] TCR signals are essential for T-cell activation. Each TCR is generated through somatic recombination and is selected to recognize linear peptide antigens when presented in the context of self-human leukocyte antigen (HLA) proteins (Chapter 5). Engagement of cognate HLA–peptide complex by the TCR elicits biochemical signal cascades and induces gene transcription that ultimately control T-cell function (Chapter 10). The second set of signals involve costimulatory receptors. Positive costimulation is delivered through receptors such as CD28 and inducible costimulator (ICOS) in response to binding to B7 (CD80 or CD86) or ICOSL, respectively, on neighboring cells. Negative or inhibitory costimulatory signals can also be received by the T cell through receptors such as cytotoxic T-lymphocyte-associated antigen-4 (CTLA-4) or programmed death-1 (PD-1) by engagement of B7 or PD-L1. Costimulation and cytokines regulate T-cell activation by quantitatively modifying TCR-induced signals or qualitatively by inducing

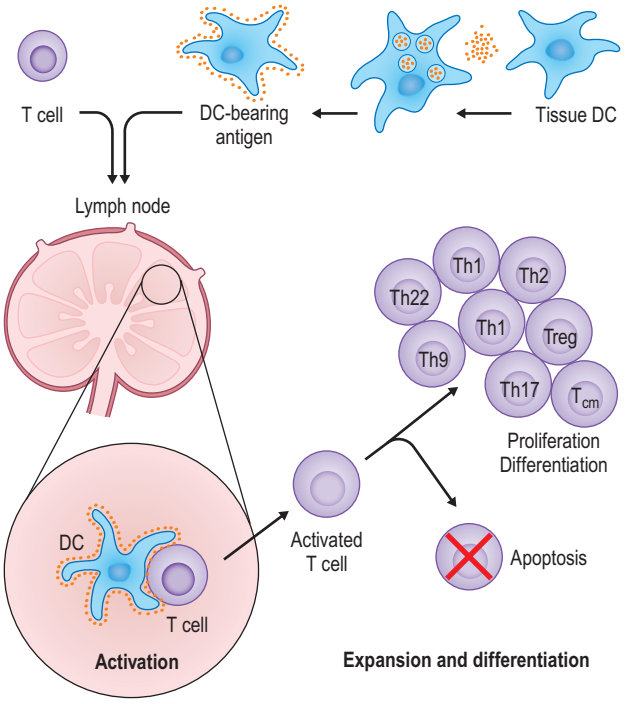


FIG. 11.1 T-Cell Expansion and Differentiation. Following emigration from the thymus, CD4 T cells travel through the lymph nodes. Inflammation results in the recruitment of tissue dendritic cells *(DCs)* into the draining lymph nodes. In the lymph nodes, DCs present antigen to naïve T cells. Effective antigen presentation stimulates the activation, proliferation, and differentiation of T cells into Th1, Th2, Th9, Th17, Th22, or uncommitted T central memory (T_{cm}) cells.

distinct biochemical cascades. Ultimately, the effects of costimulation and cytokine signals are detected at the level of gene expression where they may promote, inhibit, or otherwise alter TCR-induced cell activation.[4]

TCR costimulation and cytokine signals are delivered during interactions between T cells and APCs (Chapter 6) including dendritic cells (DC) and B cells. DCs are considered the primary APC involved in T-cell activation.[5] Mature DCs are potent stimulators of naïve T cells in the lymph nodes and strong activators of memory and effector cell subsets in tissues. In response to infection or damage, DCs migrate from the tissue into the T-cell–rich areas of regional lymph nodes.[6] T cells interact with numerous DCs as they scan for cognate antigen. In the absence of TCR signaling, these interactions are of short duration; however, TCR engagement elicits prolonged T cell–DC engagement leading to full activation.[7] Incomplete stimulation through the TCR, costimulation, or cytokine pathways may prevent T cells from participating in inflammation through altered effector differentiation, induction of apoptosis, or causing functional unresponsiveness such as anergy or exhaustion. Thus, the requirement for TCR stimulation dictates specificity of T-cell responses, and the need for multiple signals during T-cell activation prevents inappropriate T-cell activation. T cells also interact with DCs in non-lymphatic tissues and at sites of inflammation. During this interaction, T cells become activated to elicit effector functions.[8]

Defects in T-cell activation pathways contribute to several immunodeficiency diseases. Patients with mutations in the proximal TCR signaling components lck and Zap70 show defective T-cell activation. T cells from patients with Wiskott-Aldrich syndrome have defects in the formation of stable TCR signaling complexes. T cells in leukocyte adhesion deficiency 1 have a reduced activation capacity due to defects in integrin β_2. By contrast, mutations in inhibitory receptors, including CTLA-4 and PD-1, are thought to contribute to susceptibility in autoimmune diseases.

The T-cell activation process can be targeted for therapeutic immunosuppression. Pharmacological targeting of key biochemical pathways required for activation has been the primary approach to prevent rejection in solid organ transplantation, to treat graft-versus-host disease (GvHD) in bone marrow transplantation, and to reduce the severity of autoimmune diseases or other immune-mediated disorders. Cyclosporine and tacrolimus block calcineurin and thereby inhibit TCR-induced nuclear factor of activated T cells (NFATs) signals. Rapamycin inhibits the kinase mammalian target of rapamycin (mTOR), which is induced by TCR and CD28 or by IL-2 receptor signals.[9] A fusion of the extracellular domain of CTLA-4 with the Fc portion of immunoglobulin G (IgG; CTLA-4-Ig) prevents T-cell activation by blocking the interactions between CD28 and its B7 ligands. Antibodies against the IL-2R block IL-2–driven T-cell activation. Conversely, antibodies blocking the inhibitory receptors PD-1 and CTLA-4 promote or prolong T-cell activation and thereby enhance antitumor T-cell responses.[10]

CLONAL EXPANSION

T-cell activation occurs within hours of antigen exposure; however, days are required until T-cell responses can be detected from a naïve host. This delay partially reflects the need for T-cell expansion to produce a detectable and effective response. The initial frequency of T cells for a single epitope is estimated at approximately one per million total CD4 T cells or approximately 1×10^5 cells per human.[11] It takes several days following antigen exposure for the first cell divisions to occur in the lymph nodes and the spleen; thereafter cell numbers increase rapidly between days 2 and 7, when many T cells have undergone as many as eight rounds of division. This process is known as *clonal expansion* (see Fig. 11.1). A single precursor T cell following eight rounds of division can yield 256 daughter cells, each possessing the same TCR. Through the combined proliferative responses of multiple T-cell clones to the same or different epitopes produced by the same pathogen, a rapid net increase in pathogen-reactive T cells occurs.

In the clinical setting, T-cell proliferation is a target for immunosuppression. Some drugs, such as azathioprine, methotrexate, and mycophenolate mofetil impair T-cell proliferation by inhibiting nucleotide synthesis. Cyclophosphamide suppresses DNA replication by alkylating guanine residues. These agents are used in the treatment of various autoimmune diseases and for the prevention of graft loss in transplantation.

TRAFFICKING

Following activation, T cells change their migration patterns (Chapter 16). T-cell migration relies on the actions of selectins, chemokines (Chapter 15), integrins, and matrix proteases. Together, these permit vessel transmigration into lymphatic tissues or sites of inflammation.[12] The differential trafficking of naïve and effector T cells results from upregulated expression of selectins, chemokine receptors, and integrins upon activation. The circulation of naïve T cells is restricted to blood, lymph nodes, and lymphatic ducts. This pattern of movement relies on the expression of chemokine receptor 7 (CCR7), L-selectin

(CD62L), and leukocyte function–associated antigen-1 (LFA-1; $\alpha_L\beta_2$ integrin, CD11a, CD18). CCR7-expressing T cells follow gradients of CCL21 into T-cell–rich areas in the lymph nodes, Peyer patches, and spleen. L-selectin allows rolling on the high endothelial venule (HEV). LFA-1 and intercellular adhesion molecule-1 (ICAM-1) permit firm adhesion and extravasation. Migration to the Peyer patches also requires $\alpha_4\beta_7$ integrin in addition to LFA-1. T-cell retention in and egress from lymph nodes is regulated by sphingosine 1 phosphate (S1P), a secreted phospholipid. S1P sensitivity is mediated by sphingosine 1 phosphate receptor 1 (S1P1), a G protein–coupled receptor. Recirculating T cells downregulate S1P1. As a result of low S1P levels in the lymph node, S1P1 expression allows exit from the lymph nodes 12 to 18 hours after entry.[13]

During activation, T cells alter the patterns of homing receptors to reduce the cell's ability to migrate within the secondary lymphatic system while gaining receptors needed for trafficking in the peripheral tissues. Very early in the activation process, T cells downregulate S1P1 and express CD69, which suppresses chemotactic migration in response to S1P, shed L-selectin, and express the ligands for E- and P-selectin. Additionally, T cells downregulate CCR7 expression and upregulate expression of a new set of chemokine receptors. Following activation, expression of LFA-1 and additional integrin molecules is markedly increased. One such integrin, very late antigen-4 (VLA-4), is induced by activation and permits T-cell migration into the central nervous system (CNS), lungs, and intestines.[14]

Differential expression of homing receptors by naïve and activated T cells facilitates their diverse functions. Naïve T cells in the lymphatic tissues are in close proximity to DCs from many tissues facilitating antigen encounter. The function of activated T cells in immune surveillance, in contrast, is facilitated by migration through diverse tissues.

The multistep process of migration and the tissue selectivity of migration receptors have been targeted for the treatment of autoimmune diseases. Antibodies targeting CD62L and LFA-1 have been developed to prevent the entry of T cells into lymph nodes. In addition, the S1P1 inhibitor drug fingolimod has been recently approved for the treatment of multiple sclerosis (MS). Fingolimod is thought to prevent the egress of recently activated T cells from the lymph nodes. Natalizumab, a monoclonal antibody against α_4 integrin, a potent inhibitor of T-cell migration into the CNS and intestines, is utilized in the treatment of MS and Crohn disease to prevent the migration of T cells to the CNS and intestines, respectively.

DIFFERENTIATION OF CD4 Th SUBSETS

Naïve T Cells

Naïve T cells are precursors for effector and memory T-cell subsets. Phenotypically, naïve T cells are small cells with little cytoplasm; they express surface markers, such as CD45RA, CCR7, CD62L, CD127, and CD132. They lack expression of markers of previous activation, such as CD25, CD44, CD69, CD45RO, or HLA-DR. Naïve T cells have low metabolism and are unable to mediate effector responses or produce proinflammatory cytokines.[15] Naïve T cells migrate within the secondary lymphoid organs, where they interact with DCs. Because of their role as precursors for all T-effector subsets, their loss is thought to be an important contributor to the immunosenescence seen in older adults.

Early studies demonstrated that CD4 T-cell clones and cell lines would express reproducible patterns of cytokine expression designated T-helper cell-1 (Th1) and T-helper cell-2 (Th2).[16] It is now appreciated that CD4 effector T cells can differentiate into many functional phenotypes (see Fig. 11.1). These can be grouped into four general categories: (i) those possessing proinflammatory effector characteristics, (ii) those possessing regulatory or anti-inflammatory activities, (iii) those that promote B-cell follicle development, and (iv) those that provide long-term memory. Our understanding of the phenotype diversification process is rapidly expanding. T-effector cell differentiation occurs in three steps. First, TCR signals lead to activation and elicit the expression of cytokine receptors.[17] Second, signals through specific cytokine receptors differentially promote the expression of lineage-specific "master" transcription factors that promote expression of genes associated with a particular T-cell phenotype and suppress the expression of genes associated with other T-cell phenotypes. Third, the phenotype-specific transcription factors induce epigenetic changes that control gene accessibility and maintain the T-cell phenotype in a cell-intrinsic manner.

EFFECTOR CELL PHENOTYPES

T-effector cells promote inflammation through the release of cytokines (Fig. 11.2). Cytokine release elicits the activity of accessory cells, which ultimately mediate the inflammation to clear the antigen. The pattern of cytokine production ultimately dictates the type of inflammation (Table 11.1). Effector T cells are divided into five basic groups: Th1, Th2, Th9, Th17, and Th22. Table 11.2 summarizes the T-effector responses associated with selected pathogens.

Th1

Th1 cells are defined by production of the cytokines IFN-γ, GM-CSF, IL-2, and lymphotoxin (LT). Th1 differentiation

KEY CONCEPTS

CD4 Effector Phenotypes

Th1
- Produces interferon (IFN)-γ, IL-2, tumor necrosis factor (TNF), and lymphotoxin (LT)
- Stimulates IgG1 and IgG3 class switching
- Responses mediated by macrophage activity
- Promotes phagocytic activity through:
  - FcγRIII cross-linking
  - Complement deposition
  - Opsonization
  - IFN-γ-mediated macrophage activation

Th2
- Produces IL-4, IL-5, IL-6, and IL-10
- Stimulates IgG4 and IgE class switching
- Responses mediated by mast cells and eosinophils
- Increases degranulation through:
  - FcεRI cross-linking
  - IL-5–mediated eosinophil activation

Th17
- Produces IL-6, IL-17, granulocyte macrophage–colony-stimulating factor (GM-CSF), and TNF
- Activates local endothelium
- Induces cytokine and chemokine production
- Increases infiltration by neutrophils
- Activates cell-mediated inflammation

is elicited in response to infection by intracellular bacteria, fungi, and viruses, products of which stimulate surface toll-like receptors (TLR), leading to the production of cytokines by DCs and natural killer (NK) cells. Th1 differentiation is promoted by IL-12, IL-18, IFN-γ, and type 1 interferons and is inhibited by IL-4, IL-10, and transforming growth factor-β (TGF-β; Fig. 11.3A). During activation, IFN-γ receptor signals activate signal transducer and activator of transcription 1 (STAT1) and promote the expression of the Th1-restricted transcription factor T-bet and IL-12R expression. IL-12R signals through STAT4, which drives the expression of high levels of T-bet. T-bet serves to reinforce the Th1 phenotype by promoting IFN-γ and IL-12Rβ_2 expression.[17] IL-18 plays a dual role in Th1 function by promoting Th1 commitment and eliciting IFN-γ production by fully differentiated Th1 cells.

Th1 cells promote cell-mediated inflammatory responses through inducing the activation of macrophages, NK cells, B cells, and CD8 T cells. Th1 cells regulate macrophage function at several levels. GM-CSF promotes the production of monocyte lineage cells from the bone marrow. IFN-γ is a potent macrophage activator, enhancing microbicidal activity by initiating nitric oxide (NO) production, upregulating production of oxygen radicals, and increasing phagocytic function. IFN-γ also promotes antigen presentation by upregulating major histocompatibility complex (MHC) class I molecules, MHC class II molecules, and costimulatory molecule expression by macrophages (Fig. 11.2). IFN-γ can activate NK cells and also promote humoral responses in B cells to mediate antibody class switching to an IgG1 (IgG2a in mice) isotype.[18] IgG1 activates the classical complement pathway and can bind Fcγ receptors expressed on phagocytic cells, thereby promoting opsonization. Finally, IFN-γ acts in conjunction with another Th1 cytokine, IL-2, to promote the differentiation of CD8 cells into cytotoxic effector cells (Chapter 12). Macrophage-dependent Th1-mediated inflammatory responses are known as *delayed-type hypersensitivity* (DTH) responses. In vivo, DTH responses are critical for protection from intracellular pathogens including bacteria, fungi, and viruses. Individuals with defects in components of the IFN-γ or IL-12 signaling pathways are susceptible to atypical mycobacterial infections and have altered responses to other microorganisms. Additionally, Th1 cells are thought to contribute to the pathogenesis of autoimmune diseases, including MS, type 1 diabetes mellitus, rheumatoid arthritis, and Crohn disease. The Th1 pathway has been explored therapeutically. IL-12 has been studied as a way to boost vaccine immune responses against microorganisms and antitumor immunity. IL-12 neutralizing mAb-based therapies have been investigated in inflammatory and autoimmune diseases, including psoriasis, arthritis, Crohn disease, and MS.

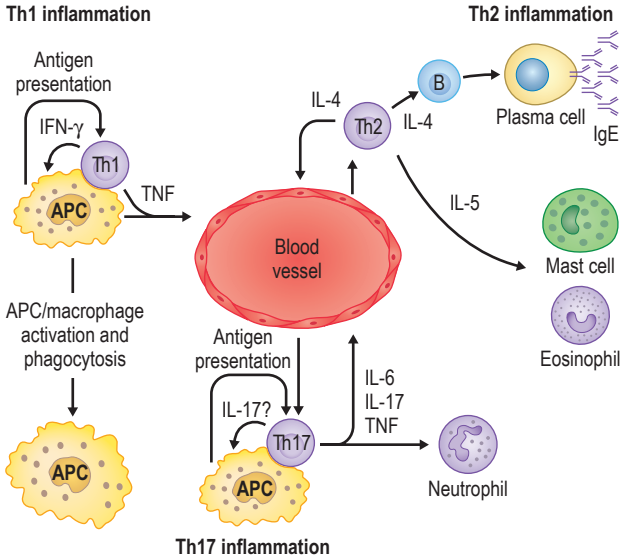


FIG. 11.2 Generalized Model for T Helper–Mediated Inflammation. *Top left.* Introduction of infectious agent stimulates the release of chemokines and tumor necrosis factor *(TNF)* from tissue macrophages, stimulating the recruitment *(upward arrows)* of T cells and monocytes through the local vasculature. Antigen recognition by T cells stimulates the local production of Th1 cytokines. Interferon *(IFN)*-γ activates macrophages, enhancing the clearance of infectious agents. *Top right.* Trafficking to sites of Th2 responses is stimulated by local chemokine expression, leading to T-cell recruitment. Antigen recognition leads to interleukin *(IL)*-4 production by Th2 cells, which stimulates B-cell immunoglobulin *(Ig)E* class switching. Production of IL-5 activates eosinophils. Cross-linking of FcεR1 molecules bound to IgE leads to the degranulation of mast cells and eosinophils. *Bottom.* Recruitment of Th17 cells and restimulation by antigen results in the release of IL-6, IL-17, and TNF, which promotes the recruitment, activation, and function of many cells, particularly neutrophils. *APC,* Antigen-presenting cells.

Th2

Th2 cells are defined by their production of IL-4, IL-5, IL-9, IL-10, and IL-13. In vitro, the critical step in Th2-cell differentiation is the presence of exogenous IL-4 and the absence of IFN-γ during T-cell activation. In vivo, Th2 differentiation is thought to require IL-4 produced by basophils, eosinophils, mast cells, NKT cells, or even previously differentiated Th2 cells. Naïve T cells express the IL-4R, and the combination of TCR, costimulatory (CD28 and ICOS), and IL-4R/STAT6 signaling promotes IL-4 transcription and the production of the transcription factors c-Maf and GATA3. c-Maf helps to establish Th2 polarity by promoting IL-4 and suppressing IFN-γ production. GATA3 also appears to inhibit IFN-γ production, but additionally plays a critical role in establishing Th2 cells by promoting IL-5 and IL-13 production (Fig. 11.3B).[19]

Th2 cells release IL-4 and IL-5, which attract and activate the function of eosinophils and mast cells. Th2-type cytokines enhance B-cell class switching toward IgG2, IgE, and sIgA. High levels of IgE in Th2 reactions combined with antigen exposure and FcεR1 receptor expression by eosinophils or mast cells results in triggering and release of inflammatory factors, such as histamine, platelet-activating factor, prostaglandins, and leukotrienes (Fig. 11.2). These factors act on the local environment, producing vascular dilation and leakage, bronchial constriction, and intestinal hypermotility. On a more systemic level, anaphylaxis may be produced. Eosinophil- and mast cell–dependent reactions are known as *immediate-type hypersensitivity* (ITH) responses. ITH responses are important for ridding the body of intestinal helminths; in fact, components of helminth eggs strongly promote Th2 differentiation. Th2 responses are also associated with atopy and hyperresponsive airway conditions, such as asthma and allergies.

TABLE 11.1 T-Helper Cell Effector Function Through Cytokine Secretion

Th1	
Interleukin (IL)-2	T-cell growth and potentiation of Fas-mediated apoptosis
	Natural killer (NK) cell growth and cytolytic activity
	B-cell growth and antibody production
Interferon (IFN)-γ	Increases class I and II molecule expression on numerous cell types
	Promotes Th1 differentiation
	Activates macrophages; stimulates phagocytic killing and oxidative bursts
	Induces immunoglobulin (Ig)G2α and IgG3 class switching
	Inhibits IgG1 and IgE class switching
	Inhibits Th2 proliferation
	Activates neutrophils
	Activates endothelium to promote CD4 T-cell adhesion
	Stimulates NK cell cytolytic activity
	Required for the activation of CD8 cytotoxic T lymphocytes (CTLs)
Tumor necrosis factor (TNF)	Activates vascular endothelial cells; enhances leukocyte recruitment
	Activates neutrophils, eosinophils, and macrophages
	Stimulates IL-1, IL-6, TNF, and chemokine expression by macrophages
	Protects against viral infections, similar to interferons (IFNs)
	Increases class I molecule expression
	Induces fever
	Activates coagulation
Lymphotoxin (LT)	Activates neutrophils and osteoclasts
	Activates vascular endothelial cells; enhances leukocyte recruitment
	Cytotoxic activity against tumor cells
	Stimulates adhesion molecule expression
	Increases class I molecule expression
Th2	
IL-4	Growth and differentiation of Th2 cells
	B-cell growth and class II molecule upregulation
	Induces IgG1 class switching (IgG4 in humans) and IgE
	Inhibits IgG2a and IgG3 class switching
	Mast cell growth
	Inhibits macrophage activation
	Induces vascular cell adhesion molecule (VCAM) expression on endothelium; recruits eosinophils/monocytes
IL-5	B-cell growth and activation
	Eosinophil differentiation
	Eosinophil activation and survival
IL-10	Inhibits cytokine and chemokine production in monocytes, especially TNF, IL-1, and IL-12
	Inhibits macrophage activation and function
	Inhibits T-cell–mediated inflammation
	Induces IgG4 class switching (IgG1 in mice)
IL-13	Upregulates class II molecules on monocytes and B cells
	Inhibits proinflammatory cytokine production by monocytes
	B-cell costimulation
	Induces IgG4 class switching (IgG1 in mice) and IgE
	Increases class I molecule expression
Th9	
IL-9	Mast cell expansion and recruitment
	T-cell growth factor
IL-21	Promotes Th17 differentiation
	B-cell survival; antibody class switching
	Enhances NK cell proliferation; promotes killer function
Th17	
IL-17	Increases T-cell proliferation
	Promotes neutrophil recruitment and activity
	Promotes cytokine production including IL-6 and TNF
	Induces chemokine production
Granulocyte macrophage–colony-stimulating factor (GM-CSF)	Promotes myeloid cell development and maturation
	Promotes dendritic cell differentiation and survival
	Enhances macrophage activation
IL-6	Activates acute-phase response, inducing fever and antibacterial responses
	Promotes Th2 and Th17 differentiation
	Activates and elicits NK responses
	Promotes plasma cell differentiation and Ig production
IL-21	See above
TNF	See above
Th22	
IL-22	Activates acute phase response, inducing fever and antibacterial responses
IL-13	See above
TNF	See above

TABLE 11.2 T-Cell Effector Responses to Selected Pathogens

Organism	Nature of the Immune Response
Bacteria	
Borrelia burgdorferi	Th1 and Th17 responses associated with protection and joint pathology
Chlamydia trachomatis	Th1 and Th17 responses protective and source of pathology
Helicobacter pylori	Th1 and Th17 responses contribute to protection and pathology. Th2 responses lead to persistence
Legionella pneumophila	Th1 responses associated with immunity
Listeria monocytogenes	Th1 responses are protective: interferon (IFN)-γ from γδ T cells and CD8 T cells is important
Mycobacterium leprae	Severity and phenotype of disease depends on Th1 and Th2 predominance
Mycobacterium tuberculosis	Th1 responses control infection
Treponema pallidum	Th1 resolves infection; Th2 chronic
Yersinia pestis	Th1 and Th17 responses associated with immunity
Fungi	
Aspergillus fumigatus	Th2 production predominates; does Th1 offer protection?
Blastomyces dermatitidis	Th1 protects; Th2 switch in progressive disease
Candida albicans	Th17 responses are protective
Cryptococcus neoformans	Susceptibility associated with Th2 response; Th17 response associated with protection
Paracoccidioides brasiliensis	Infection stimulates Th2 response, but Th1 response protects
Parasites	
Leishmania spp.	Th1 responses are protective; Th2 responses allow chronic infection
Filaria	Initiates Th2 production; Th1 response appears protective
Schistosoma mansoni	Th1 and humoral responses protect; typically Th2 responses directed against eggs
Trypanosoma cruzi	Th1 responses inhibit parasite replication, but protection is not completely CD4 dependent
Giardia lamblia	Th1 and Th2 responses protect
Viruses	
Measles	Th1 responses are protective
Hepatitis B	Th1 responses seen in spontaneous recovering patients
Human immunodeficiency virus	Rapid loss of effector and memory T cells correlates with susceptibility to pathogens.
Respiratory syncytial virus	Th2, Th9, and Th17 responses are associated with reduced lung function

Th17

Th17 T cells are characterized by the production of IL-17a/f, IL-21, IL-22, IL-26, GM-CSF, and TNF. Th17 differentiation is promoted by IL-1β, IL-6, IL-23, and TGF-β, and the absence of type 1 interferons, IFN-γ, and IL-4 (Fig. 11.3C). Much of the potency attributed to Th17 cells derives from the production of IL-17 and GM-CSF. IL-17, together with TNF, strongly promotes inflammation by inducing the expression of adhesion molecules, proinflammatory cytokines (including IL-6, GM-CSF, granulocyte–colony-stimulating factor [G-CSF]), chemokines, prostaglandin E2, and matrix metalloproteinases (see Fig. 11.2). IL-21 regulates B-cell, T-cell, and NK-cell functions.[20]

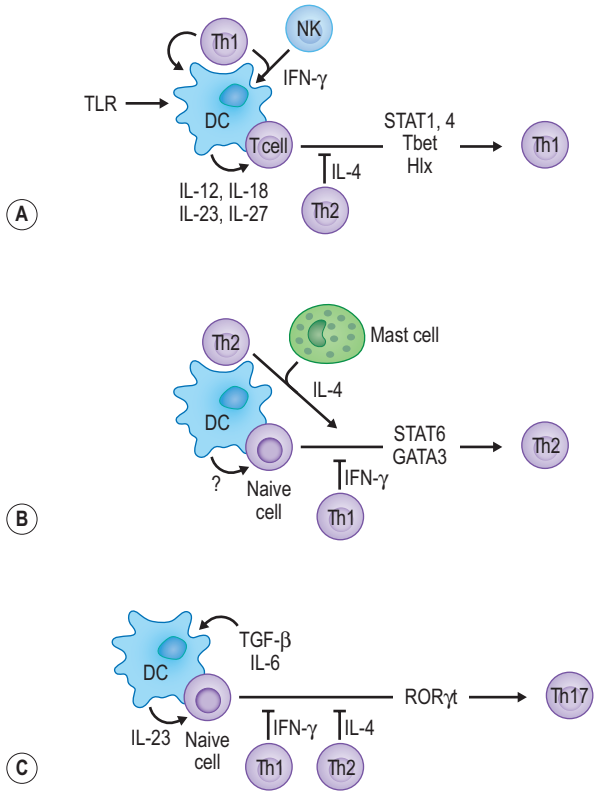


FIG. 11.3 Factors Influencing T-Effector Differentiation. Cytokine exposure during the activation stage of naïve T cells strongly influences T-effector differentiation. Depicted here are the factors promoting and inhibiting Th1 (A), Th2 (B), Th17 (C), and T-follicular helper (Tfh) cells following functional activation of undifferentiated T cells. DC, Dendritic cells; IFN-γ, interferon-γ; IL, interleukin; NK, natural killer; *RORgammat*, retinoic acid related-orphan nuclear receptor gamma; STAT, signal transducer and activator of transcription; TGF-β, transforming growth factor-β. *TLR*, Toll-like receptor.

In B cells, IL-21 has been found to regulate plasma cell differentiation and promote IgM and IgG1 antibody production. IL-21 is a T-cell growth factor, and in the presence of TCR signals, IL-21 promotes T-cell activation, proliferation, and survival. It therefore appears that Th17 cells are important for the recruitment of effector cells, including neutrophils and monocytes.

Th17 cells are potent cell-mediated effectors. Th17 responses are elicited in response to infection with extracellular bacteria and fungi. In addition, Th17 cells have been implicated in transplant rejection, atopic dermatitis, Crohn disease, psoriasis, and MS. Experimental models demonstrate that Th17 cells, via the production of GM-CSF and IL-17, are important mediators of autoimmune diseases, including experimental autoimmune encephalomyelitis (EAE).[21] Patients with impaired IL-17 expression or mutations causing Th17 differentiation experience recurrent infections with *Staphylococcus aureus* and *Candida albicans*. Alleles of the IL-23R have been found to confer protection or susceptibility to Crohn disease and psoriasis. Clinical trials are underway to examine the effects of disrupting Th17 immune responses in patients with autoimmune diseases. Therapeutic mAbs to IL12/23 p40, IL23 p19, and IL-17a have been approved for treatment of psoriasis and are being investigated for application in other conditions.

Th9

IL-9–producing T-cell subsets are a separate lineage. Human Th9 cells produce IL-9 and IL-21. IL-9 is an important growth factor for T cells, mast cells, and hematopoietic stem cells (HSCs). IL-9 functions by preventing apoptosis and by recruiting DC. The differentiation of naïve T cells into a Th9 phenotype is thought to be dependent on the presence of IL-4 and TGF-β and is enhanced by signaling through TNF receptor 2 and costimulation by TNFRSF4 (OX40).[22] Th9 cells express the transcription factor PU.1, which is thought to repress Th2-associated GATA3 and Th1-associated T-bet expression. Similar to Th2 cells, Th9 cells are thought to play a role in protection from helminths and promote airway restriction and mucus secretion in allergic asthma. Experimental and clinical studies indicate that Th9 cells may contribute to immune pathogenesis in inflammatory bowel disease, MS, rheumatoid arthritis, and other autoimmune diseases.

Th22

Although IL-22 was originally described as a Th17 cytokine, research has shown that a distinct CD4 T-cell population expressed IL-22 independently of IL-17. Further work identified IL-22, IL-13, and TNF secreting cells in the skin of patients with inflammatory skin disorders.[3] These IL-22–expressing cell populations also express the cutaneous homing receptors CCR4, CCR6, and CCR10. These Th22 cells have been subsequently found to differentiate in the presence of IL-6 and TNF.[23] Phenotypically, Th22 cells express a different gene profile than other Th cell subtypes, suggesting that they are a distinct effector population. They also produce fibroblast growth factors that promote epidermal repair. IL-22 is a member of the IL-10 family with important functions in regulating skin homeostasis and protection from infection by eliciting anti-bacterial protein production. Experimental models suggest that IL-22 and Th22 cells may play beneficial roles in diverse epithelial tissues.

KEY CONCEPTS

Regulation of T-Cell Responses

- Cell death pathways
  - TNF family receptors: Fas/FasL, TNF-related apoptosis inducing ligand (TRAIL), and TNFR1
- Inhibitory receptor activity
  - CTLA-4 and PD-1
  - Inhibition of TCR signaling
- Cytokine-mediated regulation
  - Anti-inflammatory cytokine production (IL-10, TGF-β)
  - Cytokines that promote and inhibit inflammation at different stages of a response (IFN-γ, IL-2, IL-27)
  - Regulation of apoptosis (LT, TNF)
  - Growth factor withdrawal (IL-2, IL-7, IL-15)

REGULATORY T CELLS

Regulatory T cells (Tregs) possess the ability to suppress or otherwise downregulate the function of other proinflammatory T cells. Although multiple T-cell subsets may possess regulatory T-cell activity, our understanding of regulatory function was greatly enhanced by the discovery of CD25 and FoxP3 as markers for a regulatory subpopulation of CD4 T cells (Chapter 13). This discussion focuses on the involvement of two populations of peripherally derived Treg phenotypes: adaptive Treg and Tr1 cells.

Adaptive Tregs

In mice, the majority of $CD4^+$ $CD25^+$ $FoxP3^+$ Tregs (natural Tregs) are thought to develop regulatory ability in the thymus. Peripherally derived adaptive Tregs are similar in phenotype to natural Tregs in that they express CD4, CD25, CD38, CD62L, CD103, and FoxP3. Adaptive Tregs are thought to be generated in response to prolonged exposure to antigen and are influenced by antigen presentation by immature DCs, plasmacytoid DCs, or nonprofessional APCs. The presence of inhibitory cytokines, such as IL-10 and TGF-β, also favors adaptive Treg differentiation. Similar to natural Tregs, adaptive Tregs are capable of suppressing CD4 and CD8 T-cell responses. Unlike natural Tregs, adaptive Tregs can function through both cell contact–dependent interactions and through the secretion of anti-inflammatory cytokines. These cytokines may directly regulate effector T cells or may inhibit DC activity. The generation of adaptive Tregs is of great interest as a cell-based therapy for inflammatory diseases, including autoimmunity.[24]

Tr1 Cells

Tr1 cells have been defined as a population of CD4 T cells that produce IL-10 in response to stimulation. Unlike adaptive Tregs, Tr1 cells do not express FoxP3 and do not necessarily express CD25. Tr1 cells are produced in the respiratory tract when antigen exposure is coupled with IL-10 and TGF-β. Experimentally, Tr1 cells can be induced by repeated intranasal antigen delivery.[25] Respiratory tract infections with *Bordetella pertussis* have also been found to elicit Tr1 cells in vivo.[26] Tr1 cells function by producing IL-10 to block proliferation of naïve T cells, prevent Th1 differentiation by blocking IL-12 production, and enhance differentiation of T cells toward the Tr1 phenotype.

FOLLICULAR HELPER T CELLS

One of the earliest described roles of CD4 T cells is their ability to promote or "help" generate effective antibody production by B cells. Th1 and Th2 cells have the ability to promote B-cell class switching and effector responses. Evidence supports the role of a distinct subpopulation of T cells, known as *T-follicular helper* (Tfh) cells, as playing a key role in regulating T-cell–dependent B-cell responses. Tfh cells are identified by their surface expression of CXCR5, ICOS), PD-1, CD200, B- and T-lymphocyte attenuator (BTLA), OX40, and serum amyloid P (SAP), while lacking the expression of CD127 and CCR7.[27] Tfh cells also express the transcription factor BCL6, which is essential for their differentiation and function. Interestingly, BCL6 can be co-expressed with T-bet, GATA3, or RORγt, but BCL6 modifies many of their effector lineage functions. The fact that there is an incomplete separation between Tfh and other T-effector lineages might suggest that the Tfh phenotype is generated in parallel with T-effector cell differentiation and in response to inflammatory cues present in lymph nodes.

Tfh differentiation is promoted by signals through ICOS, OX-40, and the presence of IL-6, IL-12, and IL-21. Tfh cells upregulate CXCR5, which enables homing to the interface between the T-cell and B-cell areas of lymph nodes, where they interact with recently activated B cells. This interaction with activated B cells provides signals to the Tfh cell through ICOS and IL-6, while the Tfh cells provide CD40L and IL-21 to the

B cells. Through successive interactions with B cells, Tfh cells promote the survival, proliferation, Ig class switching, affinity maturation and differentiation of B cells into memory B cells, and antibody-producing plasma cells through the germinal center reaction.[28]

Mutations affecting Tfh-associated pathways have been identified in B-cell immunodeficiency disorders. Common variable immune deficiency (CVID; Chapter 33) is often linked to mutations in the ICOS pathway. Recurrent infections in patients with CVID are associated with hypogammaglobulinemia and reduced numbers of Tfh cells. X-linked lymphoproliferative disease type 1 (XLP1) has been associated with mutations in the *SH2D1A* gene. Interestingly, patients with XLP1 have defects in B-cell germinal center formation and decreased memory B-cell numbers, but have normal Tfh numbers. XLP1 Tfh cells have reduced ability to support B-cell function.[29] Tfh cells have been implicated in autoimmune diseases, such as myasthenia gravis, Graves disease, Hashimoto thyroiditis, lupus, Sjögren syndrome, and rheumatoid arthritis, where dysregulated antibody responses are thought to contribute to disease progression. In addition, therapies to target or passively transfer Tfh cells are being studied for treatment of autoimmune disease and some forms of lymphoma.[29]

MEMORY T CELLS

Following an immune response, small numbers of T cells persist over the long term and are called *memory T cells*. Memory T cells protect the host from reinfection by the same microorganisms. Protection correlates with the number of specific memory cells present in the host. Memory cells accelerate responses against repeat antigen for several reasons. First, memory cells are maintained at higher frequencies than are naïve cells. Second, a large portion of memory T cells are maintained in peripheral tissues, allowing for rapid local responses to infection. Third, memory T cells proliferate and produce cytokines in response to stimulation with lower doses of antigen, less costimulation, and much faster kinetics compared with naïve T cells. Fourth, memory cells can promote APC function and produce cytokines that promote the effector differentiation of naïve T cells.

Diverse populations of memory T cells can be defined by surface marker expression, localization, and patterns of migration. Two populations of memory T cells are identified on the basis of their re-expression of the lymph node homing receptors CCR7 and L-selectin following activation. Central memory (T_{cm}) cells are identified as $CD45RA^-$ $CD45RO^+$, $CCR7^+$, $CD62L^+$, and stem cell memory cells (T_{scm}) are identified by the markers $CD45RA^+$ $CD45RO^-$, $CCR7^+$, $CD62L^+$, $CD95^+$.[30] By expressing CCR7 and CD62L, these memory T cells permit recirculation through secondary lymphatic organs, allowing interactions with DCs from diverse tissue sites effectively expanding the area of immune surveillance. T_{cm} and T_{scm} produce IL-2 and rapidly enter the cell cycle upon activation. They are able to acquire effector phenotypes, suggesting that they function as a precursor pool for subsequent immune responses.

Two additional memory cell subsets have been described on the basis of their ability to migrate through peripheral tissues. Effector memory T cells (T_{em}) are identified by the expression pattern: $CD45RO^+$, $CCR7^-$, $CD62L^-$. These cells express chemokine receptors, selectins, and integrins facilitating their recirculation through peripheral tissues. T_{em} cells have previously undergone effector differentiation, and upon stimulation they rapidly express restricted cytokine patterns reflecting their prior phenotype. Additional populations of nonmigratory memory T cells have been identified. These are known as tissue resident memory T cells (T_{rm}). T_{rm} cells are identified by the markers $CD45RO^+$, $CCR7^-$, $CD69^+$. They have been identified in close proximity to epithelial sites in multiple tissues. T_{rm} cells provide rapid and highly localized cytokine production in response to infection. Thus T_{em} and T_{rm} cells bolster immunity by providing immune surveillance at distinct sites.[31]

The processes underlying memory T-cell differentiation are incompletely understood. A number of signaling pathways and APC populations have been proposed to be involved in memory differentiation. It is clear that T_{cm} and T_{em} cells are generated early in the T-cell response. The pathways that lead to the differentiation of Th1, Th2, Th17, and Tfh effector cells are also important in the generation of T_{em} populations.

GENERAL CONSIDERATIONS IN EFFECTOR T-CELL DIFFERENTIATION

Several key concepts of Th maturation should be noted with regard to the cellular differentiation of CD4 T cells. First and foremost, all T-effector, memory, and adaptive Tregs are thought to arise from naïve thymic emigrant precursors. In addition, T-cell phenotype specification is regulated independent of TCR specificity.

Second, differentiation to a specific effector phenotype is self-promoting and inhibitory to other phenotypes through production of effector cytokines (see Fig. 11.3). Early in Th1 or Th2 differentiation, for example, expression of the transcription factors T-bet or GATA3 promote pro-effector differentiation while suppressing alternative cell fates. Once established, T-cell phenotypes are maintained through active signaling and chromosomal modification.

Third, polarization is incomplete at the population level. Following exposure to antigen, responding T cells may collectively possess multiple phenotypes. Heterogeneity of Th phenotypes is a common feature of some diseases, such as asthma, where Th2, Th9, and Th17 responses are thought to contribute to disease; in psoriasis, where Th1, Th17, and Th22 cell types are detectable; and in MS, where Th1, Th17, and Th22 cells are thought to mediate inflammation. A portion of the T cells may persist in a naïve phenotype or as uncommitted T_{cm} or T_{scm}. These cells retain features of stem cells in that they may give rise to T cells of diverse phenotype upon antigen re-exposure.

Fourth, there is limited plasticity in T-cell effector phenotypes. Effector phenotypes are reinforced by "master" transcription factors and epigenetic modifications (Fig. 11.4). For example, established Th1 cells do not express Th2 cytokines despite extended culture in Th2-skewing conditions. There is, however, growing evidence of interconversion between certain T-helper subsets.[32] Experiments tracing T-cell lineage commitment have shown that T cells that once expressed FoxP3 could become effector cells. It is not clear if plasticity occurs between all phenotypes or if phenotype conversion can occur at any stage of phenotype commitment (see Fig. 11.4).

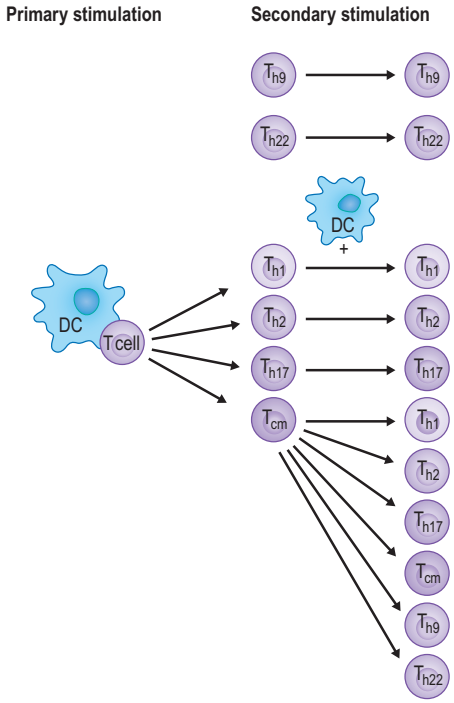


FIG. 11.4 Mechanism of T-Helper Phenotype Shift. Under appropriate conditions, T-cell activation may result in daughter cells adopting one of several T-effector phenotypes. Within a polarized population of Th cells, a subset of undifferentiated T lymphocytes may exist (here referred to in the general category as T central memory *[T_{cm}]*). Fully committed Th1, Th2, or Th17 cells upon restimulation will produce Th1, Th2, or Th17, respectively. T_{cm} or other uncommitted cell types retain the potential to differentiate into T effectors of any phenotype, depending on the context of the secondary stimulation. *DC*, Dendritic cells.

TERMINATION OF T-CELL RESPONSES

CLINICAL RELEVANCE

Diseases Associated With CD4 T-Effector Responses

Th1
- Multiple sclerosis
- Psoriasis
- Type 1 diabetes mellitus
- Tuberculoid leprosy
- Transplant rejection

Th2
- Allergy and asthma
- Helminthic infections
- Lepromatous leprosy

Th9
- Asthma
- Helminthic infections

Th17
- Asthma
- Multiple sclerosis
- Psoriasis
- Rheumatoid arthritis
- Transplant rejection

Th22
- Psoriasis
- Wound healing

As a result of the potential for unregulated inflammation, retention of activated effector T cells presents three risks to the organism. First, with each succeeding infection, the body would be increasingly burdened by the energy requirements of maintaining larger T-cell numbers. Second, there would be loss of TCR diversity resulting from the overrepresentation of clonally expanded T cells. Third, there is an increased risk of damage to self by large populations of activated effector T cells. Thus a return to homeostasis is required at the end of an immune response through reductions in T-cell numbers and activity. Homeostasis is restored through the loss of the initiating stimulus and by active control processes, including cell death, inhibitory signaling pathways, and anti-inflammatory cytokines.

CELL DEATH PATHWAYS IN T-CELL HOMEOSTASIS

Several cell death pathways regulate the processes of activation-induced cell death (AICD; Chapter 17). The primary pathway involved in AICD involves Fas (CD95) and Fas ligand (FasL, CD95L). Fas and FasL are expressed following T-cell activation. Fas-mediated apoptosis can be stimulated by FasL on neighboring cells ("death signal") or on the same cell ("suicide signal"). Mutations in Fas and FasL result in the accumulation of activated lymphocytes and autoimmunity. Another TNF family member, TRAIL, has also been associated with AICD. Mice lacking TRAIL have enhanced susceptibility to autoimmune disease. Some evidence suggests that TRAIL regulates AICD primarily in Th2 cells.[33]

ACTION OF INHIBITORY RECEPTORS

T cells express several Ig family transmembrane proteins containing immune tyrosine inhibitory motifs (ITIMs). Two of these, CTLA4 and PD-1, play important roles in terminating T-cell responses in vivo. CTLA4 is expressed by T cells following activation and regulates T-cell activity in two ways. First, CTLA4 binds to B7 with high affinity and competes with CD28 costimulation. Second, CTLA4 inhibits proximal TCR signaling by recruiting the phosphatases SHP-2 and PP2A. Mice deficient in CTLA4 develop a severe CD28-dependent lymphoproliferative disease, tissue infiltration, and early death.[34] PD-1 interacts with two ligands, PD-L1 and PD-L2, which are differentially expressed in immune and peripheral tissues. PD-1 engagement limits proximal TCR signals by recruiting SHP-2, thereby suppressing activation, cytokine production, and proliferation. Loss of PD-1 has been associated with autoimmune cardiomyopathy, arthritis, and a lupus-like glomerulonephritis.[35] Therapeutic disruption of CTLA-4 and PD-1 signals are exploited as checkpoint inhibitors for treatment of certain cancers.

CYTOKINE-MEDIATED INHIBITION

Cytokines function to terminate immune responses in three basic ways: (i) by loss or withdrawal of growth factors, (ii) by inducing cell death, and (iii) by direct anti-inflammatory properties (Chapter 14). IL-2 functions as a growth factor in early immune responses. Loss of IL-2 signals through decreased production or loss of receptor expression deprives the cell of survival signals, resulting in AICD. Similarly, loss of IL-7 or IL-7 receptor results in increased apoptosis among naive and memory T-cell populations.[36] As discussed in Chapter 17, cell death by growth factor withdrawal results from activation of the mitochondrial pathway of apoptosis.

Cytokines can suppress T-cell responses. This function of cytokines is exemplified by IL-10, IL-27, TGF-β, and TNF. IL-10

is produced by adaptive Tregs and Tr1 cells, as well as B cells, monocytes, and macrophages. IL-10 suppresses inflammation by inhibiting macrophage activation, downregulating chemokine production, and suppressing costimulatory molecule expression by APCs. IL-10–deficient mice develop inflammatory bowel disease caused by dysregulated Th1 responses. Mice genetically deficient for the IL-27 receptor develop exaggerated CD4 T-cell responses and inflammatory diseases. These effects of IL-27 can be attributed to its role in suppressing Th1 responses and CD4 T-cell proliferation.[37] TGF-β_1 is another anti-inflammatory cytokine that is produced by subsets of Tregs and a variety of other cell types throughout the body. Unlike IL-10, which targets APCs, TGF-β_1 directly inhibits T-cell proliferation and Th1 differentiation.[38] Mice deficient in TGF-β_1 have multi-tissue infiltration by activated lymphocytes and macrophages. TNF is well characterized as possessing proapoptotic function. TNFR1 possesses a death domain, and binding of TNF results in the activation of the caspase pathway and apoptosis.

SUMMARY OF THERAPEUTIC REGULATION OF T-CELL RESPONSES FOR TREATMENT OF IMMUNE-MEDIATED DISEASES

T-cell–mediated inflammation is central to immune-mediated diseases. The most effective techniques to control inflammation are aimed at the processes of T-cell activation, differentiation, trafficking, or effector functions (Fig. 11.5). Among many potential approaches currently being employed or in development are the following: (i) blockade of appropriate costimulatory signals or intracellular signaling molecules to prevent T-cell activation (*i.e.*, FK504) or to promote ongoing effector inflammatory responses (*i.e.*, anti-PD-L1); (ii) preventing recruitment of effector T cells into inflammatory sites by targeting pathways required for cellular recruitment (i.e., VLA-4); (iii) neutralizing the cytokines that promote T-effector differentiation (e.g., IL-6, IL-12, or IL-23); (iv) blocking proinflammatory cytokines (*e.g.*, IFN-γ, IL-17, GM-CSF, or TNF) and other effector molecules (e.g., iNOS, reactive oxygen intermediates, etc.) involved in the effector limb of destructive inflammatory processes; (v) the use of regulatory cytokines, such as TGF-β, IL-4, or IL-10 to regulate proinflammatory responses; and (vi) in vivo induction of Treg activity using tolerance strategies or ex vivo expansion and administration of protective or regulatory T-cell subsets.

ON THE HORIZON

- Development of effective techniques to identify antigens targeted in organ-specific autoimmune and allergic diseases and in transplant rejection
- Development and implementation of an effective tolerance-inducing strategy capable of regulating T-cell responses in an antigen-specific manner
- Identification of tolerance-specific biomarkers or assays to clearly delineate successful therapy of inflammatory diseases

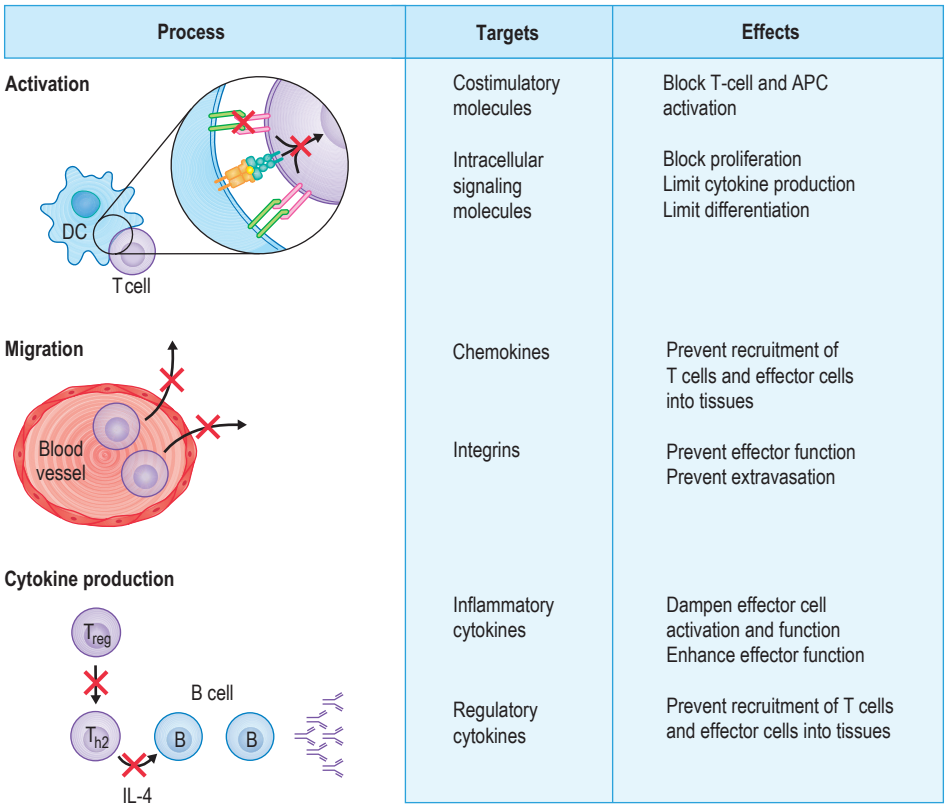


FIG. 11.5 Therapeutic Regulation of Inflammation. Several techniques have been employed to reduce T-cell responses in cases of inflammation. Among these are inhibiting T-cell activation and differentiation by blocking costimulatory interactions with antigen-presenting cells *(APCs) (top)*; inhibiting T-cell and effector-cell trafficking by blocking molecules required for chemotaxis *(middle)*; and limiting the effect of T-helper responses by decreasing the availability of inflammatory cytokines *(bottom)*. *DC*, Dendritic cells; *IL*, interleukin; *Treg*, regulatory T cells.

REFERENCES

1. Fischer A. Recent advances in understanding the pathophysiology of primary T cell immunodeficiencies. *Trends Mol Med*. 2015;21:408–416.
2. Azizi G, Ghanavatinejad A, Abolhassani H, et al. Autoimmunity in primary T-cell immunodeficiencies. *Expert Rev Clin Immunol*. 2016;12:989–1006.
3. Smith-Garvin JE, Koretzky GA, Jordan MS. T cell activation. *Annu Rev Immunol*. 2009;27:591–619.
4. Friedl P, den Boer AT, Gunzer M. Tuning immune responses: diversity and adaptation of the immunological synapse. *Nat Rev Immunol*. 2005;5:532–545.
5. Watts C, West MA, Zaru R. TLR signalling regulated antigen presentation in dendritic cells. *Curr Opin Immunol*. 2010;22:124–130.
6. Randolph GJ, Angeli V, Swartz MA. Dendritic-cell trafficking to lymph nodes through lymphatic vessels. *Nat Rev Immunol*. 2005;5:617–628.
7. Mempel TR, Henrickson SE, Von Andrian UH. T-cell priming by dendritic cells in lymph nodes occurs in three distinct phases. *Nature*. 2004;427:154–159.
8. Manicassamy S, Pulendran B. Dendritic cell control of tolerogenic responses. *Immunol Rev*. 2011;241:206–227.
9. Thomson AW, Turnquist HR, Raimondi G. Immunoregulatory functions of mTOR inhibition. *Nat Rev Immunol*. 2009;9:324–337.
10. Quezada SA, Peggs KS, Simpson TR, et al. Shifting the equilibrium in cancer immunoediting: from tumor tolerance to eradication. *Immunol Rev*. 2011;241:104–118.
11. Blattman JN, Antia R, Sourdive DJ, et al. Estimating the precursor frequency of naive antigen-specific CD8 T cells. *J Exp Med*. 2002;195:657–664.
12. Nguyen QP, Deng TZ, Witherden DA, et al. Origins of $CD4^+$ circulating and tissue-resident memory T-cells. *Immunology*. 2019;157:3–12.
13. Cyster JG. Chemokines, sphingosine-1-phosphate, and cell migration in secondary lymphoid organs. *Annu Rev Immunol*. 2005;23:127–159.
14. Rebenko-Moll NM, Liu L, Cardona A, et al. Chemokines, mononuclear cells and the nervous system: heaven (or hell) is in the details. *Curr Opin Immunol*. 2006;18:683–689.
15. Takada K, Jameson SC. Naive T cell homeostasis: from awareness of space to a sense of place. *Nat Rev Immunol*. 2009;9:823–832.
16. Mosmann TR, Cherwinski H, Bond MW, et al. Two types of murine helper T cell clone. I. Definition according to profiles of lymphokine activities and secreted proteins. *J Immunol*. 1986;136:2348–2357.
17. Zhu J, Paul WE. Peripheral CD4 T-cell differentiation regulated by networks of cytokines and transcription factors. *Immunol Rev*. 2010;238:247–262.
18. Coffman RL, Lebman DA, Rothman P. Mechanism and regulation of immunoglobulin isotype switching. *Adv Immunol*. 1993;54:229–270.
19. Paul WE, Zhu J. How are T(H)2-type immune responses initiated and amplified? *Nat Rev Immunol*. 2010;10:225–235.
20. Leonard WJ, Spolski R. Interleukin-21: a modulator of lymphoid proliferation, apoptosis and differentiation. *Nat Rev Immunol*. 2005;5:688–698.
21. Dong C. Diversification of T-helper-cell lineages: finding the family root of IL-17-producing cells. *Nat Rev Immunol*. 2006;6:329–333.
22. Soroosh P, Doherty TA. Th9 and allergic disease. *Immunology*. 2009;127:450–458.
23. Eyerich S, Eyerich K, Pennino D, et al. Th22 cells represent a distinct human T cell subset involved in epidermal immunity and remodeling. *J Clin Invest*. 2009;119:3573–3585.
24. Masteller EL, Tang Q, Bluestone JA. Antigen-specific regulatory T cells—ex vivo expansion and therapeutic potential. *Semin Immunol*. 2006;18:103–110.
25. Roncarolo MG, Gregori S, Battaglia M, et al. Interleukin-10-secreting type 1 regulatory T cells in rodents and humans. *Immunol Rev*. 2006;212:28–50.
26. Byrne P, McGuirk P, Todryk S, et al. Depletion of NK cells results in disseminating lethal infection with Bordetella pertussis associated with a reduction of antigen-specific Th1 and enhancement of Th2, but not Tr1 cells. *Eur J Immunol*. 2004;34:2579–2588.
27. Nutt SL, Tarlinton DM. Germinal center B and follicular helper T cells: siblings, cousins or just good friends? *Nat Immunol*. 2011;12:472–477.
28. Crotty S. T follicular helper cell differentiation, function, and roles in disease. *Immunity.*. 2014;41:529–542.
29. Ma CS, Uzel G, Tangye SG. Human T follicular helper cells in primary immunodeficiencies. *Curr Opin Pediatr*. 2014;26:720–726.
30. Lanzavecchia A, Sallusto F. Understanding the generation and function of memory T cell subsets. *Curr Opin Immunol*. 2005;17:326–332.
31. Carbone FR. Tissue-resident memory T cells and fixed immune surveillance in nonlymphoid organs. *J Immunol*. 2015;195:17–22.
32. O'Shea JJ, Paul WE. Mechanisms underlying lineage commitment and plasticity of helper $CD4^+$ T cells. *Science*. 2010;327:1098–1102.
33. Roberts AI, Devadas S, Zhang X, et al. The role of activation-induced cell death in the differentiation of T-helper-cell subsets. *Immunol Res*. 2003;28:285–293.
34. Chikuma S, Bluestone JA. CTLA-4 and tolerance: the biochemical point of view. *Immunol Res*. 2003;28:241–253.
35. Okazaki T, Honjo T. The PD-1-PD-L pathway in immunological tolerance. *Trends Immunol*. 2006;27:195–201.
36. Rochman Y, Spolski R, Leonard WJ. New insights into the regulation of T cells by gamma(c) family cytokines. *Nat Rev Immunol*. 2009;9:480–490.
37. Hunter CA. New IL-12-family members: IL-23 and IL-27, cytokines with divergent functions. *Nat Rev Immunol*. 2005;5:521–531.
38. Li MO, Wan YY, Sanjabi S, et al. Transforming growth factor-beta regulation of immune responses. *Annu Rev Immunol*. 2006;24:99–146.

12

Cytotoxic T Lymphocytes and Natural Killer Cells

Stephen L. Nutt and Nicholas D. Huntington

Cytotoxic T lymphocytes (CTLs) and natural killer (NK) cells represent two distinct but functionally related lineages that contribute to pathogen and tumor immunity. Although the approaches by which CTLs and NK cells kill their target cells and produce immunomodulatory cytokines are quite similar, the mechanisms by which they recognize their targets are distinctly different. CTLs are CD8 T cells (Chapter 9) that recognize targets via the interaction of a diverse repertoire of polyclonally rearranged T-cell receptors (TCRs) (Chapter 4) with a peptide–major histocompatibility complex (MHC) class I complex (Chapter 6) and are components of the adaptive immune response. MHC class I molecules are expressed on virtually all cells in the body and allow CTLs to scan the tissues for cells expressing foreign or cancer-associated peptides. In contrast, NK cells are members of the innate immune system (Chapter 3) and use an array of invariant activating and inhibitory receptors to control their activity and specificity.[1] These fundamentally distinct approaches to the recognition of antigen allow for complementary functions, with CTLs being specialized in detecting cancerous cells or those infected with intracellular pathogens, such as viruses, whereas a prominent function of NK cells is to eliminate those cells where a pathogen or oncogene has impaired MHC class I expression and/or has induced the expression of stress ligands. As one of the principal immune-evasion mechanisms of viruses and tumors is suppression of MHC class I expression, NK cells provide a key line of defense against this strategy.

CLINICAL RELEVANCE

Functions of Cytotoxic T Lymphocytes and/or Natural Killer Cells

Protective functions include:
- Host defense against:
  - Viruses, including human immunodeficiency virus (HIV), Epstein–Barr virus (EBV), pox virus, herpes simplex virus (HSV), and cytomegalovirus (CMV)
  - Bacteria, including *Listeria monocytogenes*
  - Parasites, including *Plasmodium falciparum* and *Toxoplasma gondii*
  - Primary and metastatic tumors
- Positive regulation of:
  - Graft-*versus*-leukemia (GvL) effect
  - Tumor inflammation by tumor-resident natural killer (NK) cells
  - Placental vascularization by uterine NK cells

Uncontrolled cytotoxicity contributes to:
- Some autoimmune disease, including diabetes and rheumatoid arthritis
- Hypersensitivity reactions
- Graft-versus-host disease (GvHD)
- Transplant rejection

The importance of the lytic function of CTLs and NK cells has been demonstrated in animal models as well as in patients with defective cytotoxicity. A number of recessive genetic syndromes that affect cytotoxic function have been reported, including familial hemophagocytic lymphohistiocytosis (FHL), which results from mutations in the perforin gene.[2] Patients with FHL present with severe immunodeficiency that is often associated with uncontrolled viral infections, including cytomegalovirus (CMV), herpes simplex virus (HSV), and Epstein–Barr virus (EBV) infections. Similarly, mice lacking CTLs and NK cells are overtly susceptible to viral pathogens and display impaired immunosurveillance.[2]

With its potent ability to control pathogen-infected and malignant cells, it is not surprising that modulation of cytolytic activity is an aim of many immunotherapies. These strategies involve either dampening of CTL function in such situations as transplantation or autoimmunity or enhancement of CTL and NK-cell function via vaccination, blocking antibodies to inhibitory receptors, adoptive cell transfer, or cytokine therapy. Tight controls need to be maintained over these effector cells, as deregulated CTL activity can promote autoimmune diseases, hypersensitivity reactions, graft-versus-host disease (GvHD), and transplant rejection. To maintain the discrimination between killing damaged or infected cells and not killing healthy neighboring cells, numerous layers of regulation operate to control cytotoxic functions.

EFFECTOR FUNCTIONS/MECHANISMS

KEY CONCEPTS

Cytotoxic T-Lymphocyte and/or Natural Killer Cell Effector Mechanisms

Cytotoxicity
- Killing by the perforin/granzyme pathway
- Death receptor–mediated apoptosis, including Fas and tumor necrosis factor (TNF)–related apoptosis–inducing ligand (TRAIL)

Immune modulation
- Inflammatory cytokine production, including interferon (IFN)-γ and TNF
- Chemokine secretion
- Immunomodulatory cytokines, including interleukin-10 (IL-10), transforming growth factor-β (TGF-β), and granulocyte macrophage–colony-stimulating factor (GM-CSF)

Cytotoxicity

Cytotoxic cells kill their targets via two major pathways: perforin/granzyme–mediated lysis and death receptor–induced apoptosis. Both require intimate contact between the lytic cell and its target (Fig. 12.1).[3] Although the processes are similar

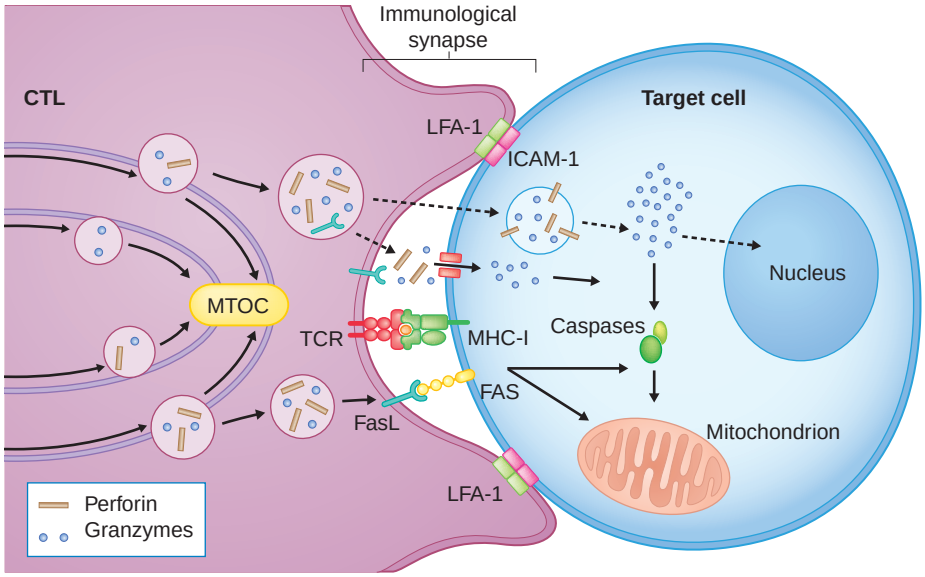


FIG. 12.1 Mechanisms of Cytotoxic T-Lymphocyte *(CTL)*–Induced Cell Death. The CTL recognizes its target via the interaction of the T-cell receptor *(TCR)* and the peptide–major histocompatibility complex class I *(MHC-I)* complex on the target cell. TCR signaling induces the formation of an immunological synapse that is stabilized by the binding of leukocyte function–associated antigen-1 *(LFA-1)* to the intercellular adhesion molecule *(ICAM)* on the target cell. Lytic granules containing perforin, granzymes, and FasL are polarized along microtubules and move toward the microtubule-organizing center *(MTOC)*. Lytic granules are then secreted into the synapse, allowing perforin to form pores in the target cell membrane, which facilitates granzyme access to the target cells. Granzymes induce apoptosis by caspase-dependent and independent pathways that result in DNA cleavage and mitochondrial damage. Membrane-bound FasL can bind to its receptor on target cells and induce apoptosis through an independent pathway. *FasL*, Fas ligand.

for CTLs and NK cells, CTL lytic activity is acquired only after activation and differentiation, whereas NK cells can spontaneously kill target cells. Despite this, NK-cell killing is significantly increased through prior activation by cytokines or inflammatory signals. Both cell types also produce cytokines that further enhance the immune response, most notably interferon-γ (IFN-γ).

Perforin–Granzyme Pathway

Perforin is a membrane-disrupting protein, which, together with a family of serine proteases (granzymes), forms the bulk of lytic granules. The process of lysis has been most extensively studied in CTLs, where, upon interaction between the TCR and an appropriate MHC class I peptide, a synaptic complex forms between the CTL and its target.[3] Lytic granules can then be observed moving along a microtubule network toward the microtubule-organizing center that localizes at the synapse (see Fig. 12.1). This process allows the polarized secretion of lytic granules precisely at the CTL–target cell interface. Perforin forms a pore that disrupts the target cell membrane, including either the plasma membrane or the lysosomal membrane. Once inside the target cell, it is granzymes that are the initiators of cell death. Granzymes function directly by cleaving substrates, such as nuclear proteins and DNA, or indirectly via initiating a protease cascade. One substrate of granzymes is the proapoptotic protein BID (BH3-interacting domain death agonist), which induces cell death via mitochondrial mediators, such as cytochrome c.

Death Receptor–Induced Apoptosis

Cytotoxic cells also have a receptor-based system to induce apoptosis of target cells (Chapter 17). This pathway uses members of the tumor necrosis factor receptor (TNFR) superfamily that are expressed on the target cells. These receptors have an intracellular signaling motif, called the *death domain*, which recruits molecules that transduce the death signal, such as FADD (Fas-associated death domain). The two most prominent apoptosis-inducing TNFR family members are Fas (CD95) and TRAIL (TNF-related apoptosis–inducing ligand).[4] Fas is expressed on a wide variety of tissues, whereas Fas ligand (FasL) expression is restricted to activated CTLs and NK cells, where it is stored in lytic granules and, upon activation, released to the effector cell membrane. The Fas/FasL pathway is important in controlling T-cell numbers through activation-induced cell death, as well as in the rejection of some tumors. Cytotoxic cells also express TRAIL, which, upon binding to TRAIL receptors, induces apoptosis in a wider selection of targets.[4] Of particular therapeutic interest is the fact that a subset of tumor cells are exquisitely sensitive to TRAIL. While each CTL can kill hundreds of tumor cells, most NK cells have the ability to kill <10 tumor cells before their cytotoxicity is lost. A small subpopulation of NK cells are capable of substantial serial killing (>30 tumor cells), primarily through the secretion of TNF superfamily ligands including TRAIL.[5]

Cytokines

Antigen-stimulated CTLs and activated NK cells modulate the immune response by their ability to produce a variety of cytokines, most notably IFN-γ and TNF. These potent inflammatory cytokines activate macrophages, dendritic cells (DCs), and lymphocytes at the site of infection. IFN-γ helps establish a T-helper cell-1 (Th1) response (Chapter 11) and further stimulates differentiated CTLs. NK cells are also a potent source of a diverse

range of cytokines, including GM-CSF, interleukin (IL)-10, and IL-13 (Chapter 14). The capacity to secrete a broad spectrum of cytokines provides NK cells with a wide range of regulatory capabilities. CTLs and NK cells are capable of secreting a number of chemokines (Chapter 15), including chemokine ligand 3 (CCL3), CCL4, and CCL5. These chemokines help recruit additional immune cells such as other lymphocytes and DCs to the response.

CYTOTOXIC T CELLS

Development and Tissue Distribution of Cytotoxic T Lymphocytes

CD8 T lymphocytes develop in the thymus, where they are selected for their ability to recognize nonself peptides in the context of MHC class I molecules (Chapter 9). Upon thymic export, these cells acquire a quiescent state and are referred to as *naïve*. Naïve CD8 T cells circulate between peripheral lymphoid organs, such as the spleen and lymph nodes, via the arterial and lymphatic systems. The tissue distribution of lymphocytes is determined by targeting proteins, which can be divided into three categories: selectins, chemokine receptors, and integrins.[6] Naïve and activated CD8 T cells display distinct sets of these targeting proteins, allowing for the differential homing abilities of these cells (Chapter 16). Naïve CD8 T cells express high levels of the lymph node homing receptor L-selectin (CD62L) and CCR7, a chemokine receptor that recognizes CCL19 and CCL21, which are produced in the T-cell areas of secondary lymphoid organs (Table 12.1; Chapter 2).[6] Here, naïve T cells interact with antigen-presenting cells (APCs). If a naïve CD8 T cell does not encounter its specific antigen, it leaves the lymph node.[6] If, however, a CD8 T cell encounters the correctly presented peptide–MHC I complex, a dramatic change in its localization and homing properties ensues. These cells shut down their egress program and undergo multiple rounds of proliferation to become activated CTLs. After the proliferative phase, the CTLs then reacquire egress capacity and travel via the circulation to nonlymphoid sites, where they tether to endothelial cells and extravasate into tissue. This transmigration occurs in both inflamed and noninflamed sites, such as skin, the gut, or the lung. Many of the effector memory CTLs are retained in nonlymphoid tissues, where they are poised to respond rapidly should the antigen be encountered again. The distinct types of memory T cells are discussed below.

THE CYTOTOXIC T-LYMPHOCYTE RESPONSE

The CTL response to an acute infection consists of three phases: first, the initial activation and proliferation of the CTL; second, the contraction of effector populations; and third, the long-term maintenance of memory cells.

Initial Activation

Naïve T cells constantly circulate through secondary lymphoid organs, where antigen encounter occurs. For a CTL response, antigen is brought to the lymph node via the lymphatic system by APCs (Fig. 12.2). These APCs are typically DCs that mature after antigen acquisition in nonlymphoid tissues and migrate to the lymph node. These antigens are recognized only when complexed to MHC molecules. APCs efficiently degrade self or foreign (pathogen-derived) proteins into shorter fragments (generally 8 to 10 amino acids in length) by the action of proteases in the cytosol. They are then transported into the lumen of the endoplasmic reticulum, where they are loaded onto newly synthesized MHC class I molecules for presentation at the cell surface (Chapter 6). This then enables the APCs to communicate with the antigen-specific CD8 T cells via interactions between the TCR and MHC molecules.

In the lymph node, CTLs scan the APCs for the presence of antigenic peptides complexed with the MHC class I molecules, a process termed *immune surveillance*. In the absence of a specific recognition by the TCR, the encounter is only transient, and the T cell continues on to another APC to repeat the process. If the MHC class I–peptide complex is bound by the TCR and initiates signaling, a more lasting interaction occurs.

TCR activation promotes polarization of the T cell and formation of the "immunological synapse."[3] The immunological synapse is a highly structured body that functions to concentrate TCR signaling in a defined area. It is associated with the selective recruitment of signaling molecules and exclusion of negative regulators. The synapse is stabilized by a ring of adhesion molecules, including, for example, LFA-1, which binds to ICAM1 on the APC (see Fig. 12.1). For a T cell to become fully active, costimulation through a second signaling pathway is required.[3] Many costimulators that have been identified share the common characteristic of being transmembrane receptors, often of the TNFR superfamily, that bind transmembrane ligands on the APC. The most important costimulator, CD28, binds the ligands CD80 and CD86, on activated APCs. Costimulation

TABLE 12.1 Properties of Cytotoxic T-Lymphocyte Populations

Marker	Naïve CD8 T Cell	Effector CTL	Effector Memory (T_{EM})	Central Memory (T_{CM})	Tissue-Resident Memory (T_{RM})
CD69	–	++	–	–	++
CD62L	++	–	–	++	–
CD44	+	+++	+++	+++	+++
CCR7	+	–	–	+	–
IL-7R (CD127)	++	+	+	++	+
IL-2Rβ (CD122)	+	+	++	+++	+
Main tissue distribution	Lymph nodes, spleen, blood	Lymph nodes, spleen, blood, nonlymphoid tissue (e.g., lung, liver)	Nonlymphoid tissue (e.g., lung, liver), spleen	Lymph nodes, spleen, blood	Nonlymphoid tissue (e.g., lung, liver, skin)
Cytotoxic function	–	+++	++	–	++
IFN-γ	–	+++	+++	+	+++

CTL, Cytotoxic T lymphocyte.

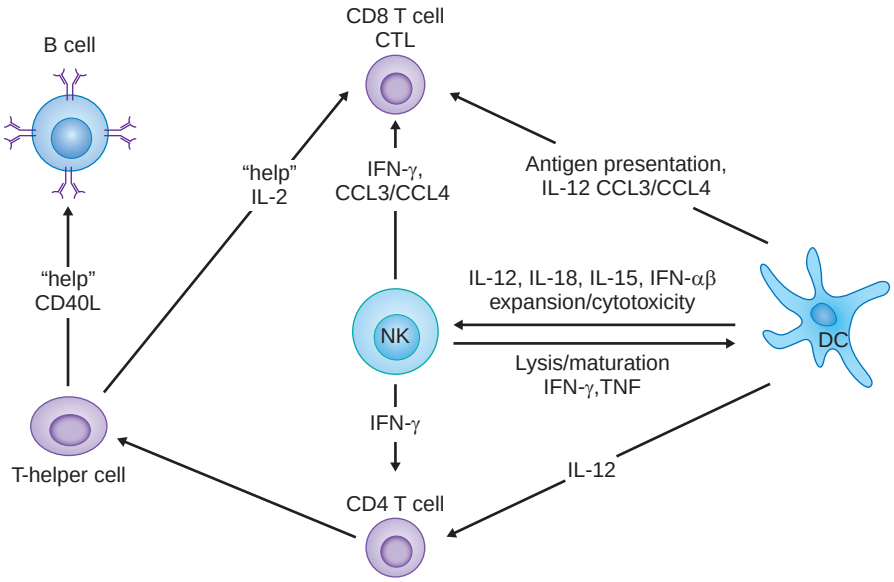


FIG. 12.2 Cellular Interactions During an Immune Response in the Lymph Node. After their encounter with antigen, dendritic cells *(DCs)* move to the draining lymph node, where they initiate an antigen-specific immune response. In the very early stages, natural killer *(NK)* cells are recruited to the lymph node and modulate various aspects of this response. NK cells provide cytokines that induce the maturation of DCs, which enables these cells to efficiently present the antigen to T cells in the context of costimulatory signals. NK cells also have the ability to eliminate immature DCs and to provide "early" interferon *(IFN)*-β for the initiation of a T-helper 1 (Th1)-type CD4 T-cell response. Finally, NK cells produce a range of chemokines, most notably chemokine ligand 3 *(CCL3)*/*CCL4*, which are crucial for the recruitment of CD8 T cells into the immune response. *CTL*, Cytotoxic T lymphocyte; *TNF*, tumor necrosis factor.

results in the clonal expansion of CTLs with the selected antigen specificity. The expression of CD80/86 occurs only after an APC receives activation signals, such as inflammatory cytokines, or components of the pathogens such as bacterial lipopolysaccharide. Naïve T cells that receive TCR stimulation in the absence of costimulatory signals can become nonresponsive to antigen, a state termed *anergy*.

Cross-Presentation and Priming

After it was established that only direct interaction with an APC and appropriate costimulation led to full CTL activity, a problem arose in explaining the mechanism by which antigens in non-APCs were recognized by CTLs. This dilemma was resolved by the discovery that there are two distinct mechanisms by which CTLs encounter peptide–MHC class I complexes.[7] If the APC expresses the antigen—for example, if a virus infects it—then the APC can process the antigen via the endogenous MHC class I pathway for presentation. The more intriguing situation arises when the APC does not express the antigen. In this case, the APC acquires and assimilates the antigen via a process termed *cross-presentation* (Chapter 6).[7] Cross-presentation is initiated by the capture of foreign or exogenous antigens by phagosomes. The antigens are then processed by an unusual mechanism that directs the peptides to the MHC class I pathway and presentation on the cell surface. An encounter of a CTL with an antigen processed in this pathway is termed *cross-priming*.[7] DCs can also acquire preformed peptide–MHC complexes from the plasma membrane through a process called *trogocytosis*. This mechanism acts to significantly amplify the magnitude of priming of T cells, and, like cross-presentation, potentially circumvents immune evasion strategies used by pathogens.[7]

Contraction of Effector Populations

After activation of CTLs in secondary lymphoid organs, an immune response is characterized by the rapid proliferation of antigen-specific cells and their acquisition of effector functions. CTLs proliferate at one of the fastest rates known for mammalian cells, with a cell cycle time of approximately 6 hours. Infection leads to a dramatic increase in the numbers of pathogen-specific CTLs, from almost undetectable initial levels to several million cells in the course of a single week. The expansion phase is followed by a contraction of the CTL population that is independent of both the magnitude of the response and clearance of the antigen. This phase is essential to prevent nonspecific tissue damage through uncontrolled cytokine release and cytolytic activity. Contraction also preserves the flexibility of the T-cell response to new infections while memory of previously encountered antigens is maintained. Typically, less than 5% of the expanded antigen-specific population survives in the long-lived memory pool.

Maintenance of Memory Cells

The production of long-lasting memory cells is essential for a rapid response should reinfection occur. CTL memory provides a more vigorous response than the primary challenge for both quantitative and qualitative reasons.[8] Quantitatively, because of the substantial clonal expansion during a primary infection, the precursor frequency of antigen-specific CTLs is vastly higher in immune individuals than in naïve subjects, thus allowing for a stronger response. Qualitatively, memory CTLs exhibit striking efficiency in elaborating the effector functions associated with the rapid production of IFN-γ. This enhanced response is the result of reprogramming of gene expression profiles by epigenetic changes in DNA methylation or chromatin structure.

The CTL memory compartment is composed of three cell types: effector memory (T_{EM}) cells, central memory (T_{CM}) cells, and tissue-resident memory (T_{RM}) cells. These subsets differ in their surface molecule expression and in their ability to exhibit effector functions (see Table 12.1). Like their naïve counterparts, T_{CM} cells express high levels of CD62L and CCR7 and reside primarily in secondary lymphoid organs. T_{CM} cells are capable of prolonged homeostatic self-renewal in the absence of antigen. T_{EM} cells, in contrast, are characterized by low expression of CD62L and CCR7 and are distributed throughout the body, including peripheral tissues, such as the lung and the gut, where they can immediately confront invading pathogens. T_{RM} cells are the most recently identified memory population that resides in peripheral tissues long after an infection is cleared, providing potent early response to reinfection of the same tissue.[6] CD4 T-cell help and cytokines, including IL-15 and IL-7 and their receptors, have been identified as crucial for the survival and maintenance of the memory T-cell pool.[6,8]

CD4 T-CELL HELP

The final player in the initial activation of CTLs is the "help" provided by CD4 T cells specific for an antigen linked to the CTL epitope (Chapter 11). The processes by which help is provided are poorly understood. It is likely that cytokines, such as IL-2 and IL-21, are involved and that the CD4 T cells influence both DCs and CTLs.[9,10] The cytokines, provided by the CD4 T cells, promote the survival, proliferation, and programming of memory CTLs. CD4 T-cell–deficient mice have been developed to study these CTL responses. Interestingly, "helpless" CD8 T cells resemble CTLs in chronic infections and cancer, in which the targets are not cleared despite a robust CTL response. PD-1 (*p*rogrammed *d*eath 1), an inhibitory receptor, is expressed on both helpless CTLs and on CTL cells during chronic infection and cancer. Blocking the interaction of PD-1 with its ligands greatly enhances the numbers and functions of impaired CTLs, a finding that is the basis for many current immunotherapy regimes in clinical oncology.[11,12]

DETECTION AND ANALYSIS OF CYTOTOXIC T-LYMPHOCYTE FUNCTION

Much progress in our understanding of the generation of CTLs in an immune response has resulted from the development of accurate and sensitive assays for CTL function. Traditional CTL assays were performed on bulk populations of effector cells. In these assays, target cells are labeled with ^{51}Cr or a nonradioactive dye, and then pulsed with peptide-antigen. Peptide-specific CTLs are incubated with the target cells. Lytic activity is then measured by the release of label into the culture supernatant or the loss of target cell viability directly. Although such an approach provides a powerful quantitative assay for CTL activity, it has the disadvantage of requiring prestimulation of the CTL population for 1 to 2 weeks to expand the numbers of antigen-specific CTLs to detectable levels. In mouse models, this limitation was overcome by the development of mouse strains that express a single transgenic TCR that recognizes either MHC class I (CD8)– or MHC class II (CD4)–specific epitopes. Such T cells, which are specific for a single peptide, have proven extremely useful in the study of CTL responses, as antigen-specific cells can be easily detected.

TCR transgenic mouse models do, however, have limitations, as they do not recapitulate the diversity of the normal immune response, and so represent an approach that cannot be broadly used in human studies, although human adoptive T-cell therapy using chimeric antigen receptors (CARs; a form of transgenic TCR) against an antigen specific to B cells has been successfully used to treat some lymphomas (Chapter 81).[13] The development of labeled MHC class I–peptide complexes that bind to the endogenous TCR has made a major contribution to overcoming these previous limitations, as it allows for the detection of rare antigen-specific CTLs within a polyclonal population from patient material or animal tissue.[14] This technique is simple and broadly applicable (Fig. 12.3). The ability to identify antigen-specific CTLs has been combined with single-cell functional assays, such as ELISpot or intracellular cytokine assays.

NATURAL KILLER CELLS

Properties of Natural Killer Cells

NK cells were the founding member of the innate lymphoid cell (ILC) family, which, in contrast to B and T cells, lack somatic antigen–receptor rearrangements. NK cells develop in bone marrow from a common ILC progenitor.[15] IL-15 is essential for most aspects of NK-cell biology, including differentiation, survival, proliferation, and activation in vivo. The IL-15R consists of three components: the IL-2/15β chain (CD122), a unique IL-15Rα chain, and the common γ chain, which contains the intracellular signaling component of the receptor.[16] The importance of this receptor complex is emphasized by the lack of NK cells in patients with X-linked severe combined immunodeficiency (SCID) that have mutations in the common γ-chain gene. The IL-15/IL-15R complex does not function like other cytokines that are produced as soluble ligands and bind to their receptor in a paracrine or autocrine manner. IL-15 is virtually undetectable in body fluids or cell culture supernatants, despite the broad distribution of its mRNA. The solution to this paradoxical observation is that IL-15 function requires the presence of IL-15Rα in the same cell. The IL-15/IL-15Rα complex is then presented on the cell surface in *trans* to the NK cell expressing the IL-15Rβγ complex.[16] In this manner, IL-15 maintains the homeostatic frequency of NK cells in the body. Once NK cells have acquired IL-15 responsiveness in bone marrow, they proceed through an ordered differentiation process that results in an expansion phase and the acquisition of a panel of germline-encoded activating and inhibitory receptors (Table 12.2).[1] It was previously assumed that these bone marrow–derived NK cells were fully functional, but more recent data suggest that multiple mature NK-cell subsets exist, with varying levels of effector function.[15]

Tissue Distribution and Diversity of Natural Killer Cells

In line with their surveillance function, NK cells are found at many sites in the body, including bone marrow, peripheral blood, lymph nodes and the spleen. The broad range of tissue distribution suggests that there is diversity in the function of mature NK cells. Furthermore, NK cells are also found in nonlymphoid organs, such as the liver and small intestine, and in the decidual lining of the uterus along with a closely related innate lymphocyte population, named *group 1 innate lymphoid cells* (ILC1; Chapter 2), which likely have a role distinct from that of NK cells in

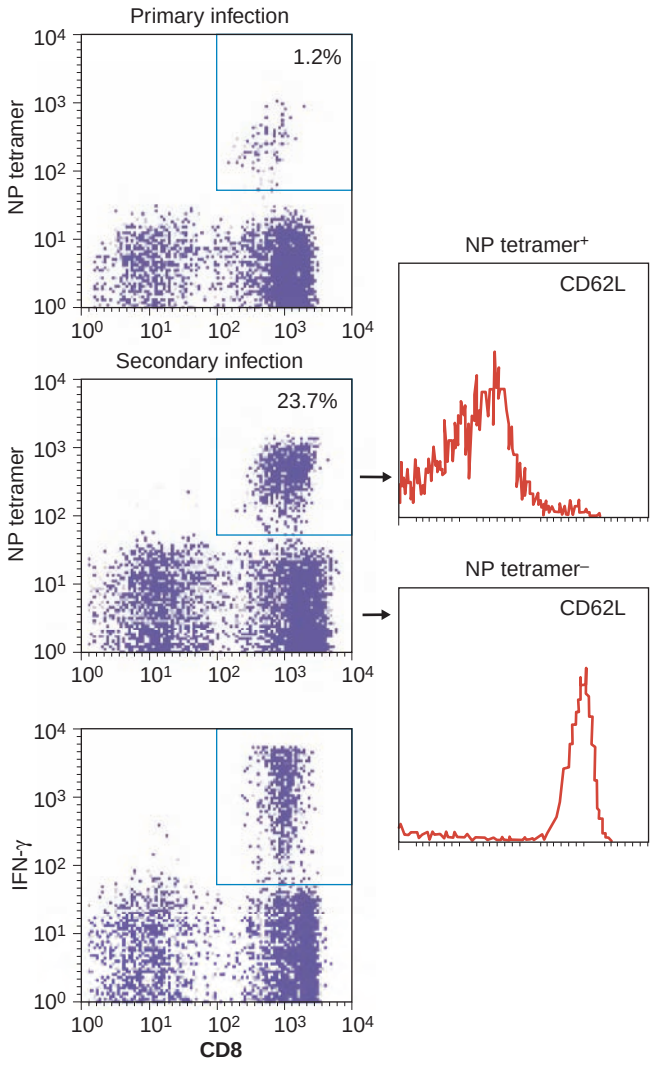


FIG. 12.3 Monitoring a Virus-Specific Cytotoxic T-Lymphocyte (CTL) Response. A fluorescent-labeled tetramer complex comprising a major histocompatibility complex (MHC) class I molecule and a virus-specific nucleoprotein *(NP)* antigen peptide is used to detect NP-specific CD8 T cells in the spleens of mice infected with influenza virus. During the primary response, virus-specific CTLs expand in the draining lymph node of the lung and later exit to the lung and spleen *(upper left)*. After the infection is resolved, a small population of virus-specific memory CTL resides in the spleen. A second infection with influenza results in very high numbers of virus-specific CTLs in the spleen *(middle left)* and the lung. Specific CTLs can be re-stimulated with the NP peptide in vitro, inducing interferon *(IFN)-β* secretion that is detected by intracellular staining with a labeled anti-IFN-β antibody *(lower left)*. Staining with an antibody to CD62L shows the typical profile of effector CD8 T cells in the NP-tetramer–positive CTLs *(upper right)* and of naïve CD8 T cells in the tetramer-negative population *(lower right)*.

TABLE 12.2 Human and Mouse Natural Killer Cell Receptors (Partial List)

Receptor	Species (H, Human; M, Mouse)	Function	Ligands (H, Human; M, Mouse)
KIR2DL1	H	Inhibitory	HLA-C (C2)
KIR2DL2	H	Inhibitory	HLA-B, C (C1, C2)
KIR2DL3	H	Inhibitory	HLA-B, C (C1)
KIR2DL4	H	Activating	HLA-G
KIR2DL5	H	Inhibitory	?
KIR3DL1	H	Inhibitory	HLA-A, B (Bw4)
KIR3DL2	H	Inhibitory	HLA-A (A3, A11)
KIR3DL3	H	Inhibitory	?
KIR2DS1	H	Activating	HLA-C (C2)
KIR2DS2	H	Activating	HLA-C (C1)
KIR2DS3	H	Activating	?
KIR2DS4	H	Activating	HLA-A, C (C1, C2, A11)
KIR2DS5	H	Activating	HLA-C (C2)
KIR3DS1	H	Activating	HLA-F
LILRB1	H, M	Inhibitory	HLA-class I
CD94/NKG2A	H, M	Inhibitory	H; HLA-E M; Qa-1b
CD94/NKG2C, E	H, M	Activating	H; HLA-E, M; Qa-1b
NKG2D	H, M	Activating	H; MICA/B, ULBP1–4, M; H60, MULT1, RAE1
CD16 (FcγRIII)	H, M	Activating	Immune complexes
CD27	H, M	Activating	CD70
CD244 (2B4)	H, M	Activating/Inhibitory	CD48
Ly49A-C, E-G, I-O	M	Inhibitory	MHC class I
Ly49D	M	Activating	H-2D^{d}
Ly49H	M	Activating	MCMV m157
Ly49P	M	Activating	H-2D^{k}/MCMV m04
KLRG1	H, M	Inhibitory	E-, R-, N-cadherins
NKR-P1A	H	Inhibitory	LLT1 (CLEC2D)
NKR-P1A, B, C, E, F	M	Activating/Inhibitory	Clr family
NKR-P1B, D	M	Inhibitory	Clr-b, Clr-g
PILRα/PILRβ	M	Activating/Inhibitory	O-glycosylated CD99
NKp46	H, M	Activating	Pathogen proteins; viral, parasite, fungal. Heparan sulfate glycosaminoglycans
DNAM-1 (CD226)	H, M	Activating	CD112, CD155

host defense. Human NK cells can be divided into subsets based on the expression of CD16 (FcγRIII) and the adhesion molecule CD56 (Fig. 12.4).[17] CD56^{dim} NK cells represent around 90% of peripheral blood NK cells, are CD16^{+}KIR^{+}, and display greater cytotoxicity, especially amongst the CD57^{+} subset. In contrast, their less mature counterparts, CD56^{bright} NK cells, show greater proliferative potential and cytokine production and are the principal NK cell population in secondary lymphoid organs. NK cell progenitors can be found in cord blood and lymph nodes and develop into CD56^{dim} cells via CD56^{bright} intermediaries; this differentiation is also dependent on IL-15.[17] One important limitation in understanding the functional role played by NK-cell subsets has been the inability to identify populations corresponding to the CD56^{bright} and CD56^{dim} populations in rodents, as this molecule is not expressed in mouse NK cells. NK-cell heterogeneity, based on the differential expression of CD11b and CD27, has been dissected in mice. Developing NK cells in bone marrow are CD27^{+}CD11b^{dim}, whereas

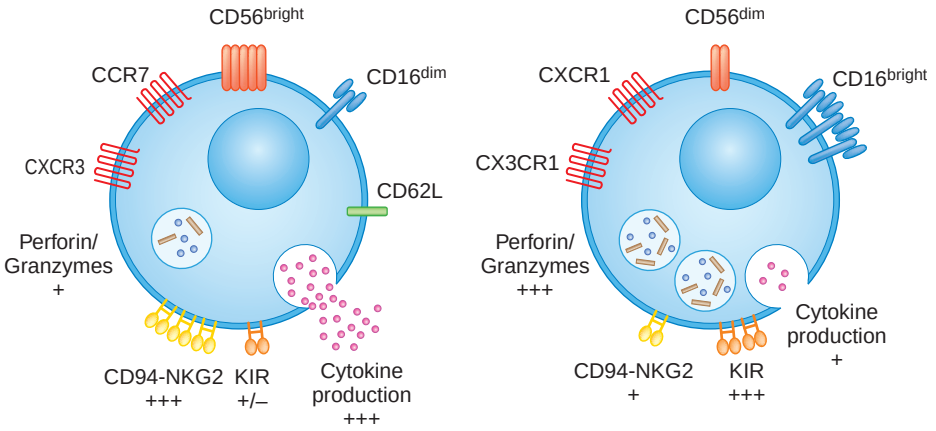


FIG. 12.4 Schematic Representation of Human Natural Killer (NK) Cells. The human NK-cell subsets show distinct receptor expression and effector functions. $CD56^{bright}$ NK cells produce high levels of cytokines and have patterns of chemokine and homing receptor expression that distinguish them from $CD56^{dim}$ NK cells. $CD56^{dim}$ NK cells express high levels of killer immunoglobulin-like receptor *(KIR)* and cytotoxic activity. The relative level of receptors and effector molecules is indicated on an arbitrary scale, with +/– being weak and +++ being strong expression.

these cells differentiate into $CD27^{+}CD11b^{bright}$ and $CD27^{-}CD11b^{bright}$ stages in the spleen. Interestingly, $CD27^{+}$ cells are found in lymph nodes, whereas the $CD27^{-}$ NK cells are localized predominantly in peripheral blood and the lung.[15] Comparative studies of the transcriptomes of human and mouse NK cells have confirmed the similarities between the human and mouse subsets (human $CD56^{bright}$ NK cells correspond to mouse $CD27^{+}$ NK cells, whereas human $CD56^{dim}$ NK cells are the equivalent of mouse $CD27^{-}$ NK cells).[18]

CYTOKINE REGULATION OF NATURAL KILLER–CELL ACTIVATION, FUNCTION, AND HOMEOSTASIS

Mature NK cells also respond to other cytokines in addition to IL-15, with their role being more important for function than homeostasis.[19] IL-2R shares its β and γ subunits with the constitutively expressed IL-15R (Chapter 14). The dimeric IL-2Rβγ is able to respond to the high levels of IL-2 supplied in vitro to induce proliferation of NK cells. IL-2 activation will also induce the IL-2Rα chain to complete the high-affinity trimeric receptor. IL-4 activates human NK cells and promotes the proliferation of a fraction of NK cells characterized by their ability to produce IL-13.

The cytokines IL-12 and IL-18 also have profound effects on NK-cell function.[20] IL-12 and IL-18 are produced by macrophages and DCs during inflammatory immune activation, such as viral infection and within the tumors. Although NK cells are present in IL-12– or IL-18–deficient mice and only marginally reduced in double mutant animals, cytotoxic activity and the ability to proliferate and produce IFN-γ in response to infections, such as mouse CMV (MCMV) infection, is dramatically impaired. Cultivation of NK cells in IL-12 and/or IL-18 induces short-term activation, proliferation, cytotoxicity, and IFN-γ production, whereas longer cultures produce more specialized cytokine-producing cells, sometimes referred to as cytokine-induced memory NK cells.[21]

Another NK cell–activating cytokine is IL-21, a common γ chain family cytokine produced by CD4 T cells. Mouse NK cells treated with IL-21 display a broad-spectrum increase in cytotoxic function and produce cytokines, including IFN-γ and IL-10,[19] whereas IL-21 also promotes the proliferation of human NK cells. Of note, mice treated with IL-21 show a marked increase in NK cell–mediated tumor rejection, highlighting the potential for use of this cytokine as an anticancer therapeutic.[22]

NATURAL KILLER–CELL RECEPTORS

NK cells differ from CTLs in that they do not require the expression of MHC class I to recognize target cells. In fact, the reintroduction of allogeneic MHC class I molecules into previously susceptible cell lines confers resistance to NK cell–mediated killing. These observations led to the missing-self hypothesis, which proposes that NK cells survey tissues for the usually ubiquitous MHC class I expression and react against cells that do not express it.

KEY CONCEPTS

Natural Killer–Cell Receptors

Inhibitory receptors:
- Recognize mostly major histocompatibility complex (MHC) class I ligands with high affinity
- Signal via immunoreceptor tyrosine-based inhibitory motifs (ITIMs)
- Recruit phosphatases (SHP and SHIP) to prevent a cytotoxic response
- Required for natural killer (NK)–cell licensing

Activating receptors:
- Do not bind MHC class I molecules with high affinity
- Ligands include viral molecules and stress-induced proteins
- Signal via immunoreceptor tyrosine-based activation motifs (ITAMs)
- Use several signaling adaptors, including DAP12

The missing-self hypothesis, although altered over time to encompass other observations, has been extremely useful in providing a predictive framework by which to investigate NK-cell receptors and the recognition of target cells (Fig. 12.5).

In the past two decades, a large number of NK-cell receptors and their ligands have been identified, which are classified into either activating or inhibitory types (see Table 12.2).[23]

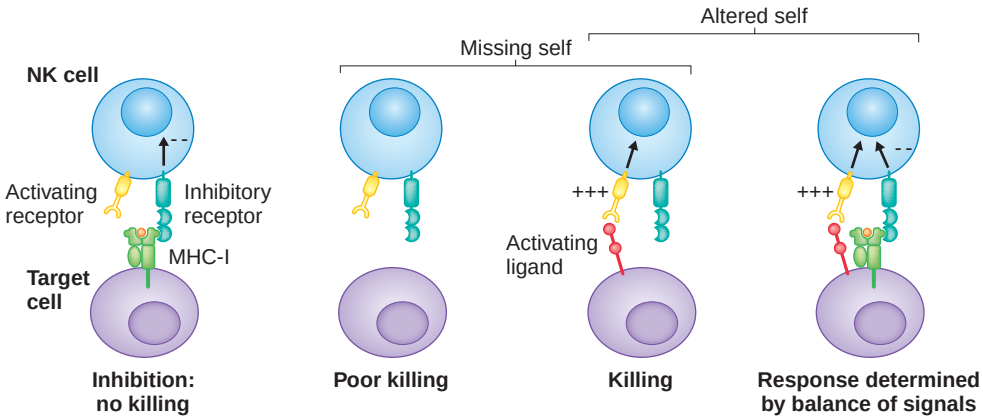


FIG. 12.5 Natural Killer *(NK)*–Cell Recognition of Target Cells. NK cells express inhibitory receptors for major histocompatibility complex class I *(MHC-I)* and activating receptors for a variety of cellular, viral, and stress-induced ligands that alter the outcome of encounter with a target cell. The missing-self hypothesis initially predicted that NK cells would be activated to kill in the absence of MHC-I inhibition. However, an activating signal is also required. This activating signal can be provided by cellular ligands or by viral or stress-induced proteins, termed *altered self*. In the presence of both inhibitory and activating signals, the outcome is determined by quantitative differences in signal strength between the two.

Although many of these NK-cell receptors recognize MHC class I molecules, as predicted by the missing-self hypothesis, there are many other classes of ligands. Interestingly, although this strategy of target recognition is conserved in all mammals tested, in rodents and humans, the receptors have evolved from two independent gene families, killer immunoglobulin-like receptors (KIRs) in humans and the Ly49 family in mice (see Table 12.2).

Natural Killer–Cell Receptor Signaling

The signals derived from NK receptors are defined as *inhibitory* or *activating* in terms of their effect on NK-cell function. Most characterized inhibitory receptors carry an immunoreceptor tyrosine-based inhibitory motif (ITIM) in their intracellular domain. Ligation of ITIM-containing receptors causes tyrosine phosphorylation and the recruitment of a variety of phosphatases, including SHP and SHIP, that act to dampen downstream signaling pathways and NK-cell effector functions.[23] In contrast, most activating receptors use immunoreceptor tyrosine-based activation motifs (ITAMs) to transduce stimulatory signals. Engagement of an ITAM-containing receptor results in tyrosine phosphorylation and recruitment of adaptor molecules, including FcεR1γ, CD3ζ, or DAP12/DAP10. The best-characterized activating receptor is CD16, an Fc receptor that binds IgG and is responsible for the antibody-dependent cellular cytotoxicity (ADCC) of human NK cells. CD16 recruits FcεR1γ and CD3ζ, which, in turn, attract the tyrosine kinases syk and ZAP70. These molecules then promote effector functions via multiple signal-transduction pathways.

NATURAL KILLER RECEPTORS THAT RECOGNIZE MAJOR HISTOCOMPATIBILITY COMPLEX CLASS I MOLECULES

NK cells recognize a wide variety of MHC class I molecules of both classic and nonconventional types. The receptors providing this recognition fit broadly into the immunoglobulin-like and lectin-like superfamilies. They show significant differences between mice and humans.

Killer Cell Immunoglobulin-Like Receptors in Humans

The *KIR* genes are a family of 14 coding genes that are physically linked on chromosome 19.[24] The locus shows a high degree of variation in humans, with both the number of *KIR* genes varying between individuals and extensive allelic polymorphisms. As expected, the inhibitory KIRs contain an ITIM, whereas activating KIRs utilize DAP12 to transduce signals from an ITAM.[23] Surface expression of KIRs is acquired with maturation, driven by IL-15, and as such, NK cells lacking KIRs or expressing combinations of multiple activating and inhibitory KIRs can be found in human blood. Similar to most other NK-cell receptors, some T cells can also express KIRs after activation, in particular in response to IL-15.

KIRs recognize the human MHC class I molecules human leukocyte antigens (HLA, see Table 12.2).[23,24] The specificity of the inhibitory KIRs has been extensively characterized, with, for example, the different KIR2D members that collectively recognize all known alleles of HLA-C. Activating KIRs also bind to HLA molecules, but at a much lower affinity than their inhibitory counterparts. KIR2DL4 is the most evolutionarily distinct member of the family and appears to be expressed in all activated NK cells in culture and on the $CD56^{bright}$ subset in peripheral blood. KIR2DL4 also has some distinct structural features and may act as an unconventional activating receptor binding HLA-G.

There are epidemiological data implicating particular KIRs in a variety of autoimmune pathologies and viral responses.[25] For example, individuals with KIR2DS2 and some HLA-C alleles are predisposed to rheumatoid arthritis with vascular complications. Conversely, individuals with human immunodeficiency virus (HIV) infection who are homozygous for HLA-Bw4 progress to acquired immunodeficiency syndrome (AIDS) more slowly than those with other haplotypes, especially when they have the *KIR3DS1* gene. This activity suggests that KIRs recognize an HIV-associated peptide in the context of HLA-Bw4.[26]

The Ly49 Family in Rodents

In a remarkable example of the power of selection to shape the evolution of the immune system, a multigenic locus with

functional properties almost identical to those of the *KIR* genes evolved independently in rodents.[23] Mice, which have only two *KIR* genes of unknown function, have an analogous cluster of type II transmembrane-spanning lectin-like genes called the Ly49 family on chromosome 6. This family consists of more than 20 members and is highly polymorphic between mouse strains. Like KIR molecules, Ly49 receptors are also variegated in their expression in NK cells. *Ly49* genes encode activating and inhibitory receptors that bind to MHC class I molecules and signal via ITAMs and ITIMs, respectively (see Table 12.2).[23] Ligand-binding studies have revealed that inhibitory Ly49 receptors, such as A, C, and I, function to prevent self-reactivity by NK cells by binding to MHC class I molecules. The function of the activating receptors has proven more difficult to elucidate. Ly49D is known to have high affinity for the MHC class I H-2D^{d} allele and is involved in rejecting bone marrow allografts expressing H-2D^{d}, but the function of Ly49D in the normal immune response is unclear. For one activating receptor, Ly49H, the physiological function is known, as it recognizes the m157 molecule of MCMV and is important in early viral control.[1]

CD94/NKG2 Family

Unlike the KIR or Ly49 families, the CD94/NKG2 complex is present both in rodents and humans. CD94/NKG2 receptors recognize nonclassical MHC class I ligands, such as HLA-E in humans and Qa1^{b} in mice.[25] A single *CD94* gene is physically linked to four *NKG2* (A, C, E, and a truncated F) genes in humans. CD94 is found on the cell surface, either alone or with activating (C, E) or inhibitory (A) forms of NKG2. Interestingly, both the activating and the inhibitory complexes recognize HLA-E, which presents predominantly the leader peptides of other HLA molecules, but not HLA-E itself.[24] This system may provide a mechanism for NK cells to monitor the expression of multiple MHC class I proteins using HLA-E as a surrogate. Human NKG2C-expressing NK cells are enriched in individuals infected with CMV and display fine specificity towards different CMV strains and are often termed *adaptive NK cells* given their ability to rapidly proliferate when stimulated with CMV peptides presented by HLA-E.[24]

NKG2D

NKG2D, which is only distantly related to the NKG2 family, is a single, nonpolymorphic gene that is expressed on all NK cells.[27] In the mouse, NKG2D signals through both DAP12 and another adaptor, DAP10, whereas human NK cells use only DAP10. Activation of NKG2D using a specific antibody results in enhanced cytotoxicity and cytokine secretion. The ligands of NKG2D are a family of molecules with structural similarity to MHC class I proteins, including MICA/B in humans and RAE-1 in mice. These ligands represent a diverse array of sequences, yet all bind with high affinity to NKG2D.[27] Interestingly, the MIC family in humans is highly polymorphic, suggesting that in this receptor–ligand interaction the diversity comes from ligands rather than receptors. Transfection of otherwise resistant tumor cells with NKG2D ligands restores the susceptibility of these cells to NK-cell cytotoxic function. The key to understanding NKG2D function lies in the fact that the ligands are inducible and provide a mechanism for NK cells to detect stressed tissue, such as virally infected or malignant cells, a phenomenon that has been termed *altered self* (see Fig. 12.5).[27] Conversely, certain cancers have been found to downregulate or shed NKG2D ligand expression to evade NK cell immunity. Thus, therapeutic approaches to modulate this pathway in human cancer are being investigated, including antibodies targeting the proteolytic region of NKG2D ligands to block their shedding.[28]

NATURAL KILLER–CELL RECEPTORS THAT RECOGNIZE NON–MAJOR HISTOCOMPATIBILITY COMPLEX I MOLECULES

Beyond the multiple systems of activating and inhibitory receptors that NK cells have evolved to recognize MHC class I molecules and their structural variants, they also have several other receptor families that bind non–MHC class I ligands.[23–25] These include the NKR-P1 family, which is a polymorphic multigene family in rodents consisting of activating (A, C, F) and inhibitory (D) forms, but consists of only a single member in humans, whose activity is inhibitory. The ligands for some family members have been recently reported and are themselves lectin family receptors, including Clr-b and Clr-g in mice and LLT1 in humans. Another NK receptor that is conserved between species is CD244 (2B4, SLAMF4), a pan-NK-cell–expressed molecule whose ligand is CD48 (SLAMF2). SLAM family members are able to switch between inhibitory and activating signaling by recruiting distinct intracellular adaptor proteins. A number of additional activating receptors exist, including the natural cytotoxicity receptors NKp30, NKp44, and NKp46. NKp30 and NKp46 are broadly expressed on human NK cells, and NKp46 is expressed on all mouse NK cells.[29] NKp44 is specifically expressed on activated human NK cells. These receptors can be activated by antibody cross-linking, and although their ligands are unknown, there is evidence that NKp46 binds many pathogen-derived ligands, including hemagglutinin on influenza virus–infected cells and the extracellular matrix component heparan sulfate glycosaminoglycans. Ligands for NKp46 have also been detected on numerous tumor cell lines and NKp46-deficient murine NK cells are poorer at killing these tumor cells, indicating that NKp46 has a direct role in NK-cell activation to various pathogens and cancers.[29]

NATURAL KILLER–CELL LICENSING AND SELF-TOLERANCE

The plethora of MHC class I–binding inhibitory receptors, as proposed by the missing-self hypothesis, explains the influence of MHC class I on NK-cell lytic function against tumors. However, how self-tolerance is achieved has been less clear. The initial theory to explain self-tolerance was the "at least one receptor" model, which proposed that NK cells must express at least one self–MHC class I inhibitory receptor. A second model suggested that the receptor repertoire is shaped by selection by the specific MHC haplotype and the presence of self-ligands. The observation that NK cells are not autoreactive in the absence of any inhibitory ligands (MHC class I–deficient mice) and are actually poor killers suggests that the situation is more complex than these models allow. A concept termed *licensing* was proposed to account for these observations.[30] Under this model, NK cells are initially unresponsive or "unlicensed" and acquire functional competency through binding of at least one inhibitory receptor before they can be activated and display cytotoxic function. An alternative "disarming" model proposes that all NK cells are initially responsive but that chronic stimulation by

normal cells renders these cells unresponsive, or "anergic," unless the stimulation is opposed by MHC class I–specific inhibitory receptors. More recently a third "rheostat" model has also been proposed, whereby the number and affinity of inhibitory receptors it expresses tune NK-cell reactivity in a quantitative manner.[30] Regardless of the exact tolerance mechanism, it is interesting that unlike with T cells, there is no evidence for clonal deletion of autoreactive NK cells.

SPECIFIC NATURAL KILLER–CELL FUNCTIONS

The ability to separate NK cells from T cells both phenotypically and genetically has greatly enhanced the understanding of their functions. Although some cytotoxic and immunomodulatory capacities overlap with those of T cells, it is also apparent that some functions of NK cells are unique. Specific examples of NK-cell functions are discussed below.

Control of Viral Infections

NK-cell activity rises early in the course of viral infection, partly driven by the release of IL-12, -18, and IFN-α, which stimulates activation.[31] The evidence that NK cells are essential for host defense against viruses comes directly from patients and mice lacking NK-cell function and indirectly from viral strategies to avoid NK-cell recognition. Human patients with selective deficiencies in NK cells show a pronounced susceptibility to recurrent severe infections, especially with HSV and CMV.

A powerful example of the role of NK cell–activating receptors in viral control is NK cell–mediated resistance to MCMV. Mouse strains that lack Ly49H are highly susceptible to MCMV, leading to uncontrolled viral replication and death. Importantly, this protection is mediated by the recognition of the m157 protein of MCMV by Ly49H. The rapid accumulation of $Ly49H^+$ NK cells during MCMV infection is the first example of clonal expansion of NK cells in a manner similar to that of B and T cells.[1] These studies also found compelling evidence for the maintenance of a long-lived NK-cell "memory"-like population (described below), blurring the line between innate (NK cells) and adaptive (B and T cells) immune cells.

As mentioned earlier, there is also evidence that NK cells have direct effects on the progression of HIV infection.[32] NK cells are able to lyse HIV-infected target cells either directly or by ADCC. Despite this ability, NK-cell responses are impaired in patients with HIV infection, as infected T-cell blasts selectively downregulate some HLA genes to avoid CTL activity but remain resistant to NK-cell cytotoxicity. These findings are supported by studies that show that high-risk, but uninfected, individuals appear to have increased NK-cell activity and that the combination of the expression of HLA-Bw4 and the *KIR3DS1* gene is associated with delayed progression to AIDS. Finally, HIV viremia induces several functional abnormalities on NK cells, suggesting that this complex virus and NK cells interact at multiple levels.[32]

Control of Malignant Cells

NK cells were named as such during the early 1970s for their potent ability to kill leukemia cell lines and since then have been the focus of immunotherapies to enhance cancer lysis and tumor regression in humans. Evading the immune system is a hallmark of cancer, but the contribution of NK cells in tumor immune surveillance has been difficult to assay, as specific NK cell–deficient mouse models have only recently been generated. There is abundant evidence that NK cells can reduce tumor burden in animal models, and that the administration of cytokines that enhance NK-cell function or number or those that induce IFN-α production, are protective against metastasis.[20]

Melanoma is at the forefront of immunotherapy research in patients since it has high immune infiltration. Bulk and single-cell transcriptomic data have been used to define an NK cell gene signature that infers the relative frequency of tumor-resident NK cell infiltration within metastatic melanoma samples. This study found that greater abundance of tumor-resident NK cells was strongly associated with long-term overall survival in patients with melanoma, and this appears also to be true for head and neck squamous cell carcinoma and lung adenocarcinoma.[31] In light of this, many clinical trials have been conducted to assess either cytokine treatment or the injection of ex vivo cultured healthy or tumor-resident NK cells.[20] Unfortunately, the high doses of IL-2 and IL-15 required for efficacy are relatively toxic, and transferred NK cells have proven difficult to target to tumors. Despite this, some successes have been demonstrated for melanoma, leukemia, and kidney cancer. Numerous immunotherapy trials are ongoing that aim to exploit and boost NK cell–mediated tumor killing, including antibodies against inhibitory receptors on NK cells, modified cytokines, bi- or trispecific antibodies that bind NK cells and tumor antigens, tumor-specific antibodies, in vitro expanded patient-derived NK cells, and irradiated NK cell lines, to name a few.[31]

Role of Natural Killer Cells in Hematopoietic Stem Cell Transplantation

CLINICAL PEARLS

Exploiting Natural Killer Cells in Leukemia Therapy

- Hematopoietic stem cell transplantation requires donor and recipient human leukocyte antigen (HLA) matching to reduce graft-versus-host disease (GvHD) mediated by transplanted cytotoxic T lymphocytes (CTLs).
- Haploidentical donors and recipients (those that share only one HLA haplotype) represent 50% of unrelated transplants and undergo a stronger conditioning regimen to avoid GvHD.
- Alloreactive natural killer (NK) cells are present after haploidentical transplant and provide a potent graft-versus-leukemia (GvL) effect in animal models.
- Transplantation from NK cell–alloreactive donors controls leukemia relapse and improves engraftment without causing GvHD.

In humans, allogeneic bone marrow transplantation can cure leukemia through the reaction of donor CTLs in the graft against the residual leukemic cells. These transferred T cells also mediate GvHD. The need to prevent GvHD as a result of strong immunosuppression is the major cause of transplantation failure because of infection and cancer relapse. It has been proposed that obtaining the transplant from a haploidentical donor (identical at one HLA haplotype and fully mismatched on the other, for example, a parent) provides allogeneic NK cells with an HLA haplotype that supplies more KIR ligand than a matched recipient would provide. Hence it would yield a stronger graft-versus-leukemia (GvL) effect.[33] Indeed, mice treated with alloreactive NK cells tolerate 30 times the lethal dose of mismatched bone marrow cells without developing GvHD, and alloreactive NK cells eradicated human acute myeloid leukemia (AML) transplanted into immune-deficient mice. Retrospective

studies on patients with AML and acute lymphoblastic leukemia who received haploidentical grafts revealed that transplants with alloreactive NK cells showed better engraftment and GvHD protection and less relapse.[33]

These studies suggest that therapeutic treatment with alloreactive NK cells will be effective in eliminating residual cancer cells following frontline treatments or in preventing cancer relapse. Feasibility studies have shown that the production of clinical-grade cultured human NK cells is possible and that the transferred cells persist for some time in the patient. An alternative strategy is the use of monoclonal antibodies directed against inhibitory KIRs. Despite the strong preclinical data for the strong inhibitory effect of KIR2DL1, KIR2DL2, and KIR2DL3 signaling on NK cells when bound to HLA-C-expressing tumor cells,[34] a phase II clinical trial of a KIR2DL1/L2/L3 blocking antibody in AML failed to show any benefit to patients, suggesting that a stronger basic understanding of NK cell target recognition is still required.

ON THE HORIZON

Clinical Trials of Transferred Alloreactive Natural Killer Cells/Chimeric Antigen Receptors

The success of chimeric antigen receptor (CAR) T-cell therapy in treating relapsed refractory B-cell malignancies has given rise to much interest in applying this approach to natural killer (NK) cells. A clear advantage of CAR NK cells is patient safety, as NK cells are well tolerated in the allogeneic setting and, unlike CAR T cells, do not produce cytokine release syndrome, immune effector cell-associated neurotoxicity syndrome, or graft-versus-host disease (GvHD).[31]

A first-in-human phase I trial of allogenic anti-CD19 CAR NK cells has recently concluded and the encouraging results saw 7 out of 11 patients experiencing a rapid response with very little toxicity reported.[35]

Preclinical data suggest that the antitumor activity and persistence of CAR NK cells can be further enhanced by genetic deletion of checkpoint genes that negatively regulate interleukin (IL)-15 signaling.[36] A new-wave of NK-cell therapies are thus on the horizon.

Natural Killer Cell Memory

During the last decade several studies have described the ability of NK cells to generate a "memory" response to pathogens and antigens. Although the original documentation of NK-cell memory was in a cutaneous contact hypersensitivity model, the concept of NK-cell memory has since been extended to viral responses, including MCMV, influenza, vaccinia virus, HSV, HIV, and vesicular stomatitis virus, where NK cells from mice pre-exposed to the virus were more efficiently activated and prolonged survival from lethal infection compared with naïve NK cells. Although the interaction between Ly49H and the viral protein m157 is required for the memory response elicited by MCMV infection, the NK cell receptor–ligands combinations involved in most other antiviral responses are not established. Inflammation, in general, and especially the cytokines IL-12, IL-18, and IFN-α, are likely to be critical for the generation of memory NK cells.

INTERACTIONS OF CYTOTOXIC T LYMPHOCYTE AND NATURAL KILLER CELLS IN THE IMMUNE RESPONSE

Although studies of CTLs and NK cells in isolation have greatly advanced our understanding of their functions, it is obvious that these immune cells function in a system that depends on numerous interactions between the various cell types at multiple levels. In the case of cytotoxic cells, these immunomodulatory interactions are becoming increasingly appreciated (see Fig. 12.2). In particular, it has become apparent that NK cells and DCs interact specifically to promote some outcomes, such as maturation and priming of NK cells. CTL activation by mature DCs is influenced directly by NK cell–derived IFN-γ and indirectly through the role of NK cells in promoting a Th1 response in CD4 T cells (Chapter 11).[37,38]

Increasing evidence has emerged to implicate DC–NK cell cross-talk in various aspects of the immune response. The interactions of mature DCs and NK cells occur at the site of infection, where DCs provide inflammatory stimuli for NK cells, including cytokines such as IL-12 and IFN-α. IFN-α production by DCs promotes MHC class I upregulation by CTLs and their protection from NK-cell killing during viral infections. The other site of encounter is the lymph node, where, during an immune reaction, NK cells are recruited by chemokines and interact with mature DCs and CD4 T cells to induce a Th1 response. This process also requires IFN-γ production from NK cells (see Fig. 12.2).

NK cell–CTL cellular interactions are also important in generating an immune response to tumors. DCs can be recruited into tumors by NK cell derived–chemokines, to influence CTL priming via antigen presentation and NK-cell function via cytokine secretion, particularly IL-12. IFN-γ produced by NK cells and CTLs is important in the rejection of primary tumors and the formation of CD8 T-cell memory to tumors. It is also likely that killing by NK cells and CTLs provides DCs with increased access to tumor antigens and promotes further adaptive immunity. Using DCs to harness the helper function of NK cells as well as the cytotoxic functions of both CTLs and NK cells offers therapeutic promise and is currently being tested in a variety of cancers.

EVASION OF THE CYTOTOXIC RESPONSE

Viruses

As one principal function of CTL and NK cells is the control of viral infections, it is not surprising that viruses have strategies to interfere with the host response (Chapter 25). The multiplicity of these evasion strategies indicates that this is an essential step for long-term viral persistence.

These strategies include:

Latency. This involves minimizing viral gene expression and thereby avoiding detection. Examples include HSV in neurons, HIV in T cells, and EBV in B cells.

Antigenic variation. Viruses possess the ability to rapidly mutate their genomes and produce escape variants that are no longer visible to CTLs. Such mutations were shown for MCMV infection in mouse and HIV infection in humans.[32]

Infection of immune inaccessible sites. Examples include infection of the central nervous system by HSV or rubella virus.

Production of viral defense molecules (immunoevasins). Many viruses, including adenovirus, CMV, and EBV, interfere with cytotoxic activity by producing proteins that either hinder Fas or TNF-mediated killing or inhibit the function of antiviral cytokines, such as IFN-α. A number of viruses, including EBV, produce homologs of antiapoptotic molecules, such as Bcl2, to inhibit killing by CTLs. Members of the poxvirus family have evolved homologs of the naturally occurring IL-18–binding protein that inhibit IL-18 activity and NK-cell function.

Modulation of molecules involved in recognition. A widely utilized viral strategy to evade the cytotoxic response is to interfere with antigen processing, presentation, or the expression of other molecules required for CTL recognition (Chapter 6).[39] Many viruses, including adenovirus and HIV, downregulate MHC class I expression on the cell surface. This can be achieved by a number of mechanisms. For example, adenovirus type 2 E3 protein forms a complex with MHC class I to prevent antigens from being processed; MCMV gpm152 protein causes retention of the MHC class I molecules in the Golgi compartment; and CMV proteins US2 and US11 promote the rapid degradation of newly synthesized MHC class I complexes.[40] An alternative approach is to interfere with antigen processing, either inhibiting the expression of the TAP protein, as is the case for HSV, or producing proteins that are resistant to antigen digestion by the proteasome, such as the EBNA-1 protein of EBV. This inhibition strategy is not restricted to MHC class I, as both human CMV and MCMV express proteins that inhibit the cell-surface expression of NKG2D ligands.

Tumor Cells

Part of the evidence that CTLs and NK cells function to control malignant cells comes from the lengths tumor cells will go to avoid cytotoxic activity. Conversely, promoting the cytotoxic response either through specific tumor antigens, blocking immune checkpoint inhibitors, or through polyclonal stimulation remain successful strategies in cancer immunotherapy (Chapter 80).

Tumors evade cytotoxic function in a number of ways:

Downregulation or loss of MHC class I expression. This strategy is common in solid tumors, including metastatic melanoma and breast cancer, where MHC class I downregulation accounts for up to 50% of samples. MHC class I downregulation is associated with changes in the regulatory mechanisms controlling antigen presentation and can often be corrected by treatment with cytokines, such as IFN-γ.

Induction of immune checkpoint ligands. One mechanism to suppress the infiltrating antitumor CLT/NK cell response is upregulation of ligands to inhibitory receptors expressed on CTLs/NK cells. Two inhibitory receptors, CTLA-4 and PD-1, are expressed on CTLs and to a less extent on NK cells, and blocking of antibodies against these has revolutionized cancer immunotherapy.[11] In particular, IFN-γ upregulates PD-1 ligands (PD-1L and PD-2L) on tumor cells, and their binding to PD-1 potently inhibits tumor-specific CTL activity. Anti-PD-1 therapy offers a high rate of sustained tumor regression in metastatic melanoma and is currently being applied to a huge range of other cancer types.

Antigenic mutation. Tumors can also avoid CTL activity by antigenic loss. This strategy takes the form of silencing or mutating epitopes that are particularly immunogenic to CTL.

Broad-spectrum tumor-derived immune suppression. Tumors express a variety of membrane-bound and soluble factors that can suppress the immune response, including FasL, which protects the tumor by inducing apoptosis in activated Fas-expressing CTLs. This model is not universally accepted, and a role for FasL in inducing the expression of inflammatory cytokines is also possible. Tumors can also express TGF-β, which acts on CTLs and NK cells to inhibit their metabolism, proliferation, and expression of effector molecules, such as perforin and granzymes. TGF-β also acts on NK cells to downregulate the expression of NKG2D. Pharmaceutical inhibitors of TGF-β signaling are being tested in patients with cancer.[20] In addition, tumors can produce soluble decoy ligands, such as MIC, which suppresses NKG2D function. CD73-mediated adenosine production also plays important immunosuppressive roles in the tumor microenvironment. CD73 is a surface receptor expressed on many tumors and catalyzes extracellular adenosine monophosphate (AMP) into adenosine that can bind its receptor (A2AR) on NK cells/CTLs to suppress their activity. As such, A2AR and CD73 antagonists are being developed, and their trials in cancer are ongoing.[20]

REFERENCES

1. Adams NM, Grassmann S, Sun JC. Clonal expansion of innate and adaptive lymphocytes. *Nat Rev Immunol*. 2020;20(11):694–707.
2. Voskoboinik I, Whisstock JC, Trapani JA. Perforin and granzymes: function, dysfunction and human pathology. *Nat Rev Immunol*. 2015;15(6):388–400.
3. Kabanova A, Zurli V, Baldari CT. Signals controlling lytic granule polarization at the cytotoxic immune synapse. *Front Immunol*. 2018;9:307. doi: 10.3389/fimmu.2018.00307.
4. Rossin A, Miloro G, Hueber AO. TRAIL and FasL functions in cancer and autoimmune diseases: towards an increasing complexity. *Cancers (Basel)*. 2019;11(5).
5. Prager I, Liesche C, van Ooijen H, et al. NK cells switch from granzyme B to death receptor-mediated cytotoxicity during serial killing. *J Exp Med*. 2019;216(9):2113–2127.
6. Groom JR. Regulators of T-cell fate: integration of cell migration, differentiation and function. *Immunol Rev*. 2019;289(1):101–114.
7. Norbury CC. Defining cross presentation for a wider audience. *Curr Opin Immunol*. 2016;40:110–116.
8. Chang JT, Wherry EJ, Goldrath AW. Molecular regulation of effector and memory T cell differentiation. *Nat Immunol*. 2014;15(12):1104–1115.
9. Bedoui S, Heath WR, Mueller SN. CD4(+) T-cell help amplifies innate signals for primary CD8(+) T-cell immunity. *Immunol Rev*. 2016;272(1):52–64.
10. Hashimoto M, Im SJ, Araki K, et al. Cytokine-mediated regulation of CD8 T-cell responses during acute and chronic viral infection. *Cold Spring Harb Perspect Biol*. 2019;11(1):a028464.
11. Sharma P, Allison JP. Dissecting the mechanisms of immune checkpoint therapy. *Nat Rev Immunol*. 2020;20(2):75–76.
12. Attanasio J, Wherry EJ. Costimulatory and coinhibitory receptor pathways in infectious disease. *Immunity*. 2016;44(5):1052–1068.
13. Guedan S, Ruella M, June CH. Emerging cellular therapies for cancer. *Annu Rev Immunol*. 2019;37:145–171.
14. Christophersen A. Peptide-MHC class I and class II tetramers: from flow to mass cytometry. *HLA*. 2020;95(3):169–178.
15. Goh W, Huntington ND. Regulation of murine natural killer cell development. *Front Immunol*. 2017;8:130.
16. Rautela J, Huntington ND. IL-15 signaling in NK cell cancer immunotherapy. *Curr Opin Immunol*. 2017;44:1–6.
17. Freud AG, Mundy-Bosse BL, Yu J, et al. The broad spectrum of human natural killer cell diversity. *Immunity*. 2017;47(5):820–833.
18. Crinier A, Milpied P, Escaliere B, et al. High-dimensional single-cell analysis identifies organ-specific signatures and conserved NK cell subsets in humans and mice. *Immunity*. 2018;49(5):971–986 e5.
19. Brady J, Carotta S, Thong RP, et al. The interactions of multiple cytokines control NK cell maturation. *J Immunol*. 2010;185(11):6679–6688.
20. Guillerey C, Huntington ND, Smyth MJ. Targeting natural killer cells in cancer immunotherapy. *Nat Immunol*. 2016;17(9):1025–1036.
21. Romee R, Schneider SE, Leong JW, et al. Cytokine activation induces human memory-like NK cells. *Blood*. 2012;120(24):4751–4760.
22. Croce M, Rigo V, Ferrini S. IL-21: a pleiotropic cytokine with potential applications in oncology. *J Immunol Res*. 2015;2015:696578.
23. Meza Guzman LG, Keating N, Nicholson SE. Natural killer cells: tumor surveillance and signaling. *Cancers (Basel)*. 2020;12(4). doi: 10.3390/cancers12040952.

24. Parham P, Guethlein LA. Genetics of natural killer cells in human health, disease, and survival. *Annu Rev Immunol*. 2018;36:519–548.
25. Das J, Khakoo SI. NK cells: tuned by peptide? *Immunol Rev*. 2015;267(1):214–227.
26. Scully E, Alter G. NK cells in HIV disease. *Curr HIV/AIDS Rep*. 2016;13(2):85–94.
27. Schmiedel D, Mandelboim O. NKG2D ligands-critical targets for cancer immune escape and therapy. *Front Immunol*. 2018;9:2040.
28. Ferrari de Andrade L, Kumar S, Luoma AM, et al. Inhibition of MICA and MICB shedding elicits NK-cell-mediated immunity against tumors resistant to cytotoxic T cells. *Cancer Immunol Res*. 2020;8(6):769–780.
29. Barrow AD, Martin CJ, Colonna M. The natural cytotoxicity receptors in health and disease. *Front Immunol*. 2019;10:909.
30. Kadri N, Thanh TL, Hoglund P. Selection, tuning, and adaptation in mouse NK cell education. *Immunol Rev*. 2015;267(1):167–177.
31. Huntington ND, Cursons J, Rautela J. The cancer-natural killer cell immunity cycle. *Nat Rev Cancer*. 2020;20(8):437–454.
32. Savoy SKA, Boudreau JE. The evolutionary arms race between virus and NK cells: diversity enables population-level virus control. *Viruses*. 2019;11(10):959. doi: 10.3390/v11100959.
33. Locatelli F, Pende D, Falco M, et al. NK cells mediate a crucial graft-versus-leukemia effect in haploidentical-HSCT to cure high-risk acute leukemia. *Trends Immunol*. 2018;39(7):577–590.
34. Romagne F, Andre P, Spee P, et al. Preclinical characterization of 1-7F9, a novel human anti-KIR receptor therapeutic antibody that augments natural killer-mediated killing of tumor cells. *Blood*. 2009;114(13):2667–2677.
35. Liu E, Marin D, Banerjee P, et al. Use of CAR-transduced natural killer cells in CD19-positive lymphoid tumors. *N Engl J Med*. 2020;382(6):545–553.
36. Zhu H, Blum RH, Bernareggi D, et al. Metabolic reprograming via deletion of CISH in human iPSC-derived NK cells promotes in vivo persistence and enhances anti-tumor activity. *Cell Stem Cell*. 2020:27(2):224–237.
37. Gonzalez H, Hagerling C, Werb Z. Roles of the immune system in cancer: from tumor initiation to metastatic progression. *Genes Dev*. 2018;32(19–20):1267–1284.
38. Bottcher JP, Bonavita E, Chakravarty P, et al. NK cells stimulate recruitment of cDC1 into the tumor microenvironment promoting cancer immune control. *Cell*. 2018;172(5):1022–1037 e14.
39. Christiaansen A, Varga SM, Spencer JV. Viral manipulation of the host immune response. *Curr Opin Immunol*. 2015;36:54–60.
40. Berry R, Watson GM, Jonjic S, et al. Modulation of innate and adaptive immunity by cytomegaloviruses. *Nat Rev Immunol*. 2020;20(2):113–127.

13

Regulatory Immune Cells

James B. Wing and Shimon Sakaguchi

The normal mammalian immune system protects the individual from a myriad of potentially pathogenic microorganisms. However, the immune system must also be tightly regulated to prevent it from attacking self-constituents and thus causing autoimmunity. This is partly achieved through central tolerance and the deletion of T cells that recognize self antigens in the thymus. However, this process is incomplete; that is, some self-reactive T cells are present in the periphery of healthy individuals and are capable of causing autoimmunity. This creates a need for peripheral tolerance mechanisms that control the action of such self-reactive cells.

The mechanisms that suppress harmful immune responses are of interest to both basic and clinical immunologists, as the failure of protective immunity can lead to increased susceptibility to infectious diseases, and loss of self-tolerance may trigger an autoimmune disorder. Furthermore, it is often clinically desirable to enhance an immune response to certain self (or quasi-self) antigens, such as tumor antigens, or to induce immune suppression for the purpose of facilitating organ transplant acceptance. Elucidation of the mechanisms responsible for immune regulation and maintenance of self-tolerance is, therefore, one of the primary goals of current medical immunology.

KEY CONCEPTS

Immunologic Self-Tolerance

Immunologic self-tolerance is actively acquired and maintained throughout life by a series of mechanisms that cooperatively and complementarily operate to prevent the maturation and activation of potentially self-reactive lymphocytes.

The mechanisms include:

- Clonal deletion
- Clonal anergy
- Clonal ignorance
- Dominant suppression

One key feature of the adaptive immune response is that once triggered, it shows essentially the same effector activity, whether the target antigen is a microbe or a self antigen, leading to elimination of the microbe or destruction of self-tissue. To prevent self-destructive immune responses while allowing protective immune responses to nonself antigens, the mammalian immune system has evolved various regulatory contrivances that inhibit the initial generation of potentially harmful self-reactive T and B lymphocytes, termed *central tolerance*, or, after lymphocyte generation, downregulate cellular activation and expansion upon encounter with self antigens, termed *peripheral tolerance*. For T cells, central tolerance is established in the thymus, where many potentially dangerous lymphocytes carrying high-affinity T-cell receptors (TCRs) for self antigens are deleted via negative selection during development. This results in the generation of a peripheral T-cell repertoire that is largely self-tolerant. However, there is abundant evidence that some autoreactive T cells escape thymic deletion, and potentially pathogenic self-reactive T cells are, indeed, present in most individuals. Nevertheless, autoimmune diseases only occur infrequently, indicating that autoreactive T cells are somehow controlled in the periphery. Such peripheral mechanisms of self-tolerance include further deletion of self-reactive T cells, seclusion of self antigen from T lymphocytes, low TCR affinity or lack of costimulation in antigen recognition (clonal ignorance), inactivation of autoreactive T lymphocytes upon encounter with antigen without costimulation (clonal anergy), and active suppression of self-reactive lymphocytes by other lymphocytes (peripheral suppression).[1]

There are various mechanisms of peripheral self-tolerance, and this chapter deals with peripheral suppression mediated by T cells and other suppressive non–T cells. Several types of T cells with regulatory activity have been described, including subpopulations of γδ T cells, natural killer T cells (NKT cells), and CD8 and CD4 T cells (Table 13.1; Fig. 13.1). Some of these suppressive T cells are constitutively produced as a separate lineage in the immune system, whereas others are induced from naïve T cells as a product of a particular mode of antigen stimulation in a particular cytokine milieu. Although it remains to be determined how each cell population is functionally stable and physiologically important, this abundance and apparent redundancy of Treg populations may not be surprising when one considers how essential it is to maintain immune homeostasis and self-tolerance.

This chapter focuses on CD4 regulatory T cells (Treg); in particular, Treg that specifically express the transcription factor Foxp3 and maintain high expression of the interleukin-2 (IL-2) receptor α chain, CD25. $Foxp3^+CD25^+CD4^+$ Treg have been the subject of the majority of recent Treg studies and may have the broadest implication to our understanding of the mechanism of various immunologic disorders. Loss of function or numerical deficiency of Treg function or number can be a primary cause of autoimmune disease, allergy, and inflammatory disorders, such as inflammatory bowel disease (IBD) in humans. Conversely, since these cells can prevent targeting of tumor tissues by conventional CD4 and CD8 T cells, inactivating them is a key goal in cancer immunotherapy. Because of their natural presence in the immune system, they are also a good target for the treatment and prevention of a variety of immunologic diseases.[2]

TABLE 13.1 Subsets of Thymic and Peripheral Foxp3⁺ Tregs and Foxp3⁻ Suppressive T Cells

	Foxp3⁺ Treg	Tr1	Qa-1–Restricted CD8⁺ Treg	NKT Cells	γδ T Cells
Site of generation	Thymus/periphery	Periphery	Periphery	Periphery	Periphery
Marker	Foxp3, CD25, CTLA-4, GITR	IL-10 TGF-β	Nonclassic MHC Ib Qa-1	Invariant TCR chain Va14 (mouse), Va24 (human)	Various subsets Vg5⁺ (mouse) Vg1⁺ (human)
Specificity	Peptide plus MHC class II	Peptide plus MHC class II	Peptide plus MHC class Ib	Glycolipids plus CD1d	Glycolipids plus CD1, Peptide plus MHC class Ib
Target cell	T cells, B cells, APCs, NK cells, NKT cells	T cells	T cells	T cells, APCs	T cells, APCs, epithelial cells
Suppressive mechanisms	Cell-contact Costimulation modification Cytokine production	IL-10 TGF-β	Perforin	IL-10, Th2 cytokines	Lysis, CD95-CD95 ligand pathway, thymosin-b4
Reported suppressive function	Autoimmunity Transplantation Allergy Infection Cancer	Autoimmunity Transplantation Allergy	Autoimmunity	Autoimmunity Transplantation Cancer	Autoimmunity Allergy (dermatitis) Infection

APCs, Antigen-presenting cells; *GITR*, glucocorticoid-induced tumor necrosis factor receptor protein; *ILT*, immunoglobulin transcript; *IL-10*, interleukin-10; *MHC*, major histocompatibility complex; *NKT cells*, natural killer T cells; *TCR*, T-cell receptors; *Tr1*, regulatory type 1 cell; *Th2*, T-helper 2 cell; *Treg*, regulatory T cell; *TGF-β*, transforming growth factor-β.

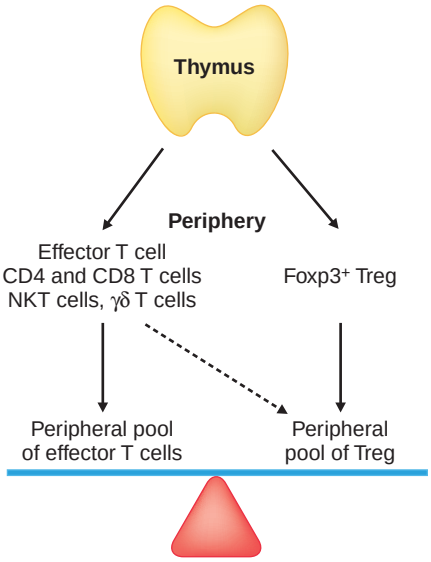


FIG. 13.1 Developmental Pathways of Regulatory T Cells *(Treg).* Treg cells can develop either in the thymus or the periphery and are vital for maintaining tolerance as a counterbalance to effector T cells. Thymically generated Treg express Foxp3 and develop within the thymus following to a specialized combination of T-cell receptor and costimulatory signals. Extrathymic development of CD4⁺ Treg, and CD8⁺ Treg can ensue from a host of different conditions, such as high concentrations of transforming growth factor-β, interleukin-10, or other particular circumstances surrounding antigen priming. The signals that control differentiation of γδ T cells and natural killer T cells *(NKT cells)* to cells with regulatory properties are less well defined.

CD4 REGULATORY T CELLS

The nomenclature used to define Treg has been inconsistent over the years, but recent efforts have been made to standardize the naming of these cells.[3] It is recommended that Foxp3⁺ Treg be clearly separated into thymus-derived tTreg, peripherally in vivo–induced pTreg and in vitro–induced iTreg. Readers should be aware that in the past tTreg have often been referred to as *natural Treg* (nTreg), whereas both pTreg and iTreg were grouped together as *induced* or *adaptive Treg*. Where there is no specific identification allowing confirmation of the cells being specifically tTreg or pTreg, the simple term *Treg* can be used to refer to Foxp3⁺Treg as a group. In this chapter, to avoid confusion, we reserve the term *Treg* for the CD4⁺Foxp3⁺ population rather than other T cells with suppressive function, such as IL-10–secreting type 1 regulatory T cells (Tr1),[4] since they differ in their development, phenotype, and function, and they are thus considered to be of separate lineages (see Table 13.1 and Fig. 13.1).

Thymus-Derived Regulatory T Cells

The first report of autoimmune-preventive, thymus-derived T cells in the normal immune system occurred about 40 years ago when thymectomy on day 3 of life (d3Tx) was shown to cause organ-specific autoimmune diseases, such as oophoritis, in otherwise healthy mice, a finding later attributed to the slightly delayed thymic emigration of tTreg in comparison with effector T cells.[5] Subsequent studies showed that the development of autoimmune diseases could be inhibited if the thymectomized animals were reconstituted with CD4⁺CD8⁻ thymocytes or CD4⁺ splenocytes from histocompatible immune-uncompromised animals. Athymic mice transferred with non-Treg or thymocytes spontaneously develop organ-specific autoimmune diseases, which can be reversed by cotransfer of tTreg from normal adult mice.[2,6] Treg can suppress the proliferation and cytokine production of conventional CD4 or CD8 T cells in vitro. tTreg are thought to arise from T-cell clones with relatively high reactivity to self antigens presented in the thymus.

While for a long time it has been considered that CD25⁺Foxp3⁻ cells in the thymus are the main Treg precursor population, recent evidence also suggests the existence of a second CD25⁻Foxp3⁺ population that forms an alternative

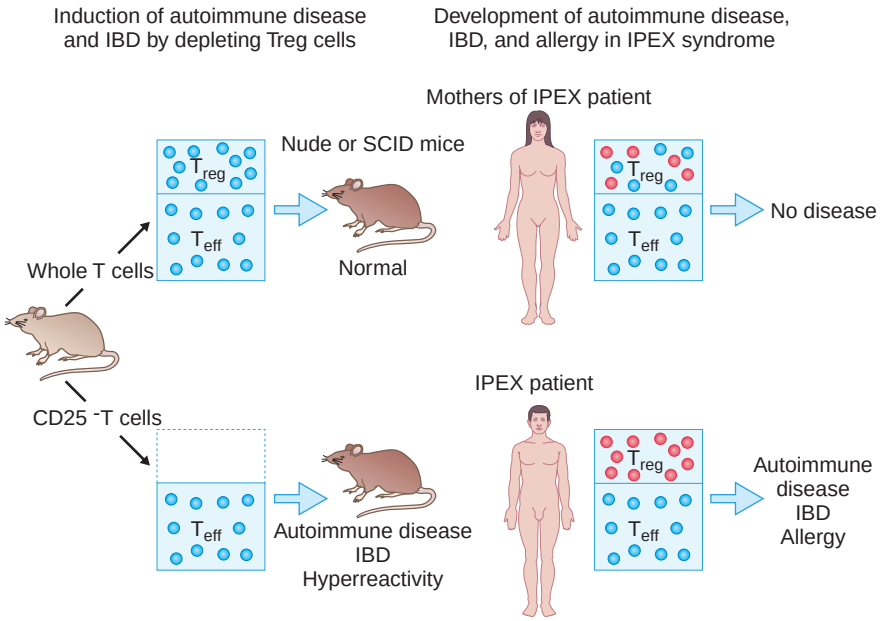


FIG. 13.2 Dominant Self-Tolerance in Rodents and Humans. Transfer of T-cell suspensions depleted of Foxp3$^+$CD4$^+$ Treg induces autoimmune disease and inflammatory bowel disease *(IBD)* and heightens immune responses to nonself antigens in athymic nude or severe combined immunodeficiency *(SCID)* mice *(left)*. Male children are afflicted by infections and immune dysfunction/polyendocrinopathy/enteropathy/X-linked *(IPEX)* syndrome. Their mothers with hemizygous defects of the *FOXP3* gene bear defective and normal Tregs as a mosaic because of random inactivation of the X chromosome in each Treg cell. They are completely normal because normal Treg cells dominantly control the activation and expansion of effector T *(T_{eff})* cells that mediate autoimmune disease, IBD, and allergy *(right)*. *Blue circles* represent intact Treg or T_{eff} cells; *red circles* represent defective Treg cells.

Treg precursor population. These two precursors do appear to result in Treg with no clear difference in their phenotype but they do harbor some differences in their TCR affinity with the Foxp3^{lo}CD25$^-$ population, having generally lower affinity, which suggests a difference in their ability to recognize antigens in the periphery once exported from the thymus.[1]

To pinpoint a specific phenotype for CD4 T cells with regulatory function, surface markers with more restricted expression patterns have been explored. As a high and stable expression of the IL-2 receptor α chain, CD25 has been found to be a useful and specific surface marker of Foxp3$^+$ Treg. Between 5% and 10% of CD4$^+$ T cells express CD25 constitutively in the thymus and periphery of mice. Importantly, transfer of CD4 lymphocytes depleted of CD25$^+$ cells induces autoimmunity in athymic nude mice, whereas cotransfer of CD4$^+$CD25$^+$ cells protects the mice from disease induction (Fig. 13.2). Other markers shown to be associated with Foxp3$^+$ Treg are cytotoxic T-lymphocyte antigen-4 (CTLA-4; CD152) and glucocorticoid-induced tumor necrosis factor receptor protein (GITR). However, these and CD25 are not truly specific markers of tTreg, since conventional T cells upregulate GITR, as well as CTLA-4 and CD25, after activation.[2,6] This dilemma is especially apparent when investigating Foxp3$^+$ Treg in humans, where a significant proportion of CD4$^+$ T cells in peripheral blood express CD25, yet only 2% to 4% of CD4$^+$ T cells, enriched among cells with the highest expression level of CD25 (CD25^{high}), have suppressive properties (Fig. 13.3).[6] The fact that CD25$^+$ Treg are not a discrete population in humans poses a problem both when obtaining cells for experimental purposes and when evaluating their role in a clinical setting. Therefore finding more specific cell-surface markers of CD4$^+$ Treg remains an important goal.

Thymus-Derived Treg Express the Transcription Factor Foxp3

Specific expression of the transcription factor Foxp3 is closely linked with the development and function of Treg.[7] The first hint as to the significance of Foxp3 was given by studies of the Scurfy mutant mouse. This mouse strain suffers from a spontaneous X-linked mutation of the *Foxp3* gene, which leads to fatal lymphoproliferative disease associated with multi-organ infiltrates and early death by 3 to 4 weeks of age in hemizygous males. Similarly, mutations in the human orthologue *FOXP3* are linked to immune dysregulation, polyendocrinopathy, enteropathy, IBD, allergic dermatitis, food allergy, hyperimmunoglobulinemia E, hematological disorders, severe infections and X-linked (IPEX) syndrome, which is an X-linked immunodeficiency associated with organ-specific autoimmune diseases, such as type 1 diabetes (see Fig. 13.2).[2,6] Common features of IPEX syndrome and scurfy mice are deficient levels of both tTreg and pTreg. CD25$^+$CD4$^+$ T cells and CD25$^+$CD4$^+$CD8$^-$ thymocytes specifically express *Foxp3* mRNA in contrast to the cell-surface markers used until now. In addition, other thymocytes/T cells, Th1, or Th2 cells scarcely express *Foxp3* even after stimulation.[7]

Although the majority of Foxp3$^+$ cells in mice reside in the CD4$^+$CD25$^+$ T-cell population, some can be found in the CD4$^+$CD25$^-$ population, particularly in non-lymphoid sites or within the germinal center. Importantly, retroviral transduction of *Foxp3* in naïve CD25$^-$ T cells can convert them to regulatory cells with at least some of the suppressive functions of true Treg. However, although it is essential, it has also become clear that Foxp3 alone is not sufficient to stably maintain the full Treg identity. Another critical factor is the presence of a Treg-type epigenetic pattern in which genes, such as *Foxp3*,

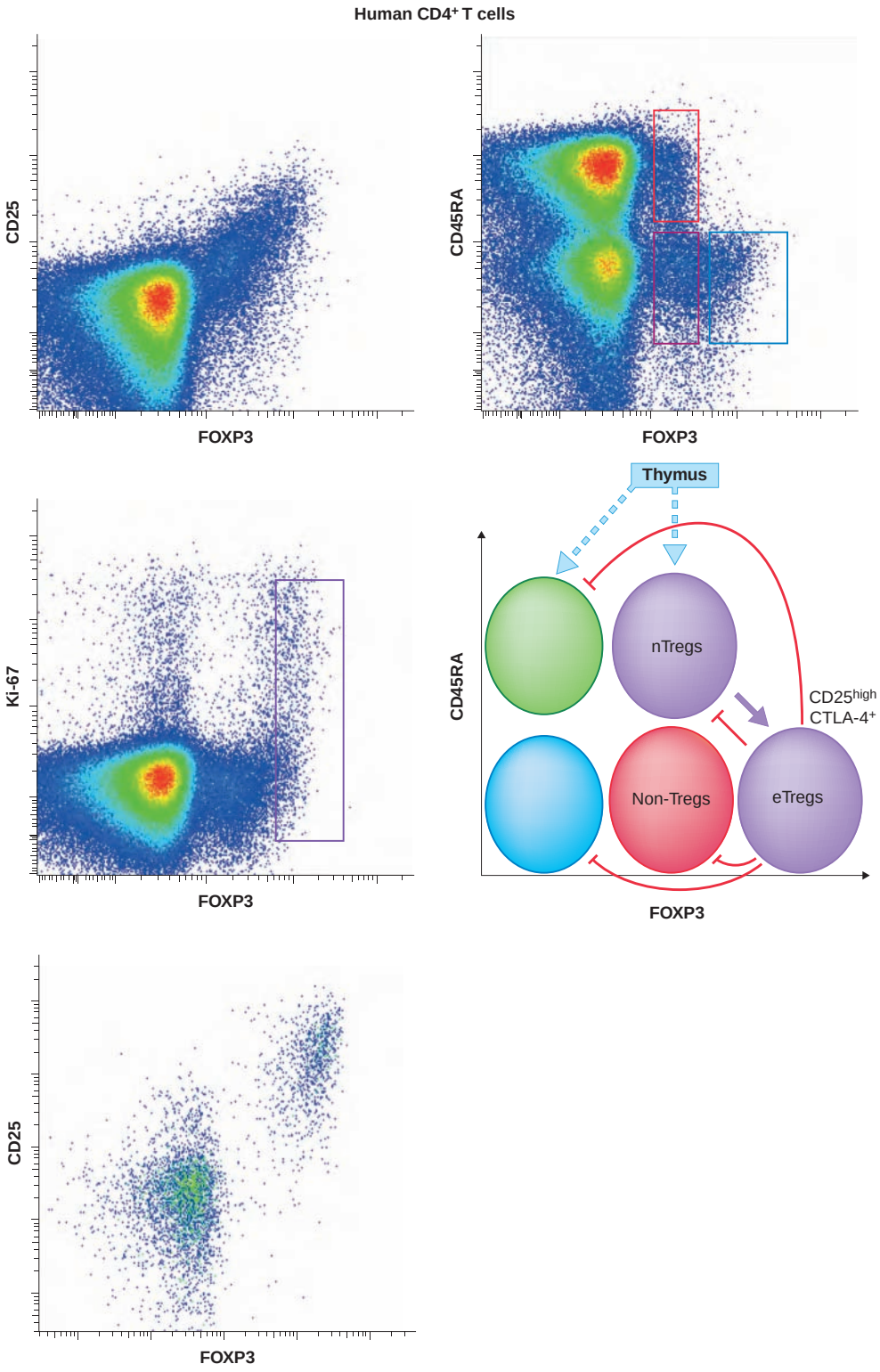


FIG. 13.3 Expression of CD25 and Foxp3 Denotes Treg in Mice and in Humans. In mice, between 5% and 10% of $CD4^+$ T cells express CD25, and among these, almost all are $Foxp3^+$. In addition, a few Foxp3-expressing cells reside among $CD4^+CD25^-$ T cells. In the peripheral blood of humans, expression of surface marker CD25 is well correlated with intracellular FOXP3, especially when the levels of CD25 and FOXP3 are high *(top left)*. CD45RA and intracellular expression of FOXP3 enables the definition of three distinct subsets within $FOXP3^+CD4^+$ T cells: $CD45RA^-$ effector Treg cells with high expression of FOXP3 *(blue box)*, $CD45RA^+$ naïve Treg cells with low expression of FOXP3 *(red box)*, and $CD45RA^-$ non-Treg cells with low expression of FOXP3 *(top right, purple box)*. Effector $FOXP3^{high}$ Treg cells correspond to actively proliferating Treg cells in vivo as they express Ki-67 *(bottom left, purple box)*. The thymus produces $CD45RA^+FOXP3^{low}$ naïve Tregs as well as naïve $CD45RA^+$ non-Treg cells. Upon antigenic stimulation, naïve Tregs differentiate to $CD45RA^+FOXP3^{high}$ effector Tregs, which are potently suppressive and capable of suppressing the proliferation of other Tregs as well as non-Tregs in vitro *(bottom right)*. *CTLA-4*, Cytotoxic T-lymphocyte antigen-4.

CTLA-4, and *GITR*, have stably demethylated CpG residues and a permissive chromatin structure allowing them to be constitutively expressed by Treg. These epigenetic modifications occur independently of Foxp3 expression and maintain a significant proportion of the *Treg* gene expression even in the absence of Foxp3 itself.[8] However, at the same time, Foxp3 expression is still essential, as shown by the severe diseases seen in humans or mice lacking functional Foxp3. A similar pattern of FOXP3 expression can be observed in humans, with most FOXP3$^+$ cells among CD4$^+$CD25^{high} T cells, and a few being CD25$^-$ or CD25^{low} (see Fig. 13.3). However, in humans, but not in mice, low levels of FOXP3 can be transiently induced by TCR stimulation in conventional T cells. These cells can be detected directly in blood as CD4$^+$ T cells with a memory phenotype and a weak expression of FOXP3, but with no suppressive function.[6] However, some true Treg such as CXCR5$^+$ T-follicular regulatory (Tfr) cells may also express lower levels of FOXP3, meaning that great care is required to separate them from non-Treg. As a result, human FOXP3$^+$ Treg with function can be divided into two relatively clearly defined subsets: CD45RA$^+$FOXP3^{low} naïve Treg (nTreg) and CD45RA$^-$FOXP3^{high} effector Treg (eTreg), and a third group of Foxp3^{lo}CD45RA$^-$ cells that contain a mixture of Treg and non-Treg (see Fig. 13.3).

CLINICAL RELEVANCE

Infections and Immune Dysfunction/ Polyendocrinopathy/Enteropathy/X-Linked (IPEX) Syndrome Is a Result of FOXP3$^+$ Treg Deficiency

When the *Foxp3/FOXP3* gene has a loss-of-function mutation, FOXP3$^+$ regulatory T cells (Treg) fail to develop, or Foxp3 protein is dysfunctional, and a fatal autoimmune/autoinflammatory disease develops. This monogenic X-linked disease directly demonstrates how crucial FOXP3$^+$ Treg are for maintaining self-tolerance and immune homeostasis.

Cardinal features of IPEX are:

- Autoimmune diseases (type 1 diabetes, thyroiditis, hemolytic anemia)
- Allergy (dermatitis, hyperimmunoglobulinemia E, food allergy)
- Inflammatory bowel disease

In addition to tTreg, there is abundant evidence from mouse studies supporting the peripheral development of T cells with suppressive properties. For example, Foxp3-expressing CD25$^+$CD4$^+$ Treg, functionally and phenotypically similar to tTreg, can be induced from naïve T cells by in vitro or in vivo antigenic stimulation in the presence of TGF-β.[9] However, it should be noted that in vitro–induced Treg lack the Treg-type epigenetic pattern and so may not be stable Treg, making it important to differentiate between in vivo peripherally induced pTreg and in vitro–induced iTreg.[8] Both murine and human thymus-derived Treg cells express Helios, an Ikaros family transcription factor, and neuropilin-1, whereas most induced Foxp3-expressing CD4 T cells do not. However, some highly activated pTreg and Foxp3$^-$ T cells may express these markers to some extent, making them useful to differentiate between tTreg and pTreg.[9] It has been suggested that tTreg are sufficient for the prevention of widespread autoimmunity but that pTreg have a more specialized function in the prevention of excessive immune responses in mucosal sites, such as the gastrointestinal tract and the lungs.[10]

Maintenance of Foxp3$^+$ Tregs

In addition to TCR interaction, it seems that accessory signals, such as costimulation through CD28-B7 or CD40-CD40L, play an important role in the production of Foxp3$^+$ Treg in the thymus, since animals that lack CD28 or CD40 expression generate only minute numbers of Foxp3$^+$ T cells in the thymus (Table 13.2).[11]

In the periphery, maintenance of Foxp3$^+$ Treg requires antigenic priming and cytokines. It is vital that Foxp3$^+$ Treg encounter specific antigens to remain in the Treg pool. For example, cell transfer experiments in mice thymectomized on day 3 (d3Tx) have demonstrated that Treg from donors of the opposite sex are better at protecting against orchitis or oophoritis compared with Treg from same-sex donors, and that Treg from ovariectomized mice are less competent in preventing oophoritis than those from normal females.[5]

IL-2 is vital for the maintenance of Treg. Thus CD25 is not only a marker but also an indispensable molecule for Treg as a component of the high-affinity IL-2 receptor. Mice genetically deficient in IL-2 or IL-2 receptor α (CD25) or β chain (CD122) develop severe lymphoproliferative disease in multiple organs, resulting in early death.[11] Moreover, genetic deficiencies of CD25 in humans generate a comparable pattern. IL-2–deficient animals have substantially reduced, although not completely depleted, numbers of Foxp3$^+$ T cells, and disease can be prevented by adoptive transfer of normal Foxp3$^+$ Treg (see Table 13.2).

In addition to antigen recognition and cytokines, Foxp3$^+$ Treg require appropriate interactions with antigen-presenting cells (APCs) for their function and survival. Several molecules

TABLE 13.2 Signals With Impact on FOXP3$^+$ Treg Induction, Maintenance, and Suppression

	Development	Maintenance/ Survival	Suppressive Function
Peptide–MHC II interaction	Yes (high affinity)	Yes	Yes, at least initially
CD28	Yes (crucial)	Yes	Not crucial for induction of suppression but high expression on APCs breaks suppression
CD40	Yes	No	No
CTLA-4	No	No	Yes
GITR	No	Modest positive effect	Breaks suppression
TLR ligands	No	Yes	TLR ligands initially break suppression, but this is followed by induction of enhanced suppression
IL-2	Yes (but not crucial)	Yes (crucial)	High levels break suppression
TGF-β	Not required for thymic differentiation but may be involved in peripheral induction	Yes	Yes (not crucial)

APCs, Antigen-presenting cells; *CTLA-4*, cytotoxic T-lymphocyte antigen 4; *GITR*, glucocorticoid-induced tumor necrosis factor receptor protein; *IL-2*, interleukin-2; *MHC II*, major histocompatibility complex II; *TGF-β*, transforming growth factor-β; *Treg*, regulatory T cell; *TLR*, Toll-like receptor.

of cell adhesion and costimulation are important for function and homeostasis of Treg, including CD18/CD11a, GITR, CD28, and CTLA-4.

Effector CD4 T cells differentiate into various subpopulations, such as Th1, Th2, Th17, and Tfh, in response to specific stimulation and cytokine milieus (Chapter 11). It has become clear that Treg somewhat mirror this process and can also differentiate in response to the same stimuli into a subpopulation specialized to control the matching effector population. Consequently, Treg express transcription factors linked to these fates, such as T-bet (Th1), RORγt (Th17), and BCL6 (Tfh), and this allows Treg to gain expression of relevant chemokine receptors, such as CXCR5, in the case of Tfr cells, enabling them to traverse to the germinal center, where they are able to suppress T-follicular helper (Tfh) cells and the germinal center reaction. Similarly, expression of CCR6 and CXCR3 allows a similar process to occur for Th17-Treg and Th1-Treg, respectively. This process lets Tregs respond to the environment, specifically suppressing Th1, Th2, Th17, and Tfh-cell responses.[6] Furthermore, to the development of Treg that specifically mirror effector T-cell subsets, it has also been demonstrated that Treg travel to nonlymphoid sites, such as skin, adipose tissue, and skeletal muscles, to suppress inflammation. Treg from these sites may represent specialized subtypes of Treg with adapted transcriptional signatures, such as adipose tissue Tregs, expressing the transcription factor peroxisome proliferator-activated receptor-γ (PPAR-γ), which regulates fatty acid metabolism and allows their homeostasis and survival.[12]

Another important aspect of Treg responsiveness to the environment is that Tregs are affected by the microbiota (Chapter 22). Recently, it has been demonstrated that certain commensal organisms, such as *Clostridium* spp. in the gut, have large effects on Treg differentiation. In some cases, this may be indirect via induction of TGF-β production, leading to induction of Treg from naïve T cells, whereas in other cases—microbial metabolites, for instance—short-chain fatty acids directly induce Treg proliferation.[9]

Suppressive Function of Foxp3$^+$ Tregs

The standard assay to analyze Treg suppression is to coculture purified fractions of T cells with Treg and then measure proliferation upon antigenic stimulation in the presence of APC. Such assays show that freshly isolated Foxp3$^+$CD25$^+$ Treg cannot suppress T-cell responses in vitro unless they are first stimulated via the TCR. Once activated, they suppress other CD4 and CD8 T cells, irrespective of their antigen specificities.[7] Foxp3$^+$ Treg can inhibit proliferation and cytokine production of naïve T cells. They can also suppress the function of differentiated Th1 and Th2 cells in vitro and reverse ongoing immunopathology, such as colitis, in vivo.[9] In addition to suppressing T-cell function, Foxp3$^+$ Treg can also suppress B cells, NK cells, and NKT cells.

Foxp3$^+$ Treg likely suppress by multiple mechanisms, including both soluble and cell surface–bound mediators (Fig. 13.4). In vitro studies have shown that Foxp3$^+$ Treg need direct cell-to-cell contact with responder cells and that suppression does not occur if Treg are separated from effector T cells by a semipermeable membrane.[2] Moreover, Foxp3$^+$CD25$^+$ Treg are not prominent producers of either IL-10 or TGF-β in vitro. These features differ from those of Tr1 or Th3 cells, which primarily rely on immunosuppressive cytokines, such as IL-10 and TGF-β, for inhibition.

Although Foxp3$^+$ Treg can suppress effector T cells under APC-free conditions in vitro, presumably by absorbing IL-2, it is reasonable that Foxp3$^+$ Treg in vivo control immune responses also by regulating APC. Indeed, by using two-photon laser-scanning microscopy, it has been shown in vivo that although there are limited contacts between Foxp3$^+$ Treg and effector T cells, stable interactions exist between Foxp3$^+$ Treg and dendritic cells (DCs) during ongoing suppression in lymph nodes. One way that Foxp3$^+$ Treg regulate immune responses may be through competition with effector T cells for access to APC. Foxp3$^+$ Treg have, relative to conventional T cells, a more activated phenotype (e.g., high expression of adhesion molecules) in normal individuals, which gives an advantage when engaging with APC and results in prevention of naïve T-cell priming. Interestingly, Foxp3$^+$ Tregs can modify APC function. APC cultured with Foxp3$^+$ Treg downregulate CD80 and CD86 via a CTLA-4–dependent mechanism and become impaired in their ability to stimulate T cells.[13] Additionally, CTLA-4–expressing Treg can induce production of the enzyme indoleamine 2,3-dioxygenase (IDO), which catalyzes tryptophan to the metabolite kynurenine, which represses T-cell responses. Importantly, CTLA-4 expression by Foxp3$^+$ Treg is important for tolerance in vivo, since mice with deletion of this coinhibitory molecule only in Foxp3-expressing cells develop lethal autoimmunity. Other proposed suppressive mechanisms that involve close contact between Foxp3$^+$ Tregs and target cells include surface-bound TGF-β, perforin, and granzyme B (see Fig. 13.4).[2,13]

Although immunosuppressive cytokines are redundant for suppression in vitro, the in vivo situation seems somewhat different. Recently, the immune suppressive cytokine IL-35 has been implicated in Treg-mediated suppression, and both TGF-β and IL-10 expressed by Foxp3$^+$ Treg are important in preventing IBD in mice. Curiously, adoptive transfer of IL-10–deficient CD25$^+$ Treg failed to protect against colitis, although it could inhibit development of gastritis. Furthermore, although perforin- or granzyme B–expressing Foxp3$^+$ Treg are rare in the spleen, they are abundant in a tumor environment. Taken together, Foxp3$^+$ Treg can use several mechanisms of suppression, depending on the local cytokine milieu and the strength and type of immune response.[13]

Tr1 Cells

Some Foxp3$^-$ suppressive T cells have also been identified. The best characterized are IL-10–secreting Tr1 cells, which are present and functional in humans following bone marrow transplantation and in response to allergens (see Table 13.1).

Tr1 cells were initially generated in vitro from CD4 T cells rendered anergic by chronic stimulation in the presence of IL-10, a potent negative regulator of inflammation and lymphoproliferation.[14] T cells so obtained produce a unique pattern of cytokines distinct from Th1 or Th2 cells, with IL-10 as their signature cytokine. Additionally, Tr1 cells secrete some TGF-β, interferon-γ (IFN-γ), and IL-5, but not IL-4 or IL-2. Tr1 cells can be identified by surface expression of CD49b and LAG-3, but lack Foxp3 expression.[14] It remains unclear if Tr1 cells have a master transcription factor but several transcription factors such as c-Maf, IRF4 and EGF2 have a role in their development and function. Also, antigenic stimulation with immature DC (i.e., low levels of costimulatory molecules) or with DC pretreated with IL-10/TGF-β, confers on naïve CD4 T cells an anergic and suppressive phenotype and has further been reported to suppress humoral immunity.[14]

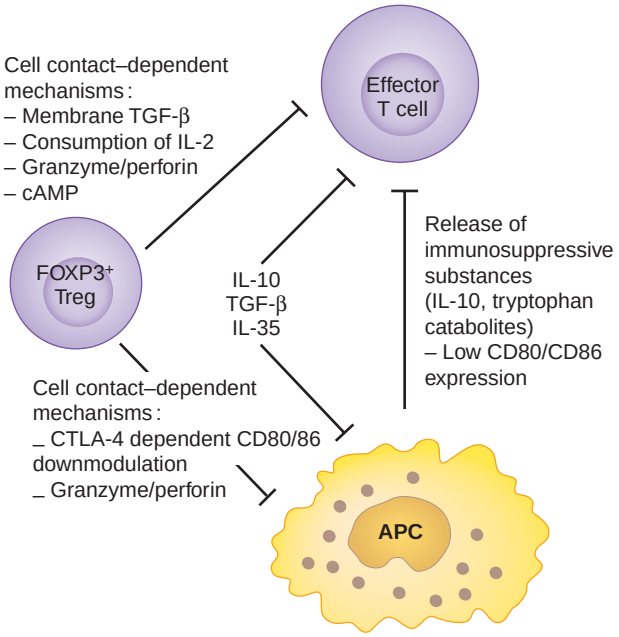


FIG. 13.4 Proposed Suppressive Mechanisms of Foxp3⁺ Regulatory T Cells *(Treg)*. FOXP3⁺CD25⁺ Treg can inhibit many different types of effector cells and can also suppress immune responses at multiple stages, including the initial priming in the lymph node and effector actions at the site of inflammation. The precise mechanism is not known, but numerous theories have been proposed, and it is likely that Foxp3⁺CD25⁺ Treg suppress by several different mechanisms. In vivo Treg can act in a cell contact–dependent manner by competing directly for stimulatory ligands on the antigen-presenting cell *(APC)*, by absorbing essential growth factors, such as interleukin-2 *(IL-2)*, or by directly transmitting an as-yet-uncharacterized negative signal to either the T cell or the APC. Alternatively, Treg can use long-range suppressive mechanisms by means of the cytokines IL-10, IL-35, and transforming growth factor-β *(TGF-β)*. *cAMP*, Cyclic adenosine monophosphate; *CTLA-4*, cytotoxic T-lymphocyte antigen-4; *TGF-β*, transforming growth factor-β.

Tr1 cells also take part in human tolerance. For example, Tr1 cells can be correlated with lack of graft-versus-host disease (GvHD) in bone marrow transplantation, and are also induced after specific immunotherapy in patients with allergies. Data from mice show that Tr1 cells can prevent IBD. Interestingly, Tr1 cells specific for *Escherichia coli* proteins can be isolated from the intestinal mucosa of healthy donors. In pemphigus vulgaris (Chapter 63), Tr1 cells specific for desmoglein-3 can be isolated from pemphigus vulgaris–prone but apparently healthy individuals, whereas patients with pemphigus rarely have such cells. Collectively, Tr1 cells can be induced by autoantigens as well as foreign antigens as a component of the mechanisms that maintain tolerance to both self and nonself.[14]

Treg-of-B cells

Several reports have suggested that repeated contacts with B cells may drive some T cells into a regulatory or suppressive state. These "Treg-of-B cells" produce express CD25 but lack Foxp3 and have IL10-dependent suppressive function.[15] In humans IL10–producing suppressive T cells have also been found among $CD25^+CXCR5^{hi}PD1^{hi}Foxp3^-$ Tfh cells, and these cells, referred to as $CD25^+$ Tf, again appear to have suppressive function dependent on IL-10 and other factors such as CTLA-4; they may have a particular role in suppression of humoral immunity. Initial analysis suggests that $CD25^+$Tf they may not be Tr1 since they lack Lag3 expression.[16] Further work is required to determine the relationship between Tr1, Treg-of-B, and $CD25^+$Tf. They appear to have considerable overlap in phenotype and function and it seems possible that Tconv cells can be driven into a similar suppressive state, with variations in starting cell type and local environment providing some variation in final phenotype.[17]

OTHER SUBSETS OF FOXP3⁻ REGULATORY T CELLS

In addition to CD4 Treg, other types of T cells with immunosuppressive properties have been found; these recognize antigens different from those typically presented to CD4 T cells via MHC class II and may therefore serve to induce tolerance in other settings (see Table 13.1). One example is CD8 T cells with TCR which recognize antigen presented by the mouse MHC class lb Qa-1 molecule (HLA-E in humans; Chapter 5). Qa-1 has limited polymorphisms and can present both foreign and self-peptides. Since Qa-1 peptide complexes on, for example, CD4 T cells can bind to both the inhibitory CD94-NKG2 receptor complex and the TCR on CD8 T cells, total loss of Qa-1 does not lead to development of spontaneous autoimmunity, since loss of CD94-NKG2 inhibitory signaling is largely compensated for by loss of TCR signaling. However, mutation of Qa-1, leading to its loss of CD94-NKG2, but not TCR binding, results in severe lupus-like autoimmunity underpinned by dysregulated Tfh-cell responses. The function of Qa-1–restricted CD8 appears to be largely dependent on perforin expression, as perforin-deficient CD8 T cells fail to mediate the suppression of Tfh cells.[18] While these cells lack expression of Foxp3, it is clear that Helios, a transcription factor also shared by thymic Treg, plays an important role in stabilization of both CD4 Treg and these suppressive CD8 cells; however, despite this area of overlap it appears that CD8 Treg have a closer phenotypic relationship with innate cells such as NKT than with CD4 Treg.[18]

Double-negative T cells are a class of αβ TCR-expressing cells that lack both CD4 and CD8, and make up around 1% of the TCR αβ cells in mice and humans.[19] Much remains unknown about the action of these cells, but it seems that they may be able to suppress CD4 and CD8 T cells by taking up antigen–MHC complexes from APC via trogocytosis, capturing plasma membrane fragments from the immunological synapse and then presenting them to other T cells in the absence of costimulatory signaling and possibly a signal-inducing apoptosis, such as Fas. They may also suppress APCs in a manner similar to Foxp3⁺ Treg by expression of CTLA-4.[19]

γδ T cells with a regulatory phenotype are a subset of the epithelial γδ T cells that can be found in mice. Mice deficient in γδ T cells do not appropriately regulate responses to various pathogens. This inappropriate regulation manifests as immunopathology in conjunction with the robust development of immunity. γδ T cell–deficient mice also show accelerated autoimmune responses in models of systemic lupus erythematosus (SLE) and spontaneously develop dermatitis when bred on certain genetic backgrounds. Commonly, these conditions are driven by αβ T cells, and γδ T cells will inhibit αβ T cells predominantly in the local environment. In humans, who lack an equivalent population of intraepithelial γδ cells, it is plausible

that this immune regulation is provided by other types of suppressive cells. However some forms of non-epithelial γδ cells have also been suggested to have an immunosuppressive role in antitumor immunity.[20]

NKT cells respond to CD1d, the nonclassic class I APC, which binds glycolipids rather than peptides. NKT cells can induce either pro-inflammatory (IFN-γ) or anti-inflammatory (IL-4, IL-10, IL-13) immune responses, but the prerequisites for this choice are ill defined. Nevertheless, under appropriate conditions, NKT cells clearly promote tolerance, which is illustrated in studies of transplantation and oral tolerance (see Table 13.1).[21]

Suppressive Non–T Cells

Recently, interest has also been focused on non–T-cell subsets with suppressive functions. One example of this is B cells with regulatory function (known as *Breg* or *B10 cells*). Breg cells produce the suppressive cytokine IL-10 in both mice and humans. Experiments in mice have demonstrated that these cells have suppressive capacity and can influence the development of autoimmune diseases, such as EAE, collagen-induced arthritis, and colitis. Breg cells appear to be induced in the periphery; to date, a clear transcriptional controller providing a clear transcriptional program has not been identified. This, together with the finding that various subsets of Breg cells differentiate from a wide variety of different B-cell populations (marginal zone, plasma blasts, plasma cells, $CD5^+CD1d^{high}$ B10 cells), has led to the suggestion that Breg cells may not be a lineage in themselves but, rather, a suppressive state that can be induced in all B cells when given the correct stimuli.[22] The suppressive function of Breg cells has largely been attributed to production of IL-10, TGF-β, and IL-35; this may either be via direct action on effector CD4 and CD8 T cells or indirectly by induction of other suppressive populations, such as $Foxp3^+$Tregs and Tr1 cells.[22]

In humans functional or numerical defects in Breg have been found in a variety of autoimmune disease settings such as SLE, rheumatoid arthritis (RA), and psoriasis; this suggests that therapeutic manipulation of Breg may be an attractive strategy for the treatment of autoimmunity.[23]

Myeloid-derived suppressor cells (MDSCs) are broadly divided into polymorphonuclear MDSCs, which are closely related to neutrophils, and monocytic MDSCs, which are related to monocytes, although in both cases, while they share common precursors, mature neutrophils and monocytes cannot convert to MDSCs. MDSCs are often associated with tumor progression and are believed to play an important role in the establishment of the immunosuppressive tumor environment. Accumulation of MDSCs in the tumor environment occurs in a two-step process, with the first being a tumor buildup of immature myeloid precursors and a block on their terminal differentiation to neutrophils or monocytes caused by tumor released growth factors. This is followed by pathological signaling from the tumor stroma, causing the conversion of immature precursors into MDSCs.[24]

MDSCs can be distinguished from other myeloid cells by the high levels of expression of NADPH oxidase (Nox2) and nitric oxide synthase 2 (*nos2*), leading to production of reactive oxygen species (ROS) and nitric oxide (NO) and the transcription factor c/EBPβ. MDSCs have decreased phagocytic capacity, which, in combination with their production of suppressive cytokines, such as IL-10 and TGF-β, leads them to have a suppressive effect on T-cell responses.[25]

DCs have also been demonstrated to have tolerogenic properties in certain circumstances. It is unclear if tolerogenic DC are of a stable lineage or, perhaps like Bregs, represent a particular state of differentiation. Antigen presentation by immature DC may, indeed, be tolerogenic because of lack of costimulation, but antigen presentation by the same cells may be immunogenic once mature and expressing greater levels of costimulatory molecules. Additionally, both pDC found in the tumor microenvironment and $CD103^+$ conventional DC in the lamina propria produce the immune suppressive molecule IDO, which has been demonstrated to aid in the induction of pTreg cells.[26]

CLINICAL RELEVANCE OF REGULATORY T CELLS

Abundant evidence strongly supports Treg as key controllers of self-tolerance, and Treg of various subsets play an active role in the control of almost all types of physiological and pathological immune responses, which also makes them suitable targets for immunotherapy (Tables 13.1 and 13.3).

THERAPEUTIC PRINCIPLES

Adjustment of the Immune Response by FOXP3+ Regulatory T Cell

- Reduction of $FOXP3^+$ Treg suppression or reducing Treg numbers
  - Enhancement of tumor immunity
  - Clearance of infections
  - Improvement of responses to vaccines
- Boosting of $FOXP3^+$ Treg function or increasing Treg numbers
  - Treatment of autoimmunity
  - Treatment of allergic responses
  - Induction of transplantation tolerance
  - Control of excessive immunopathology to foreign antigens (i.e., pathogens)
  - Maintenance of fetomaternal tolerance in pregnancy

TABLE 13.3 Potential Therapeutic Approaches for Treg-Based Therapy

Increase of Treg Numbers or Function	Reduction of Treg Numbers or Function
Ex vivo expansion of pure $FOXP3^+$ Treg with allo- or autoantigens plus growth factors, such as IL-2, and chemicals, such as rapamycin	Transient reduction of $FOXP3^+$ Treg and/or perturbation of suppression in vivo (anti-CD25 antibody, anti-CTLA-4 antibody, or anti-IL-2 antibody)
Ex vivo induction of Treg from conventional T cells by cytokines (IL-10, TGF-β), pharmacological agents, or modified DC	Render effector T cells resistant to suppression (GITR signaling)
In vivo promotion of Treg, rather than effector T cells, using monoclonal antibody treatment or pharmacological agents (anti-CD3 antibody, anti-CD40L antibody, etc.)	

CTLA-4, Cytotoxic T-lymphocyte antigen-4; *DC*, dendritic cell; *GITR*, glucocorticoid-induced tumor necrosis factor receptor protein; *IL-2*, interleukin-2; *IL-10*, interleukin-10; *TGF-β*, transforming growth factor-β; *Treg*, regulatory T cell.

Autoimmunity

As discussed above, Foxp3$^+$CD4$^+$ Treg are engaged in active suppression of autoimmune disease, since their depletion results in spontaneous development of autoimmune diseases in rodents. Furthermore, genetic anomaly of Foxp3 function can be a direct cause of autoimmune diseases in humans, as exemplified in IPEX syndrome.[2] A reduction of Foxp3$^+$ Treg in number or function has been reported in systemic autoimmune diseases, such as SLE, Sjögren syndrome, antineutrophil cytoplasmic antibody (ANCA)–associated vasculitis, Kawasaki disease, systemic sclerosis, psoriasis, autoimmune hepatitis, myasthenia gravis, and IBD.[2,27] Of note, type II autoimmune polyglandular syndrome, which resembles the systemic disease induced in nude mice reconstituted with splenocytes devoid of CD4$^+$CD25$^+$ T cells, is also characterized by Treg functional deficiency. However, studies of MS and type 1 diabetes did not detect any differences between patients and controls, and conflicting data have been reported in RA with regard to both function and numbers of CD25^{high} Treg.[27] The accuracy of the identification of Treg must be taken into account in these human studies, since CD25 and Foxp3 are also expressed to some extent by activated non-Treg T cells in humans.[6]

A general observation is that FOXP3$^+$ Treg increase in number at the site of inflammation. In the case of RA, synovial fluid from patients with ongoing RA was found to contain increased numbers of Foxp3$^+$CD25$^+$ Tregs compared with the levels in peripheral blood. CD25$^+$ Tregs from the synovial fluid of patients with RA are largely functional, although their numbers or suppressive function is apparently not enough to halt the inflammatory process. In contrast, CD25$^+$ Tregs obtained from the blood of patients with MS reportedly have decreased abilities to suppress proliferation of effector T cells. In summary, reduced levels of FOXP3$^+$CD25$^+$ Tregs in peripheral blood are not a general finding in autoimmune diseases and may not necessarily reflect the actual conditions at the site of inflammation. However, if Treg subsets defined by expression levels of CD45RA and Foxp3 are assessed, cytokine-secreting Foxp3^{low} non-Treg cells increase in active SLE.[2] Dynamic changes of Treg subsets in various autoimmune states is an area that still requires further investigation.[6]

Allergic Disease

While Th2 cells play an important role in allergic responses, it is also now clear the IL-4–producing Tfh cells are responsible for the majority of immunoglobulin E (IgE) production. FOXP3$^+$ Treg play an important role in suppressing the development of allergic reactions to innocuous environmental substances. This is best illustrated by IPEX syndrome, which is accompanied not only by organ-specific autoimmune diseases but also by severe eczematous dermatitis, high levels of serum IgE and sometimes eosinophilia.[2] Indeed, FOXP3$^+$ Treg from the blood of healthy donors without allergies suppress both proliferation and production of Th2 cytokines when challenged with specific allergens in vitro. If the same experiment is performed with Treg or Tfr from individuals with allergies, a marked difference is seen, as these Treg fail to downregulate Th2 and Tfh-related responses to allergens. IL-10 production by the FOXP3$^+$ Treg, Tr1 and Breg, has been implicated in the control of allergic responses.[28]

Since the suppressive ability for polyclonal stimuli is retained in patients with allergies, this deficiency is directly related to the allergen the individual is sensitized to, and thus probably does not reflect a general Treg deficiency. The inability to suppress Th2 and Tfh responses induced by birch or grass pollen is aggravated when the allergic reaction is ongoing and effector cells are fully activated, as during spring and summer compared with wintertime. Addition of IL-4 can attenuate the FOXP3$^+$ Treg-mediated suppression of Th2 clones in vitro in the same way as IL-2, which may provide an explanation for the insufficient control of ongoing allergic responses by Treg. Individuals with or without allergies harbor allergen-specific, IL-4–producing effector T cells, IL-10–producing Tr1 cells, and CD25$^+$ Treg, but in different proportions. The balance between Th2, Tfh, and certain Treg populations such as Tfr may, therefore, dictate whether clinical allergy will develop. Indeed, in the setting of curative specific allergy immunotherapy (SIT), allergen-specific IL-10–producing T cells can be induced. Furthermore, children who "grow out" of their allergy to cow's milk have higher numbers of CD4$^+$CD25$^+$ T cells specific for β-lactoglobulin compared with children with clinically active allergies. This suggests that certain allergies can be cured by the induction or expansion of antigen-specific Treg and that the balance between Treg and effector T cells, and Tfr and Tfh in particular, is of importance to prevent allergies.[29]

Transplantation

The ultimate goal of organ transplantation is to establish tolerance of allogeneic organ grafts as effectively and stably as self-tissues, but without the need for continuous general immunosuppression (Chapter 89).[30] Needless to say, Treg have sparked a lot of attention in this area of research. CD25$^+$ Treg were first shown to suppress GvHD in murine models of allogeneic bone marrow transplantation. Similarly, nude mice transplanted with allogeneic skin rejected the graft when reconstituted with CD4$^+$CD25$^-$ T cells alone but retained the graft when large enough numbers of CD25$^+$ Treg were cotransferred. In humans, attempts have been made to prevent GvHD in bone marrow transplantation and induce graft tolerance in organ transplantation with the use of purified FOXP3$^+$ Treg.[30] Another potential way to promote the induction of Treg in organ transplantation is to evaluate the effects of various immunosuppressants in terms of altering the balance between Treg and effector T cells. Different immunosuppressants target different pathways in cell metabolism and can therefore have different effects on cell populations that behave in dissimilar ways, such as effector T cells and Treg cells. Dosage and timing of administration, as well as specific drug combinations, seem to be a promising angle of transplantation immunotherapy with the purpose of inducing graft tolerance and preventing graft rejection. Treg have been shown to home to, and reside within, the graft, which is stably accepted once a dominance of Treg has become established. Treg-mediated transplantation tolerance is not a systemic phenomenon but rather is localized to the graft, and as such, it would not incur the dangers that accompany general immunosuppression.[30]

Tumor Immunity

It is now well known that many of the tumor-associated antigens recognized by a patient's T cells are normal self-constituents, indicating that antitumor immune responses are within the range of FOXP3$^+$ Treg control. Therefore the presence of Treg in the normal immune system may not only prevent autoimmunity but also hamper immune surveillance of cancer.[31] In fact, FOXP3$^+$ Treg, in particular highly

suppressive FOXP3^{high} effector Treg, are abundant in the tumor mass where they likely block any immune response targeting malignant cells. Studies on human malignancies have shown that the frequency of FOXP3$^+$ Treg is increased in tumors of, for example, metastatic melanoma or pancreatic and lung cancers. Moreover, high levels of FOXP3$^+$ Treg cells in the tumor are correlated with a poor prognosis and survival. Treg cells are not only involved in solid tumors but also in hematological malignancies.[31] For example, the architectural pattern of Treg cells in follicular lymphomas is associated with the prognosis of the disease. Whether the elevated levels result from migration of Treg cells into the tumor or from an expansion on the site is not clear, but evidence exists in support of both events. For example, ovarian tumor cells and infiltrating macrophages secrete the Treg-recruiting chemokine CCL22, which binds to CCR4 expressed by Treg cells, and, in addition, many tumors produce TGF-β, which contributes to the maintenance of FOXP3$^+$ Treg cells and may also induce FOXP3 expression in non-Treg cells within the tumor microenvironment. It is now evident that both effector T cells and Treg cells must be assessed in monitoring the efficacy of anticancer immunotherapy.[31]

The involvement of Treg in tumor immunity indicates that antitumor immune responses can be provoked or enhanced by depleting Treg in a host that is otherwise responding poorly. Experimental mouse models have demonstrated that simple depletion of CD25$^+$ Treg cells with anti-CD25 antibody results in tumor eradication, and similar effects can be achieved with in vivo administration of agonistic anti-GITR, anti-CTLA-4–blocking antibodies, or anti-CCR4 cell–depleting antibodies.[31] Depletion of CD25$^+$ T cells also enhances the effect of vaccination with tumor antigens.

Pharmacological agents are another possible way of altering the effector T cell / Treg ratio. For example, fludarabine was shown to selectively decrease the frequency of CD25$^+$ Treg cells in patients receiving chemotherapy. Conversely, previously used regimens, such as administration of exogenous IL-2, are now being reevaluated, since IL-2 may expand Treg cells. As expected from the role of Treg cells in self-tolerance, a caveat of Treg-based therapies of cancer is the possible development of autoimmunity, which may depend on the degree and period of in vivo systemic Treg depletion, as well as the genetic makeup of the host.[31]

In addition to Foxp3-expressing Treg cells, MDSC are believed to have an important role in the establishment and maintenance of the tumor immunosuppressive environment, largely through their production of immunosuppressive cytokines, such as IL-10 and TGF-β, and reactive nitrogen and oxygen species.[32]

Infectious Disease

Immune responses to infectious agents, such as bacteria and viruses, often result in tissue damage, which might be more severe if it were not for the involvement of Treg cells. On the downside, in many cases, Treg cells can contribute to the development of chronic infections. As previously discussed, Treg have the potential to directly respond to microbial products and are believed to be engaged in suppressing responses to infectious agents. A number of studies have shown that the outcome of an infection partly hinges on the proper balance between effector T cells and Treg cells.[33] Adoptive transfer of Treg cells prevents lethal pneumonia in T cell–deficient mice infected with *Pneumocystis jiroveci*, but at the expense of a deficient protective response and microbial clearance. Similarly, Treg suppress Th1 responses in mice infected with *Helicobacter pylori*, thereby limiting the mucosal inflammation but resulting in a higher bacterial load. Human studies have shown that Treg from carriers of *H. pylori* suppress responses to *H. pylori* antigens in vitro and have increased frequencies of CD25^{high} T cells in both the stomach and the duodenal mucosa compared with Treg cells from healthy controls. Taken together, modulation of the infectious response by Tregs can limit tissue damage but may enhance pathogen survival. This sort of compromise may not always be disadvantageous to the host. For example, in murine *Leishmania major* infection, Tregs prevent complete eradication of the parasites, which results in the persistence of low numbers of microbes that have been proven to be essential for the development of T-cell memory and prevention of reinfection. However, this delicate balance can be tipped in favor of the pathogen, which can be seen in the case of malaria and various viral infections such as human immunodeficiency virus (HIV). For example, HIV-specific CD4 and CD8 T-cell responses are substantially suppressed by Tregs in vitro in most individuals with HIV infection. Taken together, future treatments, as well as vaccine design, will need to take Tregs into account, and depending on the pathogen in question, it might be necessary to reduce or enhance the activity of Tregs to achieve a favorable outcome (see Table 13.3).[33]

TRANSLATIONAL RESEARCH

Manipulations of Foxp3$^+$ Treg cells in animal models of, for example, autoimmune disease, cancer, transplantation, and infection, have shown a great potential for modulation of immunological diseases by this subset. The findings in animals have already started to make their way into clinical practice and more is likely to come within the next 5 to 10 years.[2] In cases when immunity needs to be boosted, such as in cancer and during infection, the approach is to identify molecules that inhibit function and differentiation of Treg cells as well as molecules that locally deplete them. Particularly, cancer immunotherapy has proven to be successful (Chapter 80 and 81).[34] Such examples in current therapy are CTLA-4–specific blocking antibody (ipilimumab; MDX-010/Yervoy; Bristol-Myers Squibb, N.Y.); although CTLA-4 expression is not exclusive to Treg cells, it is much more highly expressed and critical to their suppressive function. Furthermore, recent studies have suggested that in addition to blocking the function of CTLA-4, anti-CTLA-4 antibodies may also deplete Treg cells in the tumor environment, but not in the periphery, because of antibody-dependent, cell-mediated cytotoxicity. This depletion specificity to the tumor environment may be attributed to a combination of high levels of CTLA-4 expression and the presence of large numbers of phagocytic cells on the local area.[31] Furthermore, other checkpoint blockade strategies, such as those targeting the cell intrinsic immunosuppressive molecule PD-1 or its ligand PD-L1, have also proven to be effective. Combination therapies are also of great interest. It has been found that anti-CTLA-4 treatment can lead to increased expression of PD-L1 by the tumor itself, which may dampen the effect. Additionally, PD-1 therapy can expand Tregs, enhancing their function and reversing any benefits of PD-1 blocking of effector T-cells.[2] Thus adding both anti-CTLA-4 and anti-PD1 or PD-L1 may have greater benefits, as demonstrated by a recent phase I trial of anti-CTLA-4 (ipilimumab) plus anti-PD-1 (nivolumab) in patients with melanoma.[35] Certain combinations may prove particularly effective, since they target different pathways. Again, in this case,

CTLA-4 may be acting primarily as a cell-extrinsic suppressive molecule on Treg cells, and PD-1 may be acting in *cis* to suppress PD-1–expressing CD4 or CD8 effector cells, and as a result, the combination of the two may provide synergistic effects. As might be predicted, these types of immunomodulatory treatments can lead to severe autoimmune adverse events; however, further studies on both CTLA-4 and PD-1 and other potential checkpoints, such as B- and T-lymphocyte attenuator (BTLA), inducible T-cell costimulator (ICOS), and a range of other inhibitor or stimulatory molecules,[34] either as monotherapies or in combination, may help boost efficacy while reducing adverse events.[34]

Conversely, in cases when tolerance should be reinforced (autoimmunity, transplantation, allergy, fetomaternal tolerance), molecules that either mimic or enhance Treg function and survival are under investigation. In the former case, CTLA-4-Ig (abatacept; Orencia, Bristol-Myers Squibb, N.Y.), a solubilized version of CTLA-4 that, in part, mimics the effect of Treg cells on APC by blocking their expression of B7.1 and B7.2, is currently being used for the treatment of, for example, RA. In addition, cellular therapy, by infusion of ex vivo or manipulated expanded Treg cells, is in small-scale use in patients with GvHD after hematopoietic stem cell transplantation,[2] and early trials of Treg transfer in recent-onset type 1 diabetes are underway. Further to this, transfer of Tr1 cells in patients with Crohn disease is also under investigation.[14] Cell-based therapies are, indeed, a promising strategy for managing various immunological diseases and immune responses; however, several challenges still remain, particularly those relating to the functional stability of ex vivo–expanded or –induced Treg cells.[2]

ON THE HORIZON

- It is clear that regulatory T cells (Treg) in different sites have phenotypic differences, and that a greater understanding of Treg heterogeneity is critical to clarification of their role in human autoimmunity. New approaches such as single-cell RNA sequencing, single-cell epigenomic analysis, and mass cytometry may prove critical to this.
- Specific targeting of tumor-resident Treg without depletion of peripheral Treg is critical to enhancing antitumor therapy while avoiding adverse events.
- Expansion and control of Treg subsets, such as T-follicular regulatory (Tfr) or muscle-resident Treg cells, may be critical to fine control of specific immunological disorders.

REFERENCES

1. Yi J, Kawabe T, Sprent J. New insights on T-cell self-tolerance. *Curr Opin Immunol*. 2019;63:14–20.
2. Sakaguchi S, Mikami N, Wing JB, et al. Regulatory T cells and human disease. *Annu Rev Immunol*. 2020;38:541–566.
3. Abbas AK, Benoist C, Bluestone JA, et al. Regulatory T cells: recommendations to simplify the nomenclature. *Nat Immunol*. 2013;14(4):307–308.
4. Curotto de Lafaille MA, Lafaille JJ. Natural and adaptive foxp3+ regulatory T cells: more of the same or a division of labor? *Immunity*. 2009;30(5):626–635.
5. Sakaguchi S. Regulatory T cells: history and perspective. *Methods Mol Biol*. 2011;707:3–17.
6. Wing JB, Tanaka A, Sakaguchi S. Human FOXP3(+) regulatory T cell heterogeneity and function in autoimmunity and cancer. *Immunity*. 2019;50(2):302–316.
7. Rudensky AY. Regulatory T cells and Foxp3. *Immunol Rev*. 2011;241(1):260–268.
8. Kitagawa Y, Sakaguchi S. Molecular control of regulatory T cell development and function. *Curr Opin Immunol*. 2017;49:64–70.
9. Whibley N, Tucci A, Powrie F. Regulatory T cell adaptation in the intestine and skin. *Nat Immunol*. 2019;20(4):386–396.
10. Wing JB, Tay C, Sakaguchi S. Control of regulatory T cells by co-signal molecules. *Adv Exp Med Biol*. 2019;1189:179–210.
11. Leonard WJ, Lin JX, O'Shea JJ. The gammac family of cytokines: basic biology to therapeutic ramifications. *Immunity*. 2019;50(4):832–850.
12. Panduro M, Benoist C, Mathis D. Tissue Tregs. *Annu Rev Immunol*. 2016;34:609–633.
13. Yamaguchi T, Wing JB, Sakaguchi S. Two modes of immune suppression by Foxp3(+) regulatory T cells under inflammatory or non-inflammatory conditions. *Semin Immunol*. 2011;23(6):424–430.
14. Roncarolo MG, Gregori S, Bacchetta R, et al. The biology of T regulatory type 1 cells and their therapeutic application in immune-mediated diseases. *Immunity*. 2018;49(6):1004–1019.
15. Chien CH, Chiang BL. Regulatory T cells induced by B cells: a novel subpopulation of regulatory T cells. *J Biomed Sci*. 2017;24(1):86.
16. Canete PF, Sweet RA, Gonzalez-Figueroa P, et al. Regulatory roles of IL-10-producing human follicular T cells. *J Exp Med*. 2019;216(8):1843–1856.
17. Yamaguchi T, Kishi A, Osaki M, et al. Construction of self-recognizing regulatory T cells from conventional T cells by controlling CTLA-4 and IL-2 expression. *Proc Natl Acad Sci U S A*. 2013;110(23):E2116–E2125.
18. Nakagawa H, Wang L, Cantor H, et al. New insights into the biology of CD8 regulatory T cells. *Adv Immunol*. 2018;140:1–20.
19. Ligocki AJ, Niederkorn JY. Advances on Non-CD4+ Foxp3$^+$ T regulatory cells: CD8$^+$, Tr1, and double negative T regulatory cells in organ transplantation. *Transplantation*. 2015;99(8):1553–1559.
20. Zhao Y, Niu C, Cui J. Gamma-delta (gammadelta) T cells: friend or foe in cancer development? *J Transl Med*. 2018;16(1):3.
21. Rossjohn J, Pellicci DG, Patel O, et al. Recognition of CD1d-restricted antigens by natural killer T cells. *Nat Rev Immunol*. 2012;12(12):845–857.
22. Rosser EC, Mauri C. Regulatory B cells: origin, phenotype, and function. *Immunity*. 2015;42(4):607–612.
23. Mauri C, Menon M. Human regulatory B cells in health and disease: therapeutic potential. *J Clin Invest*. 2017;127(3) 772–729.
24. Condamine T, Mastio J, Gabrilovich DI. Transcriptional regulation of myeloid-derived suppressor cells. *J Leukoc Biol*. 2015;98(6):913–922.
25. Gabrilovich DI. Myeloid-derived suppressor cells. *Cancer Immunol Res*. 2017;5(1):3–8.
26. Kim SJ, Diamond B. Modulation of tolerogenic dendritic cells and autoimmunity. *Semin Cell Dev Biol*. 2015;41:49–58.
27. Grant CR, Liberal R, Mieli-Vergani G, et al. Regulatory T-cells in autoimmune diseases: challenges, controversies and—yet—unanswered questions. *Autoimmun Rev*. 2015;14(2):105–116.
28. Akdis CA, Akdis M. Mechanisms of immune tolerance to allergens: role of IL-10 and Tregs. *J Clin Invest*. 2014;124(11):4678–4680.
29. Yao Y, Wang ZC, Wang N, et al. Allergen immunotherapy improves defective follicular regulatory T cells in patients with allergic rhinitis. *J Allergy Clin Immunol*. 2019;144(1):118–128.
30. Atif M, Conti F, Gorochov G, et al. Regulatory T cells in solid organ transplantation. *Clin Transl Immunology*. 2020;9(2):e01099.
31. Tanaka A, Sakaguchi S. Targeting Treg cells in cancer immunotherapy. *Eur J Immunol*. 2019;49(8):1140–1146.
32. Kumar V, Patel S, Tcyganov E, et al. The nature of myeloid-derived suppressor cells in the tumor microenvironment. *Trends Immunol*. 2016;37(3):208–220.
33. Berod L, Puttur F, Huehn J, et al. Tregs in infection and vaccinology: heroes or traitors? *Microbial Biotechnology*. 2012;5(2):260–269.
34. Wei SC, Duffy CR, Allison JP. Fundamental mechanisms of immune checkpoint blockade therapy. *Cancer Discov*. 2018;8(9):1069–1086.
35. Rizvi H, Sanchez-Vega F, La K, et al. Molecular determinants of response to anti-programmed cell death (PD)-1 and anti-programmed death-ligand 1 (PD-L1) blockade in patients with non-small-cell lung cancer profiled with targeted next-generation sequencing. *J Clin Oncol*. 2018;36(7):633–641.

14

Cytokines and Cytokine Receptors

John J. O'Shea, Massimo Gadina, Giuseppe Sciumè, and Françoise Meylan

Cytokines play pivotal roles in controlling the development and functions of a variety of immune and non-immune cells. These include immune regulation, disease pathogenesis, and—increasingly—modulation and treatment of immune-mediated diseases.

The term "cytokine" encompasses factors that are structurally or functionally unrelated. Included among cytokines are a number of different factors produced by lymphoid and non-lymphoid cells that mediate intercellular communication. The term "lymphokine" was originally used to denote products of lymphocytes,[1] whereas "interleukin" was introduced to emphasize the importance of these factors in communication between leukocytes.[2] Although the designation interleukin has remained in use, the term does not accurately convey that many interleukins are made by cells other than leukocytes.

KEY CONCEPTS

Cytokine Characteristics

- Cytokines have pleotropic effects—they bind more than one receptor.
- Cytokines can be redundant—their receptors often share subunits.
- Cytokines can have specific and unique functions—their receptors typically have ligand-specific subunits as well.

Cohen et al.[3] coined the word "cytokine" to emphasize the point that these secreted factors need not be made by one specific cell source. This was an important insight, because many immunologically relevant cytokines are made by non-lymphoid and even non-immune cells, though the argument is easily made that all cells participate in immune responses. Cytokines are thus defined operationally as polypeptides secreted by leukocytes and other cells that act principally on hematopoietic cells, the effects of which include modulation of immune and inflammatory responses. However, there are clear exceptions to even this broad definition (discussed below).

Some definitions distinguish cytokines from hormones and growth factors, which act on non-hematopoietic cells. Cytokines are typically characterized as factors that act locally, whereas hormones are secreted by specialized cells and act at a distance on a restricted set of target cells. Although many cytokines do act locally in an autocrine or paracrine fashion, some enter the bloodstream and can act in a typical endocrine fashion. Consequently, the boundary between cytokines and hormones is rather indistinct. In fact, classic hormones such as growth hormone (GH), prolactin (PRL), and erythropoietin (EPO), and a more recently identified hormone, leptin, are all clearly cytokines, as evidenced by the structure of their receptors and their modes of signaling. Perhaps it is just simplest to accept that cell-cell communication and host defense went hand-in-hand during evolution, and so functional and structural similarities exist among families of molecules that act on the immune, hematopoietic, endocrine, and nervous systems.

CYTOKINE CLASSIFICATION

A major challenge in discussing cytokines is how best to classify them. Due to how they were discovered, complicated nomenclatures and classifications have arisen that can be a barrier to understanding cytokines. Indeed, many were first identified by researchers in different disciplines who derived names based on their original observed functions that may not reflect the full spectrum of a given cytokine's actual biological functions.

One can legitimately group cytokines in different ways, but we have chosen to classify cytokines based on the type of receptor that they bind. Our scheme emphasizes the evolutionary relatedness of cytokines, growth factors, and hormones and highlights similarities in signal transduction. The classification used is adapted from Vilcek[4] and includes the following receptors: the so-called type I (hematopoietin family) and type II (interferon family) cytokine receptors, tumor necrosis factor (TNF) family receptors, interleukin (IL)-1 receptor and the related Toll-like receptors (TLRs), IL-17 receptors, receptor tyrosine kinases, and the transforming growth factor-β (TGF-β) family receptor serine kinases (Fig. 14.1). A seventh group, known as chemokines, forms a separate family and bind seven transmembrane domain receptors (see Chapter 15). This chapter addresses only a selected set of cytokines with important immunological functions.

TYPE I AND II CYTOKINE RECEPTORS (HEMATOPOIETIN FAMILY AND INTERFERON RECEPTORS)

Ligand and Receptor Structure

Cytokines (Table 14.1) that bind the class of receptors termed the type I or hematopoietic cytokine receptor superfamily include hormone-like factors and colony-stimulating factors (CSF) such as GH, PRL, leptin, EPO, thrombopoietin (TPO), granulocyte (G)-CSF, and granulocyte–macrophage (GM)-CSF; and interleukins (IL)-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-9, IL-11, IL-12, IL-13, IL-15, IL-21, IL-23, IL-27, IL-31, IL-35, and IL-39. Also included in this family are ciliary neurotrophic factor (CNTF), leukemia inhibitory factor (LIF), oncostatin M (OSM), and cardiotropin 1 (CT-1). Closely related are the interferons (IFNs): IFN-αs (13 in humans and 14 in mice), -β, -ε, -κ, -ω (humans), and -ζ (mice) and IL-10-related cytokines: IL-10, IL-19, IL-20, IL-22, IL-24,

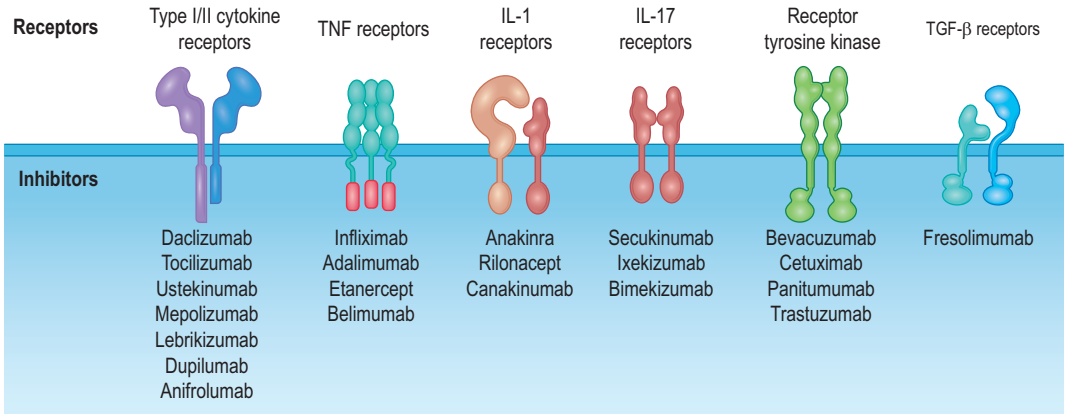


FIG. 14.1 Schematic Representation of Prototypical Receptors of the Six Major Cytokine Receptor Superfamilies and Biological Agents that Block Their Functions.

TABLE 14.1 Selected Cytokines Classified by Receptor Families

Receptor Family	Cytokine	Signaling	Predominant Source	Major Target	Actions	Knockout Phenotype/ Human Mutations
Type I (hemato-poietin)	GH	JAK2, STAT5B	Two GH genes, pituitary, placental	Diverse tissues	Growth, adipocyte differentiation	Dwarfism
	Prl	JAK2, STAT5A	Two Prl genes, pituitary, uterus	Mammary epithelium	Growth, differentiation	Infertility, lactation defects
	Epo	JAK2, STAT5	Kidney, liver	Erythroid precursors	Erythroid differentiation	Embryonic lethal, severe anemia, GOF mutations associated with erythrocytosis
	Tpo	JAK2, STAT5	Liver, kidney	Committed stem cells and megakaryocytes	Platelet growth and differentiation	Severe thrombocytopenia
	Leptin	JAK2, STAT3	Adipocytes	Hypothalamus, thyroid	Satiety, controls metabolic rate	Hyperphagia, obesity
	G-CSF	JAK2, STAT3	Many tissues, macrophages, endothelium, fibroblasts	Committed progenitors	Differentiation, activates mature granulocytes	Neutropenia
	IL-6	JAK1, JAK2, TYK2, STAT1, STAT3	Macrophages, fibroblasts, endothelium, epithelium, T cells, other cells	Liver, B cells, T cells, thymocytes myeloid cells, osteoclasts	Acute-phase reactants proliferation, differentiation co-stimulation	Reduced Ig (esp. IgA), T lymphopenia, impaired acute-phase response, impaired Th17 cell development, reduced estrogen-dependent bone loss *IL6RST* associated with hyperimmunoglobulin E syndrome
	IL-11	JAK1, JAK2, STAT3	Stromal cells, synoviocytes, osteoblasts	Hematopoietic stem cells, hepatocytes, macrophages, neurons	Proliferation	Female infertility *IL11RA* mutations associated with craniosynostosis
	IL-27	JAK1, JAK2, TYK2, STAT1, STAT3, STAT4. STAT5	Activated DCs, macrophages, epithelial cells	T and NK cells, other cells	Enhancement of Th1 responses and IL-10 production; inhibition of Th1, Th2, and Th17 responses	Fatal inflammatory disease during infection
	IL-31	JAK1, STAT3, STAT5	Th2 cells, CD8 T cells	Monocytes, epithelial cells keratinocytes, eosinophils basophils	Induces chemokines, PMN recruitment	
	CNTF[a]	JAK1, JAK2, TYK2, STAT3	Schwann cells	Neuronal	Survival	Progressive atrophy and loss of motor neurons

TABLE 14.1 **Selected Cytokines Classified by Receptor Families—Cont'd**

Receptor Family	Cytokine	Signaling	Predominant Source	Major Target	Actions	Knockout Phenotype/ Human Mutations
	LIF[a]	JAK1, JAK2, TYK2, STAT3	Uterus, macrophages, fibroblasts, endothelium, epithelium, T cells	Embryonic stem cells, neurons, hematopoietic cells	Survival	Decreased hematopoietic progenitors, defective blastocyst implantation, maintenance of stem cell pluripotency, *LIFR* mutations in humans associated with Stuve-Wiedemann syndrome and congenital abnormalities of kidneys and urinary tract
	OSM	JAK1, JAK2, TYK2, STAT3	Macrophages, fibroblasts, endothelium, epithelium	T cells, myeloid cells, liver, embryonic stem cells	Differentiation, acute-phase induction	Thymic hypoplasia, *OSMR* mutations underlie familial primary localized cutaneous amyloidosis
	CT-1	JAK1, JAK2, TYK2, STAT3	T cells, other cells, myocardium	Myocardium	Growth	Motor neuron death
	CRLF1/ CLCF1	JAK1, JAK2, TYK2, STAT3	Lymphocytes, myeloid cells	Neurons	Promotes neuronal cell survival, B-cell growth	Mutations associated with cold-induced sweating syndrome, autonomic dysfunction, kyphoscoliosis, and craniofacial abnormalities
	GM-CSF	JAK1, JAK2, STAT3	T cells, macrophages, endothelium, fibroblasts	Immature and committed myelomonocytic progenitors, macrophages, granulocytes, and DCs	Growth, differentiation, survival, activation	Pulmonary alveolar proteinosis
	IL-3	JAK1, JAK2, STAT5	T cells, macrophages, mast cells, NKT cells, eosinophils	Immature hematopoietic progenitors of multiple lineages	Growth, differentiation, survival	No defects in basal hematopoiesis
	IL-5	JAK1, JAK2, STAT5	Th2 T cells, activated eosinophils, ILCs and NKT cells	Eosinophils, B cells, basophils, mast cells	Proliferation, activation	Decreased eosinophilia, defective CD5, B1-cell development
	IL-2	JAK1, JAK3, STAT5	T cells, NKT cells, ILCs, DCs	T, B, and NK cells, ILCs, macrophages	Proliferation, cytotoxicity IFN-γ secretion, antibody production	Lymphoproliferation[a]
	IL-4[b]	JAK1, JAK3, STAT6	Th2 cells, basophils, mast cells, NKT cells, γ/δ T cells	T cells, B cells, macrophages	Proliferation, Th2 differentiation, IgG1 and IgE production, inhibition of cell-mediated immunity	Defective Th2 differentiation and IgE production, decreased allergic responses
	IL-7	JAK1, JAK3, STAT5	Bone marrow, thymic stromal cells, spleen DCs, keratinocytes, monocytes, macrophages	Thymocytes, T cells, B cells, ILCs	Growth, differentiation, survival	SCID[a]
	IL-9	JAK1, JAK3, STAT5	Th2 and Th9 T cells, ILCs, mast cells, eosinophils	T cells, B cells, mast cell precursors, goblet cells	Proliferation, Th1 inhibition	Impaired goblet cell mucus production
	IL-15[b]	JAK1, JAK3, STAT5	Many cells	T cells (especially memory T cells), NK and NKT cells	Proliferation, survival, activation	Absence of NK and memory cells
	IL-21	JAK1, JAK3, STAT3	T cells, Th17 cells, Tfh cells	T, B, and NK cells, DCs, macrophages, keratinocytes	Isotype switching, plasma cell differentiation, enhancement of CD8 and NK cell responses, promotes Th17 cell differentiation	Decreased numbers of Th17 cells, LOF mutations associated with primary immunodeficiency in humans

Continued

TABLE 14.1 Selected Cytokines Classified by Receptor Families—Cont'd

Receptor Family	Cytokine	Signaling	Predominant Source	Major Target	Actions	Knockout Phenotype/ Human Mutations
	IL-13	JAK1, TYK2, STAT6	Activated T cells, NKT, ILCs, mast cells, basophils	B cells, mast cells, macrophages, epithelial cells, smooth muscle cells	Co-stimulator of proliferation, Increases IgE, CD23 and MHC Class II expression, inhibits cytokine secretion and cell-mediated immunity	Defective Th2 responses and IgE production, decreased allergic responses
	IL-12	JAK2, TYK2, STAT1, STAT3, STAT4, STAT5	Macrophages, DCs, B cells	T cells, NK cells	Th1 differentiation, proliferation, cytotoxicity	Defective Th1 differentiation, susceptibility to bacterial infections* *IL12R* mutations associated with primary immunodeficiency
	IL-23	JAK2, TYK2, STAT1, STAT3, STAT4, STAT5	Macrophages, DCs	T cells, ILCs, macrophages	IL-17 production	Reduced arthritis, inflammation
	IL-35	STAT1 STAT2	Treg cells	T cells	Treg proliferation, suppresses proliferation and function of Th17	Reduced Treg activity
	TSLP	JAK1, JAK2, STAT1 STAT3, STAT5	Epithelial cells, keratinocytes	DCs (human), B cells (mouse)	Th2 differentiation (human)	
Type II	Type I IFNs (IFN-α/β)	JAK1, TYK2, STAT1, STAT2, IRF9	Ubiquitous, but especially plasmacytoid DCs	All cells, immunoregulatory effects on immune cells	Antiviral, anti-proliferative, increases MHC class I activation	Susceptibility to viral infections[a]
	IFN-γ	JAK1, JAK2, STAT1	Th1 T cells, NK cells	Macrophages, endothelium, T cells, NK cells	Activation, increases MHC class II expression, increases antigen presentation	Susceptibility to bacterial infections[a]
	IL-10	JAK1, TYK2, STAT3	Most leukocytes, including macrophages, DCs, T cells, NK cells, and B cells	Myeloid cells, DCs, T cells	Represses immune responses, decreases MHC class II expression, decreases antigen presentation, stimulates mast cells and eosinophils	Exaggerated inflammatory response and autoimmune disease *IL10* and *IL10R* mutations associated with IBD
	IL-19, -20, -24, -26	JAK1, TYK2, STAT1, STAT3	T cells, myeloid cells, NKT cells upregulated in psoriasis and RA	T cells, keratinocytes, epithelial cells	Induces production of inflammatory cytokines, Th2 responses, activation of epithelial cells	
	IL-22	JAK1, TYK2, STAT3	T cells, ILC3	Barrier epithelial cells	Induces anti-bacterial peptides, promotes wound repair, tissue regeneration	Increased gut inflammation
IL-1/TLR	IL-1α/β	IRAK, MyD88, TRAF6, NF-κB	Many cells, esp. macrophages	CNS, endothelial cells, liver, thymocytes, macrophages, T cells	Fever, anorexia, activation of acute-phase reactants co-stimulation, activation, cytokine secretion, differentiation of Th17 cells	Reduced inflammation, cooperates with TNF in host defense
	IL-18	IRAK, MyD88, TRAF6, NF-κB	Many cells, esp. macrophages, keratinocytes, DCs, osteoblasts	T cells, NK cells, macrophages, epithelial cells	Induction of IFN-γ, activation of NK cells, angiogenesis, tumor progression	Increased susceptibility to infection, reduced arthritis
	IL-33	IRAK, MyD88, p38	macrophages, DCs, fibroblasts, adipocytes, smooth muscle cells, endothelial cells, osteoblasts, epithelial cells	T cells, nuocytes, mast cells, basophils, granulocytes, ILC2	Enhances Th2 responses, angiogenesis	
	IL-36	IRAK, MyD88, MAPK	Skin	DCs, T cells	Induce secretion of pro-inflammatory cytokines,	Pustular psoriasis, elevated levels of pro-inflammatory cytokines in skin
	IL-37	IRAK, MyD88, AKT, STAT3	Many cell types and tissues	DCs, macrophages	Suppression of inflammation	

TABLE 14.1 Selected Cytokines Classified by Receptor Families—Cont'd

Receptor Family	Cytokine	Signaling	Predominant Source	Major Target	Actions	Knockout Phenotype/ Human Mutations
	IL-38	IRAK, MyD88, MAPK	B cells	Macrophages	Suppression of inflammation	
IL-17	IL-17A	ACT1, TRAF6, NF-κB, MAPK, C/EBP	Th17 cells, CD8 T cells, γ/δ T cells	Endothelium, many cell types and tissues	Inflammation	Susceptibility to extracellular bacteria
	IL-17B, C, D		Many cell types and tissues	Monocytes, epithelial cells	Inflammation, chondrogenesis	
	IL-17E (IL-25)	ACT1, TRAF6, TRAF2	Mast cells, Th2 cells	Th2 cells	Enhances Th2 responses	Increased susceptibility to helminths
	IL-17 F	ACT1, TRAF6, NFkB, MAPK, C/EBP	Th17 cells, CD8 T cells, γ/δ T cells	Endothelium, many cell types and tissues	Inflammation	
TGF-β receptor serine kinase family	TGF-β 1,2, 3	SMADs	T cells, macrophages, other cell types	T cells, macrophages, other cell types	Inhibits growth and activation, promotes Th17 development	
Receptor tyrosine kinases	Stem cell factor	Ras/Raf/MAPK, stromal cells	Bone marrow	Pluripotent stem cells	Activation, growth	Defective hematopoietic stem cell proliferation, melanocyte production and development
	CSF-1 (M-CSF)	Ras/Raf/MAPK	Macrophages, endothelium, fibroblasts, other cell types	Committed myelomonocytic progenitors	Differentiation, proliferation, survival	Monocytopenia, osteopetrosis, female infertility
	Flt-3 ligand	Ras/Raf/MAPK	Many cell types and tissues	Myeloid cells, especially DCs	Proliferation, differentiation	Reduced repopulating hematopoietic stem cells; reduced B-cell precursors
	IL-32	NF-κB, p38 MAPK	T cells, NK cells, monocytes, epithelia	Monocytes	Induces TNF, IL-1, IL-6, IL-8	
	IL-16		T and B cells, mast cells, eosinophils	CD4 T cells		
	IL-34	ERK, PI3K	Many cell types and tissues	Monocytes	Monocytes, macrophages and osteoclasts differentiation, and proliferation via CSF-1 receptor.	

In cases where STAT5a or STAT5b are designated, the cytokines appear to use both interchangeably.
[a]LIFR is shared by these cytokines.
[b]Note that two forms of the IL-4 and, perhaps, IL-15 receptors exist.

IL-26, and the IFN-related cytokines IL-28A (IFN-λ2), IL-28B (IFN-λ3), IL-29 (IFN-λ1), and IFN-λ4, all of which bind type II receptors. The ligands and receptors in this superfamily are structurally similar and utilize related molecules for signal transduction.[5] A central feature of type I cytokines is the presence of four anti-parallel α helices with two long and one short loop connections arranged in an up–up–down–down configuration, and thus their designation as the α-helical bundle cytokine family.

Structurally, the receptors in the type I family have conserved cysteine residues, a conserved Trp–Ser–X–Trp–Ser motif (where X indicates any amino acid), and fibronectin-like repeats in their extracellular domains. These receptors have a single transmembrane domain and divergent cytoplasmic domains. Within the cytoplasmic portion of these receptors, two segments of homology can be discerned—termed the Box 1 and Box 2 motifs. The membrane proximal domain binds Janus kinases (JAKs; see below). Some of the cytokine receptors are homodimers, such as the receptors for EPO, TPO, PRL, and possibly leptin; whereas other receptors for type I cytokines are heterodimers, containing two distinct receptor subunits. Based on this characteristic, the type I family of receptors can be divided into subfamilies. Each member of the subfamily uses a shared receptor subunit in conjunction with a ligand-specific subunit. For example, the receptors for IL-2, IL-4, IL-7, IL-9, IL-15, and IL-21 all use a common cytokine γ chain—γc (see Table 14.1)—whereas a common β chain—βc—is shared by IL-3, IL-5, and GM-CSF. Similarly, gp130 is a shared receptor subunit for IL-6 family cytokines: IL-6, IL-11, IL-27, IL-35, IL-39, ciliary neurotrophic factor (CNTF), LIF, OSM, cardiotrophin-1 (CT-1), and cardiotrophin-like cytokine factor 1 (CLCF1). IL-12 and IL-23 also share a receptor subunit, as do members of the IL-10 family.

Other levels of shared receptor usage also exist. For example, the receptors for LIF, CNTF, OSM, and CT-1 all share the LIF receptor subunit. IL-31 and OSM share one receptor chain. IL-2 and IL-15 utilize the same β and γc chains. Conversely, IL-4 can bind two different receptor complexes: one composed of IL-4Rα and γc and a second composed of IL-4Rα and IL-13Rα. IL-13 only utilizes the IL-13 receptor complex for signaling.

The utilization of common receptor subunits explains the phenomenon of shared biological activities (cytokine redundancy) between cytokines that belong to the same subfamily. Within a subfamily, actions distinct for each cytokine can be attributed, at least in part, to the ligand-specific subunits. The pleiotropic effects of a single cytokine can be accounted for by the existence of more than one receptor for a cytokine and expression of receptors on a wide range of cells.

CLINICAL RELEVANCE

Types I and II Cytokine Receptors

- Mutations of genes encoding IL-7R, γc, and JAK3 cause SCID.
- TYK2 and STAT3 mutations cause hyper-IgE syndrome.
- STAT1 mutations cause autosomal dominant chronic mucocutaneous candidiasis with increased susceptibility to mycobacterial and viral infections.
- Mutations of genes encoding IL-12, IL-12R, and IFNγR are associated with susceptibility to intracellular infections.
- Polymorphisms of IL-2R and IL-7R are associated with multiple sclerosis.
- Polymorphisms of IL-23R are associated with IBD.
- Polymorphisms of STAT4 are associated with RA and SLE.
- EPO, G-CSF, and TPO are used to treat cytopenias.
- Anti-cytokine and/or cytokine receptor mAbs are used to prevent transplant rejection and treat several autoimmune and inflammatory diseases.

Family Members and Their Actions

Homodimeric Receptors

Cytokines that use homodimeric receptors include GH, PRL, leptin, EPO, and TPO. EPO is required for erythrocyte growth and development and is widely used to treat anemia. TPO is required for megakaryocyte development and may have a use in the treatment of thrombocytopenia. G-CSF regulates the production of neutrophils (Chapter 39) by acting on committed progenitor cells, supporting the survival of mature neutrophils, and enhancing functional capacity. G-CSF is used clinically to treat patients with granulocytopenia. G-CSF-deficient mice have marked neutropenia, and mutations of the G-CSFR result in severe congenital neutropenia in humans.

Cytokine Receptors Utilizing gp130

Gp130 is ubiquitously expressed and is the shared receptor component for IL-6, IL-11, IL-27, IL-35 IL-39, LIF, OSM, CNTF, CT-1, and the compound cytokine, CRLF1/CLCF1. Targeted disruption of the gene encoding gp130, *Interleukin 6rst (Il6rst)*,[6] is lethal in early embryogenesis, causing defects in myocardial, hematological, and placental development.

Interleukin-6. The IL-6 receptor (IL-6R) consists of an IL-6 binding protein (α chain) (CD126) and membrane-bound gp130. IL-6 associates with IL-6Rα and gp130 to form a hexamer with 2:2:2 stoichiometry. While gp130 is ubiquitously expressed, IL-6Rα is expressed by hepatocytes and immune cells. However, signaling can occur in cells that express IL-6Rα and gp130 as well as cells that express only gp130. In the latter circumstance, IL-6 and the soluble form IL-6Rα can bind gp130, a mechanism termed trans-signaling.[7]

Originally identified as a B cell (Chapter 7) growth factor, IL-6 has a wide array of biological actions on both lymphoid and non-lymphoid cells with the consequences of signaling by the membrane-bound and soluble receptors being distinct. IL-6-deficient mice are susceptible to *Candida* (Chapter 28) and *Listeria* (Chapter 26) infections. IL-6 induces production of immunoglobulins (Chapter 8), including IgE; and $Il6^{-/-}$ mice have normal numbers of B cells with reduced immunoglobulin response to immunization and reduced IgA production (Chapter 33). IL-6 also promotes T-cell growth and differentiation (Chapter 9) and interleukin-6-deficient ($Il6^{-/-}$) mice have reduced numbers of thymocytes and peripheral T cells. IL-6 is important for Th17 differentiation (Chapter 11) and the cytotoxic T-cell response (Chapter 12) to viruses (Chapter 25). IL-6 functions synergistically with IL-3 in hematopoiesis, and $Il6^{-/-}$ mice have reduced numbers of progenitor cells.

IL-6 is a major inducer of fever, inflammation (Chapter 37), and the synthesis of acute-phase proteins (e.g., fibrinogen, serum amyloid A, haptoglobin, C-reactive protein) in the liver. The elevation of the erythrocyte sedimentation rate (ESR) in inflammatory disease largely reflects the accelerated synthesis of these proteins, and IL-6-deficient mice are defective in this response. IL-6 reduces synthesis of albumin and transferrin in the liver and initiates hepatocyte regeneration. IL-6 induces adrenocorticotrophic hormone and anterior pituitary hormones, such as PRL, GH, and luteinizing hormone. IL-6 also plays a role in osteoporosis by affecting osteoclast function. $Il6^{-/-}$ mice are protected from bone loss following estrogen depletion.

Levels of IL-6 in serum are low in the absence of inflammation, but are rapidly increased in response to infection or trauma. Patients with rheumatoid arthritis (RA) (Chapter 53), cardiac myxoma, Castleman disease, and other autoimmune or inflammatory diseases (Chapter 51) have high serum levels of IL-6. This cytokine may also contribute to malignancies such as multiple myeloma (Chapter 79).

IL-6 is produced by many immune and non-immune cells, including fibroblasts, keratinocytes, astrocytes, and endothelial cells. Stimulation of monocytes with IL-1, TNF, or lipopolysaccharide (LPS), in turn, stimulates the expression of IL-6, whereas IL-4 and IL-13 inhibit its production. The *IL6* gene contains binding sites for nuclear factor-κB (NF-κB), nuclear factor for IL-6 (NF-IL-6, or CCAAT element-binding protein), activator protein-1 (AP-1), cAMP response element-binding protein (CREB), and the glucocorticoid receptor.

Monoclonal antibodies (mAbs) targeting IL-6 receptor (tocilizumab, sarilumab) are approved for treatment of RA (Chapter 53), systemic juvenile idiopathic arthritis (Chapter 54), and Takayasu arteritis (Chapter 60). Tocilizumab is also approved for cytokine release syndrome associated with chimeric antigen receptor T-cell therapy (Chapter 81) and is being tested in the treatment of Covid-19-associated cytokine storm (Chapter 31). Satrilizumab is an IL-6R mAb approved neuromyelitis optica. Olamkicept is a soluble gp130 Fc fusion protein being tested presently.

IL6RST mutations are associated with hyper-IgE syndrome (HIES) (Chapter 33).

Interleukin-11. IL-11 and its receptor are widely expressed. IL-11 stimulates stem cells, megakaryocytes, myeloid precursors, and erythroid precursors and promotes B-cell differentiation. It also acts on non-hematopoietic cells, including bone and liver. IL-11 is induced by pro-inflammatory cytokines (IL-1, TNF) and by TGF-β. Oprelvekin is a recombinant IL-11 approved for the treatment of severe thrombocytopenia. *IL11RA* mutations are associated with craniosynostosis.

Interleukin-27. IL-27 is composed of two subunits, designated EBI3 and p28 (also termed IL-30), and signals through gp130 and IL-27Rα (also known as WSX-1/T-cell cytokine receptor). The receptor is expressed on naïve CD4 T cells. IL-27 promotes Th1 differentiation, but also has essential anti-inflammatory properties, inhibiting Th17 differentiation and enhancing IL-10 production.[8]

LIF binds to gp130 in association with the LIF receptor (LIFR), as do OSM, CNTF, and CT-1. Deletion of the *Lifr* gene is embryonically lethal, creating defects in placental architecture and developmental abnormalities in neural tissue and bone. Targeted disruptions of LIF lead to failure of blastocyst implantation. LIF is also critical for the maintenance of stem cell pluripotency in culture. *LIFR* mutations in humans are associated with Stuve-Wiedemann syndrome and congenital abnormalities of kidneys and urinary tract. Deletion of *Osm* in mice results in thymic hypoplasia and decreased bone marrow progenitor activity associated with autoimmunity. *OSMR* mutations underlie familial primary localized cutaneous amyloidosis (Chapter 37) and anti-OSM antibody is being tested in systemic sclerosis (Chapter 56). Mutations of *CRLF1* are associated with cold-induced sweating syndrome, autonomic dysfunction, kyphoscoliosis, and craniofacial abnormalities.

Cytokine Receptors Utilizing the βc Chain

IL-3, IL-5, and GM-CSF bind to a ligand-specific α subunit associated with the common βc receptor subunit (common β subunit, *CSF2RB*, CD131).[9] Mice, but not humans, have a second β chain—βIL3. This species-specific redundancy may explain why gene targeting of βc in the mouse did not result in loss of IL-3 responses, although βc-null mice did have reduced GM-CSF and IL-5 responses. The GM-CSF/IL3/IL5 family is not essential for steady-state production of myeloid cells.

Interleukin-3. IL-3 synergizes with other cytokines to stimulate the growth of immature progenitor cells of all lineages and is termed a multilineage CSF. It promotes survival of macrophages (Chapters 3 and 6), mast cells (Chapter 44), and megakaryocytes. IL-3 is produced mainly by lymphoid cells, mast cells, and eosinophils (Chapter 45). IL-3-deficient mice have no obvious defect in hematopoiesis, suggesting that the major role of IL-3 in vivo may be in the response to stress.

Interleukin-5. IL-5 is unusual in that it is a disulfide-linked homodimer, with each component containing three α-helical bundles. Originally identified as a growth factor for B cells and eosinophils, it is important in allergic disease. *Il5*$^{-/-}$ mice fail to develop eosinophilia in response to parasitic (Chapter 30) or aeroallergen challenge (Chapter 43) and exhibit minimal signs of inflammation and damage to the lungs. IL-5 deficiency does not affect the worm burden of infected mice, indicating that eosinophilia may not play an essential role in the host defense against helminths per se. Both IL-5 and IL-5R knockout mice have decreased numbers of CD5^{+} B cells (B-1 cells) (Chapter 7) and concomitant low serum IgM and IgG3 levels. IL-5 is produced by type 2 innate lymphoid cells (ILC2) (Chapter 3), activated helper T cells of the Th2 phenotype, mast cells, and eosinophils in an autocrine manner and is regulated by neuropeptides.

Mepolizumab and reslizumab are anti-IL-5 mAbs that have been approved for the treatment of severe eosinophilic asthma disease and eosinophilic granulomatosis with polyangiitis. Benralizumab is an anti-IL-5Rα (CD125) antibody approved for treatment of eosinophilic asthma and was granted orphan drug designation for treatment of eosinophilic esophagitis.

Granulocyte–Macrophage-CSF. GM-CSF (encoded by *CSF2*) acts on hematopoietic precursors to support myelomonocytic differentiation and causes rapid expansion of myeloid cells during inflammation. It activates mature neutrophils and macrophages, increasing their microbicidal activity and inducing the production of pro-inflammatory cytokines. Along with IL-4 and IL-13, GM-CSF is important for the in vitro production of dendritic cells (DCs), but also expands immunosuppressive myeloid cells. GM-CSF induces proliferation and activation of eosinophils and upregulates adhesion molecules on fibroblasts and endothelial cells.

Deletion of *Csf2* in mice does not affect steady-state hematopoiesis. Instead, these mice develop lymphoid hyperplasia and alveolar proteinosis due to failure to clear surfactant from the lungs. βc(*Csf2rb*)-deficient mice and humans with *CSF2RA* and *CSF2RB* mutations also develop alveolar proteinosis, characterized by the accumulation of surfactant in the lungs.

GM-CSF is produced by lymphoid cells, including natural killer (NK) (Chapter 12), invariant natural killer (iNKT), T helper 17 (Th17), and ILC3 cells. GM-CSF can be induced by pro-inflammatory cytokines and LPS and is an important driver of immune pathology in murine models of autoimmunity. GM-CSF is not ordinarily detectable in blood except under pathological conditions, such as asthma. Sargramostim is a recombinant form of GM-CSF approved for the treatment of myelosuppression, especially in the context of infection (e.g., fungal). Mavrilimumab and gimsilumab are anti-GM-CSF antibodies being studied in arthritis and Covid-19. Sipuleucel-T is a GM-CSF fusion protein approved for use in a prostate cancer vaccine. Talimogene laherparepvec is an engineered GM-CSF-expressing virus approved for the treatment of melanoma.

Cytokine Receptors Utilizing the γc Chain

IL-2, IL-4, IL-7, IL-9, IL-15, and IL-21 all bind to receptors that share a common γc receptor subunit.[10] The γc subunit and the ligand-specific subunits are expressed predominantly on lymphocytes, although they can also be found on other hematopoietic cells. Mutation of the γc gene (*IL2RG*) is responsible for X-linked severe combined immunodeficiency (SCID), characterized by a lack of T cells and NK cells, and poorly functioning B cells (T^{-}B^{+} SCID) (Chapter 34). The lack of γc abrogates signaling by all cytokines that utilize this subunit (IL-2, IL-4, IL-7, IL-9, IL-15, and IL-21). The lack of IL-7 and IL-15 signaling is largely responsible for the lack of T and NK cells, respectively.

Interleukin-2. The IL-2 receptor consists of three subunits, α, β, and γc. The latter two are members of the type I cytokine receptor family. NK cells constitutively express these latter two subunits and respond to high doses of IL-2; whereas in T cells, the IL-2Rα subunit is induced upon activation, creating a high-affinity receptor for IL-2. IL-2Rα is also highly expressed on regulatory

T cells (Treg) (Chapter 13), ILC2, and ILC3 cells (Chapter 3) and inducible in activated monocytes and B cells. IL-2Rα, however, is not a member of the type I cytokine receptor family. Rather, it resembles members of the complement family and IL-15R (see below).

IL-2—one of the first cytokines to be intensively studied—is produced by activated T, ILC2, ILC3, and DC (Chapters 3 and 6). It was first identified as an autocrine T-cell growth factor, required for in vitro T-cell proliferation. However, the generation of *Il2*$^{-/-}$ mice and their systemic autoimmune disease unexpectedly revealed the essential role of IL-2 in preserving immune tolerance. IL-2-deficient mice develop massive enlargement of peripheral lymphoid organs, hemolytic anemia, inflammatory bowel disease (IBD) (Chapter 75), and infiltrative granulopoiesis. In addition, the mice have high levels of IgG1 and IgE. The mice succumb to this widespread autoimmune and lymphoproliferative disease, pointing to the nonredundant roles of IL-2 in limiting immune responses. A similar phenotype is seen in humans with IL-2 and IL-2 receptor mutations. Thus, we now appreciate the criticality of IL-2 in promoting the expression of Foxp3 and the development of Treg cells in the thymus and periphery.[11]

Additionally, IL-2 is an important factor in determining the magnitude of T-cell and NK-cell responses in vivo, although other factors also contribute. IL-2 augments the cytolytic activity of T and NK cells, enhances IFN-γ secretion, and promotes Th2 and Th9 differentiation. IL-2 also influences fate determination between follicular helper T (Tfh) cells and non-Tfh effector T cells and inhibits Th17 differentiation.[12] IL-2 is also important in programming CD8 memory T cells, which undergo secondary expansion in viral infections. IL-2 is a growth factor for B cells and induces class switching (Chapter 7) and also activates macrophages.

The *IL2* gene has been extensively characterized and contains binding sites for nuclear factor of activated T cells (NFAT), AP-1, and NF-κB. IL-2 production is also regulated by stabilization of its mRNA.

Recombinant IL-2 (aldesleukin) is approved for the treatment of renal cancer and melanoma; however, its clinical utility is limited by its toxicity, including hepatic dysfunction and capillary or vascular leak syndrome. Given its role in promoting Treg cell homeostasis and inducing Foxp3, low-dose IL-2 is being studied for the treatment of autoimmune disease, with some success in systemic lupus erythematosus (SLE) (Chapter 52). Engineered versions of IL-2 (superkines) and antibodies that preferentially expand Treg cells versus effector T cells have been generated and have been studied in preclinical models.[13] Basiliximab and daclizumab—anti-IL-2Rα mAbs—were approved to prevent rejection of allotransplants. Polymorphisms of *IL2* and *IL2RA* are also associated with autoimmune disease.

Interleukin-4. The two types of IL-4Rs are the type I receptor comprising IL-4Rα in conjunction with γc, which is expressed on hematopoietic cells, and the type II receptor comprising IL-4Rα and IL-13Rα, which is broadly expressed. Type I receptors bind IL-4, and type II receptors bind both IL-4 and IL-13. The loss of IL-4Rα blocks the actions of both IL-4 and IL-13, explaining why gene targeting of IL-4Rα is more severe than IL-4-deficiency. The existence of two receptors helps to explain the diverse actions of IL-4 on both hematopoietic and non-hematopoietic cells. A third receptor subunit, IL13RA2, binds IL-13 with high affinity but not IL-4; it may act as a negative regulator of IL-4 and IL-13 signaling.

IL-4 was discovered as a factor that promotes B-cell differentiation and drives immunoglobulin class switching. More broadly, IL-4, which shares many actions with IL-13, promotes allergic responses and inhibits cell-mediated immune responses. IL-4 promotes differentiation of naïve CD4 T cells into T helper 2 (Th2) cells that produce IL-4, IL-13 and IL-5, and Th9 cells[14] (Chapter 11). In conjunction with CD40 engagement, IL-4 promotes B-cell proliferation and class switching—particularly to IgG1 and IgE in mice and to IgG4 and IgE in humans—and upregulates expression of IgM, MHC class II (Chapter 5), and CD23 (Chapter 77).

In conjunction with GM-CSF, IL-4 is a growth factor for mast cells and basophils, as well as a potent inducer of DC differentiation. IL-4 inhibits macrophage activation and the production of pro-inflammatory cytokines. It antagonizes the effects of IFN-γ, blocks cytokine-induced proliferation of synoviocytes, downregulates the expression of adhesion molecules, and antagonizes the induction of some acute-phase reactants in hepatocytes by IL-6.

IL-4-deficient mice have normal B lymphopoiesis but marked reductions in IgG1 and IgE production in response to parasites. These mice have residual Th2 responses because IL-13, which also binds IL-4Rα, can partially compensate for the defect.

IL-4 is made by the Th2 cells, NK1.1^{+} CD4 T cells, and basophils and mast cells (Chapter 44); ILC2 cells generally do not produce IL-4. A number of transcription factors appear to be important in regulating IL-4 production, including NFAT, NF-IL6, C/EBP, c-MAF, and GATA-3. The *IL4* gene has multiple STAT6 binding sites, consistent with the fact that IL-4 regulates its own expression. Epigenetic control and chromatin remodeling are also important aspects of IL-4 regulation.[15]

Polymorphisms of *IL4* and *IL4RA* are associated with allergy and asthma. The IL-4/IL-13 blocking antibody, dupilumab, is approved for the treatment of atopic dermatitis, asthma, and chronic rhinosinusitis. IL-4 is used to generate DC for tumor vaccines (Chapter 87).

Interleukin-7. The IL-7 receptor consists of the IL-7Rα chain (CD127) in association with γc. It is expressed on both immature and mature thymocytes. Humans with loss-of-function (LOF) mutations of IL-7Rα have T^{-}B^{+} SCID but display normal NK-cell development, unlike individuals with γc mutations (X-SCID) (Chapter 34). Gain-of-function (GOF) mutations of the IL-7Rα chain result in constitutive JAK1 signaling and cell transformation and give rise to T-cell acute lymphoblastic leukemia (Chapter 77).

IL-7 plays an important role in both developing thymocytes and mature T cells, and IL-7Rα expression is tightly regulated during thymocyte development (Chapter 9). IL-7Rα is expressed on double-negative thymocytes, downregulated in double-positive cells, and re-expressed in single-positive thymocytes and mature peripheral T cells. This regulation may relate to IL-7's anti-apoptotic effects and induction of Bcl-2 family members. IL-7 promotes the growth of thymocytes, as well as the expression and rearrangement of TCR genes and the expression of RAG1 and RAG2 (Chapter 4). IL-7Rα is expressed on cutaneous T-cell lymphomas, which also produce this cytokine; thus, the autocrine response to IL-7 may contribute to the growth of these tumors.

IL-7- and IL-7R-deficient mice exhibit impairement in T- and B-cell development. Postnatal B-cell development in *IL7*$^{-/-}$ mice is blocked at the transition to pre-B cells and is arrested even earlier in *IL7ra*$^{-/-}$ mice. Why these B-cell abnormalities do not occur in humans with IL-7Rα mutations is not clear.

IL-7 is produced by a wide variety of cells in the spleen, kidney, stroma, and epithelium. This is consistent with its role in the maintenance of function in both immature and mature lymphocytes.

Clinically, IL-7 may be useful to restore immune function in some congenital immunodeficiencies, after bone marrow transplantation (Chapter 92), or in HIV infection (Chapter 41). Polymorphisms of the *IL7R* gene are associated with multiple sclerosis (Chapter 66) and type 1 diabetes (Chapter 71).

Interleukin-9. IL-9R is expressed predominantly on ILC2, B cells, and non-lymphoid cells, including pulmonary goblet cells. IL-9 has some of the same properties as IL-4. It drives allergic inflammation as well as mucous production in the lungs. IL-9 may also contribute to immunopathology in IBD.

IL-9 synergizes with stem cell factor to promote the growth and differentiation of mast cells and regulate their function. IL-9 potentiates IgE production induced by IL-4 in B cells. IL-9 can inhibit Th1 cytokine production. Although first identified as a T-cell growth factor, a major role of IL-9 in T-cell development has not been established. IL-9 has been reported to have anti-tumor activity, but some lymphoid tumors also produce IL-9, where it may serve as an autocrine growth factor.

IL-9 is produced by activated Th2 cells, ILC2 cells, mast cells, and eosinophils. A subset of Th cells designated Th9 cells preferentially produces IL-9. Their differentiation is promoted by IL-2, IL-4, IL-25, and IL-33. Polymorphisms of *IL9* and *IL9R* are associated with allergy and asthma.

Interleukin-15. The IL-15 receptor consists of the IL-2Rβ and γc subunits in association with a unique ligand-specific subunit, IL-15Rα, which is homologous to IL-2Rα. These receptor proteins contain protein-binding motifs termed "sushi domains." In both human and mouse, the genes encoding receptors are linked.

Given their shared receptor usage, there are many similarities in the actions of IL-2 and IL-15. Like IL-2, IL-15 induces proliferation and cytokine production in T and NK cells. However, IL-15 is critical for NK cell development. In T cells, IL-15 is less efficient than IL-2 in inducing effector memory T-cell differentiation or sensitivity to apoptosis. IL-15Rα is more widely expressed than IL-2Rα, IL-2Rβ, and γc. IL-15Rα is expressed by lymphoid cells, DC, fibroblasts, epithelial, liver, intestine, and other cells and is thought to present IL-15 in *trans* to cells expressing IL-15β and γ chains. IL-15- and IL-15Rα-knockout mice are defective in NK-cell production and in the generation of memory T cells, explaining the absence of NK development in patients with γc mutations.

IL-15 mRNA is expressed in hematopoietic and non-hematopoietic cells but is not typically produced by T cells (HTLV-I-transformed T cells being an exception). Following the pattern seen in IL-7 and IL-9, there are multiple upstream AUGs in the 5′ untranslated portion of the IL-15 message that influence translational regulation. IL-15 protein is also controlled at the level of protein secretion, but this is not completely understood. High levels of IL-15 protein have been reported in patients with RA, sarcoidosis, and ulcerative colitis.

Interleukin-21. IL-21R is broadly expressed on B, T, dendritic, myeloid, and other cells. The genes encoding *IL21* and *IL2* are adjacent to each other in the genome and—like IL-2—IL-21 can activate CD8 T and NK cells, particularly when in conjunction with IL-15 or IL-7. However, IL-21 also opposes a number of actions of IL-2: it promotes Th17 differentiation and, along with IL-6, drives differentiation of Tfh cells. Tfh cells are found preferentially in B-cell follicles where, under control of the transcription factor BCL6, they regulate B-cell development, activation, and class switching. While IL-21 promotes IgG1 production, it represses IgE production. IL-21 also drives terminal B-cell differentiation to plasma cells.

IL-21 is produced by T cells, especially Tfh, Th17, and NKT cells. IL-21 appears to have some anti-cancer properties, and it has been tested in the treatment of melanoma.

Polymorphisms of *IL21* and *IL21R* are associated with multiple autoimmune diseases. LOF *IL21R* mutations are associated with primary immunodeficiency affecting T, B, and NK cells.

Other Heterodimeric Receptors

Interleukin-12. IL-12 is a heterodimer composed of two disulfide-linked polypeptides derived from two distinct genes, p35 (encoded by *IL12A*) and p40 (*IL12B*). While IL-12p35 shares homology with IL-6 and other cytokines, p40 resembles the IL-6 receptor. Thus, IL-12 can be viewed as being produced as a preformed ligand–receptor complex.[8] DC and macrophages are the major producers of IL-12 in response to various pathogens, occupancy of Toll-like receptors, and CD40. IL-12R consists of two chains denoted IL-12Rβ1 and β2, found predominantly on adaptive and innate-like T cells, NK cells, and other ILCs. IL-12Rβ1 and β2 are highly inducible upon T-cell activation, whereas both receptors are constitutively expressed on NK cells and ILC1. IL-12 plays a pivotal role in promoting cell-mediated immune responses. It induces IFN-γ, proliferation and cytolytic activity in innate and adaptive lymphocytes through activation of the transcription factor STAT4.[16,17] In uncommitted CD4 T cells, IL-12 promotes Th1 differentiation.

Humans carrying *IL-12/IL12R* mutations have blunted immune responses and are highly susceptible to infections by intracellular pathogens, typically mycobacteria (Chapter 26).[18] Because of its profound effects on cell-mediated immunity, IL-12 has been used in the treatment of infectious diseases and malignancies. Unfortunately, its utility is limited by toxicity. In this context, clinical trials aimed to selectively deliver IL-12 to tumors are currently under evaluation and IL-12 may have use in vaccines as an adjuvant. IL-12/IL-23 mAb are approved for treatment of psoriasis (Chapter 64) and IBD (Chapter 75).

Interleukin-23. IL-23 is another heterodimeric type I cytokine composed of two disulfide-linked polypeptide chains, p19 (*IL23A*) and IL-12p40 (*IL12B*), which bind to IL-12Rβ1 chain paired to IL-23R. The ability of p19 to complex IL-12p40 to IL-23 helped explain the phenotypic differences observed in mice deficient for IL-12p35 and IL-12p40. These observations ultimately led to the understanding of the role of IL-23 in driving IL-17 expression, providing the missing link between IL-23 and immune-mediated diseases.[19] The IL-23R complex is also expressed by ILCs and innate-like T cells, thereby inducing production of IL-17 and IL-22 via the activation of STAT3.

IL-23 is produced primarily by DC in response to TLR agonists. It regulates epithelial cell barrier functions, host defense against extracellular bacteria, and it contributes to the pathogenesis of autoimmune and autoinflammatory disorders. *IL23R* polymorphisms are associated with IBD, ankylosing spondylitis (Chapter 58), and multiple sclerosis.

Ustekinumab inhibits both IL-12 and IL-23 and is approved for psoriasis, psoriatic arthritis, Crohn disease, and ulcerative colitis; whereas tildrakizumab, risankizumab, and guselkumab

are approved for the treatment of psoriasis and inhibit IL-23 but not IL-12.

Interleukins-35 and -39. IL-35 is a dimer consisting of IL-12 p35 and EBI3. It is preferentially produced by Treg cells. Tregs are also the main cellular target of IL-35, where it induces proliferation and production of IL-10. A synthetic form of IL-35 can reduce the incidence of arthritis in mouse models.

IL-39, a dimer composed of IL-23p19 and EBI3, has been reported to be produced by murine B cells and acts on neutrophils. IL-39's role in humans has not been established.

Interleukin-13. IL-13 has many of the same effects as IL-4 and shares a receptor subunit with IL-4. IL-13-deficient mice have reduced levels of IL-4, IL-5, IL-10, IgE, and eosinophils. In mice deficient for both IL-4 and IL-13, Th2 responses are abolished and parasite clearance is severely impaired. These double-knockout mice default to Th1 responses with concomitant production of INF-γ, IgG2a, and IgG2b. It appears that IL-4 and IL-13 cooperate in promoting Th2 responses, and thus have both overlapping and additive roles.

IL-33 and the neuropeptide neuromedin U promote IL-13 production in ILC2s, whereas another neuropeptide, calcitonin gene-related peptide, limits IL-13 production.[20,21]

IL-13 mAbs include lebrikizumab, tralokinumab, and anrukinzumab and are currently undergoing testing for the treatment of atopic eczema, asthma, and IBD.

Interleukin-31. IL-31 signals through the heterodimeric receptor IL-31RA and oncostatin M receptor (OSMR). IL-31R is expressed on macrophages, DC, eosinophils, basophils, keratinocytes, and peripheral nerves. IL-31 is produced by activated Th2 cells, granulocytes, and mast cells, and its expression is especially prominent in skin. Overexpression of IL-31 results in atopic dermatitis, but surprisingly, IL-31RA-deficient mice showed an increased Th2 response.

Targeting IL-31 with nemolizumab appears to be efficacious in atopic dermatitis and prurigo nodularis.[22] Lokivetamab is approved for use in treating canine atopic dermatitis.

Thymic Stromal:Lymphopoietin. Thymic stromal lymphopoietin (TSLP) is an IL-7-like cytokine expressed by epithelial cells and keratinocytes. Its receptor comprises TSLPR (encoded by *CRLF2*) and IL-7Rα, which is expressed on leukocytes, especially on myeloid and B cells. TSLP exerts its effect through promotion of basophil hematopoiesis. TSLP-treated human DC promote Th2 differentiation. In mice, TSLP contributes to prenatal B-cell development. Elevated TSLP levels have been found in humans and animal models of airway inflammatory disease and atopic dermatitis. Tezepelumab is TSLP mAb being studied in severe asthma.[23]

Interferons

Type I Interferons

The type I IFNs include: IFN-αs, IFN-β, IFN-ε, IFN-κ, IFN-ω (humans), and IFN-ζ (mice).[24] IFN-β and IFN-ω are encoded by single genes, whereas IFN-αs include at least 13 separate genes in humans and 14 in mice, each encoding structurally distinct forms. These intronless genes are all clustered on the short arm of chromosome 9 and appear to have diverged from a common ancestor more than 100 million years ago. All type I IFNs signal via a heterodimeric receptor composed of two subunits—IFNAR1 and IFNAR2. The actions of type I IFNs are similar. These subunits have limited similarity to type I cytokine receptors, although they lack the WSXWS motif.

A major effect of type I IFNs is their antiviral action. Discovered in 1957, they act on all cells to inhibit viral replication and cellular proliferation. It is unclear why there are so many type I genes. Given that their relative potencies differ, it is possible that these genes evolved in response to specific viral pathogens. Alternatively, IFN gene duplication may affect the magnitude of antiviral responses. Type IFNs can also inhibit protein translation. Type I interferons upregulate MHC class I and can block the ability of interferon γ to upregulate MHC class II expression. IFN-α/β increase the cytolytic activity of NK cells. Predictably, *IfnarI* knockout mice are extremely susceptible to infections, even though lymphoid development is normal.

Interferons are produced ubiquitously, although plasmacytoid DC produce exceptionally high levels. Extracellular and intracellular foreign DNA produced by viruses and products generated by bacteria are detected by endosomal pattern recognition receptors including Toll-like receptors, RIG-I like cytosolic sensors (RIG-I/MDA5/MAVS), cyclic GMP-AMP synthase (cGAS), and stimulator of interferon gene (STING), which act to induce transcription of *IFN* genes (Chapter 3). *IFN* genes are bound by multiple transcription factors, including NF-κB, interferon regulatory factors 3 (IRF-3) and IRF7, and STAT1.

Beyond their direct antiviral actions, type I IFNs have broad actions on immune cells, including the regulation of adaptive immune responses. IFNs induce expression of class I and II MHC, co-stimulatory and adhesion molecules, and chemokines. Depending upon timing, type I IFNs can promote or inhibit T-cell proliferation and effector differentiation. B lockade of type I IFN signaling during the chronic phase of viral infection can decrease expression of negative regulatory molecules, enhance IFN-γ responses, and improve viral immunity.

Overproduction of IFNs underlies a collection of heterogenous diseases. Inherited interferonopathies include LOF mutations of genes that metabolize DNA and RNA (TREX1, SAMHD1, RNASEH2, ADAR1) or negatively regulate IFNAR signaling (ISG15, USP18), and GOF mutations in cytosolic RNA or DNA sensors (MDA5 and STING).[25] Human chromosome 21 contains four IFN receptor genes (*IFNAR1*, *IFNAR2*, *IFNGR2*, and *IL10RB*), and interferonopathy is one aspect of Down syndrome. Many autoimmune diseases exhibit high expression of IFN-inducible genes—the "interferon signature"—with lupus being a prominent example.

Anifrolumab—an anti-IFNAR1 mAb—is approved for treatment of lupus.[26] Recombinant IFN-β is approved for the treatment of multiple sclerosis. Recombinant type I IFNαs are approved for the treatment of certain infections (e.g., viral hepatitis). Owing to their anti-proliferative action, IFNαs are also used in the treatment of certain malignancies, particularly hairy cell leukemia.

Interferon-γ

IFN-γ (also termed type II IFN) is a major activator of macrophages and neutrophils, enhancing their ability to kill microorganisms by augmenting their cytolytic machinery. Like IFN-α/β, IFN-γ contributes to antiviral defenses. IFN-γ upregulates MHC class II expression and increases production of reactive oxygen intermediates, including hydrogen peroxide, nitric oxide, and indoleamine dioxygenase. IFN-γ acts on CD4 T cells to promote Th1 differentiation while inhibiting the generation of Th2 cells. It promotes the maturation of CD8 T cells to cytotoxic cells,

augments NK cytolytic activity, regulates B-cell class switching, and activates endothelial cells.

The IFN-γ receptor is a heterodimer composed of IFN-γRα and IFN-γRβ subunits. When one IFN-γ homodimer binds, a complex of two α and two β receptors is created. Mice with a disrupted IFN-γR have normal lymphoid development but are highly susceptible to viral and bacterial infections, especially intracellular microbes. Knockout mice have diminished macrophage MHC class II expression, decreased NK function, and reduced serum IgG2a concentrations. Humans with mutations of *IFNGR* subunits are also susceptible to mycobacterial and *Salmonella* infections.

IFN-γ is produced by Th1 and NK cells. Transcription factors—including STAT4, T-BET, and EOMES—play important roles in *IFNG* gene regulation. IFN-γ has been used to treat patients with immunodeficiencies (e.g., chronic granulomatous disease) and in certain patients with disseminated mycobacterial infections. Emapalumab is an anti-IFN-γ mAb approved for hemophagocytic lymphohistiocytosis (Chapter 36).

Interleukin-10 and related cytokines. The interleukin-10 (IL-10) family of cytokines comprises IL-10, IL-19, IL-20, IL-22, IL-24, IL-26, and type III IFNs and binds to two shared receptors, IL-10RB and IL-20RB.[27] IL-10 binds to IL-10RA and IL-10RB, whereas IL-22 binds to IL-22R and IL-10RB. Type III IFNs bind to IL-28R, IFNLR1, and IL-10RB; IL-26 binds to IL-20RA and IL-10RB; IL-20 and IL-24 bind IL-20RA; and IL-19, IL-20, and IL-24 bind IL-20RA and IL-20RB. The shared usage of receptors helps explain overlapping functions.

Interleukin-10

IL-10R is expressed on macrophages, mast cells, and most other hematopoietic cells. It is inducible in non-hematopoietic cells by stimuli such as LPS. Unlike other cytokines in this family, IL-10 is a disulfide-linked dimer and is made by T, B, NK and DC, macrophages, keratinocytes, bronchial epithelial cells, and other cells. LPS and TNF are inducers of IL-10.

IL-10 serves as an anti-inflammatory and immunosuppressive cytokine, limiting damage during host response to infection. It inhibits macrophage antigen presentation and decreases expression of MHC class II, adhesion molecules, and the costimulatory molecules CD80 (B7.1) and CD86 (B7.2). IL-10 has a direct inhibitory effect on memory Th17 and Th2 cells. It promotes activity of $Foxp3^+$ regulatory T cells and activates mast cells and B cells.

IL-10 is detected in the blood of patients with septic shock and other inflammatory and immune disorders. There is a correlation between levels of IL-10 and autoantibody production in SLE.

IL-10-deficient mice develop autoimmune disease manifested by severe IBD and exaggerated inflammatory responses. Mutations and polymorphism of the *IL10* and *IL10R* genes are associated with IBD in humans. Viral homologues of IL-10 may blunt the immune response to these pathogens.

Interleukin-19, Interleukin-20, Interleukin-22, Interleukin-24, and Interleukin-26. IL-22 acts on tissue epithelial cells, inducing Reg family microbial peptides and promoting tissue regeneration and wound healing. IL-22 is predominately secreted by T cells and ILC3s. IL-22's actions are antagonized by its natural inhibitor, IL-22BP, which is produced by T cells, DC, and eosinophils.

IL-20 subfamily cytokine receptors are preferentially expressed on epithelial cells, including keratinocytes and lung and intestinal epithelial cells. IL-19 and IL-20 are mainly produced by myeloid cells. Myeloid and Th2 cells are cellular sources for IL-24. IL-20 cytokines are upregulated in many diseases, including psoriasis and RA; however, their biological actions remain elusive.

IFN-λs. Type III interferons include IFN-λ1 (IFNL1, IL-29), -λ2 (IFNL2, IL-28A), -λ3 (IFNL3 IL-28B), and IFN-λ4 (IFNL4) in humans; and IFN-λ2 and -λ3 in mice.[24] Receptors for type III IFNs are expressed on barrier epithelial cells and some immune cells. Type III IFNs are less potent and produced at slower kinetics than type I IFNs. Like type I IFNs, type III IFNs induce antiviral responses, but the interferon-stimulated genes induced by type III IFNs are a subset of those induced by type I IFNs. Antiviral responses in the gut are dominated by type III rather than type I IFN signaling. Polymorphisms of *IFNL4* are associated with hepatitis C susceptibility. A frameshift mutation in the promoter of *IFNL4* that results in the loss of IFN-λ4 production is associated with improved clearance of HCV.

Signaling

Neither type I nor type II receptors exhibit intrinsic enzymatic activity. However, the conserved membrane proximal segment of each of these receptors serves as the site at which these receptors bind Janus kinases (JAKs) (Fig. 14.2, Table 14.1).[28] These JAKs play a pivotal role in the downstream signaling via this family of cytokine receptors.

Janus Kinases

There are four mammalian Janus kinases—JAK1, JAK2, JAK3, and TYK2. JAKs have a C-terminal, catalytically active kinase domain that is preceded by a segment termed the pseudokinase domain, which regulates kinase activity. The JAK amino terminus mediates their association with cytokine receptors.

Ligand binding to type I and II receptors induces the aggregation of receptor subunits, which brings JAKs into close proximity, allowing them to phosphorylate and activate each other. After activation, the JAKs phosphorylate tyrosine residues on receptor subunits that recruit proteins with SRC homology-2 (SH2) or phosphotyrosine-binding (PTB) domains. In turn, these proteins are phosphorylated by JAKs resulting in the activation of a number of biochemical pathways. Importantly, phosphorylation of cytokine receptors generates docking sites for a class of SH2-containing transcription factors termed the STATs (see below) (see Fig. 14.2).

The pivotal function of the JAKs is vividly illustrated by mice and humans that are deficient in these kinases. Consistent with the selective association of JAK3 with the common gamma chain, γc; mutations of *JAK3* cause autosomal recessive T^-B^+ SCID (Chapter 34). $Jak3^{-/-}$ mice also exhibit SCID, and mutation of either γc or JAK3 leads to the same functional defects. Multiple JAK inhibitors have been approved as immunomodulatory drugs for various diseases (Chapter 86).[29] The importance of the JAKs is also substantiated by mutations in many leukemias and lymphomas.

$Jak1^{-/-}$ mice die perinatally from neurological defects, but also have a SCID phenotype similar to $Jak3^{-/-}$ mice. This is explained by the fact that γc-containing cytokine receptors utilize JAK1 in association with their ligand-specific receptor subunit. Other cytokines that are dependent on JAK1 include those that use gp130 cytokine receptors and type II receptors (IL-10, IFN-γ, and IFN-α/β). *JAK1* LOF mutations in humans are associated with primary immunodeficiency, whereas GOF mutations are associated with systemic autoimmunity and hypereosinophilic syndrome.

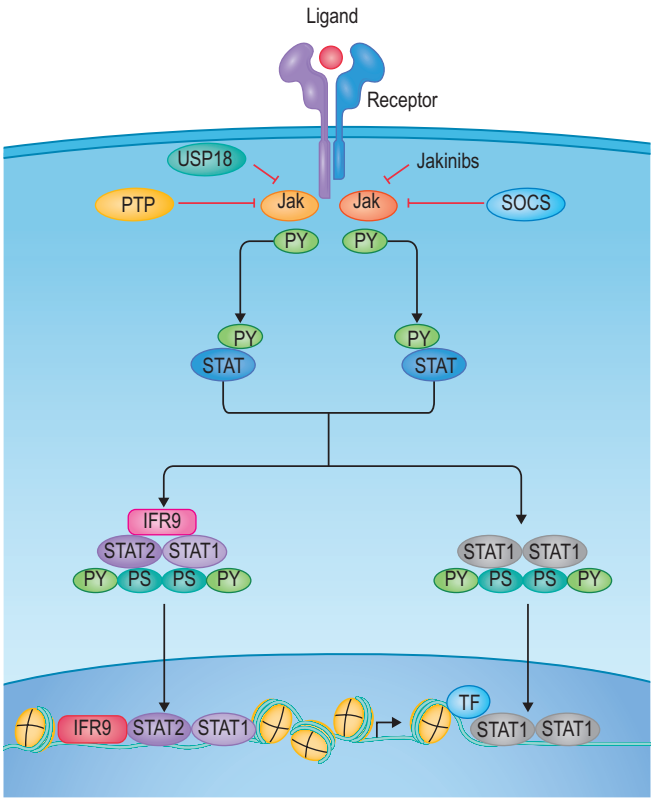


FIG. 14.2 Signal transduction cascade downstream of type I and II cytokine receptors: the role of Janus kinases *(JAKs)*, STATs, and the molecular machinery upstream of cytokine-driven gene regulation. Cytokine binding to this class of receptors activates receptor-associated JAKs. Activated JAKs tyrosine phosphorylate each other and the cytokine receptor *(PY)*, allowing docking of signal transducers and activators of transcription *(STATs)*, which are also tyrosine phosphorylated. STATs dimerize, translocate to the nucleus, bind DNA and regulate transcription. Types I and III interferon activate a complex of STAT1/STAT2/IRF9. Signaling is inhibited by suppressor of cytokine signaling *(SOCS)* proteins, protein tyrosine phosphatases *(PTP)*, and USP18. Janus kinase inhibitors *(Jakinibs)* are small molecule inhibitors approved for the treatment of multiple autoimmune disorders.

Gene-targeting of *Jak2* is embryonically lethal, principally because JAK2 is essential for EPO function and the mice fail to form blood. Conversely, GOF mutations in the pseudokinase domain of *JAK2* underlie most cases of polycythemia vera. A large number of other cytokines signal via JAK2 (Table 14.1).

Tyk2$^{-/-}$ mice have increased viral susceptibility and impaired IL-12 and IL-23 signaling. *TYK2* LOF mutations are associated with immunodeficiency, and *TYK2* polymorphisms reduce the risk of SLE.

STATs

Members of the signal transducer and activator of transcription (STAT) family of DNA-binding proteins serve a key role in transducing signals from cytokine receptors on the cell surface to the nucleus, where they regulate gene transcription. STATs are latent, cytosolic transcription factors that have SH2 domains (phosphotyrosine-binding modules) that allow them to be recruited to phosphorylated cytokine receptors (see Fig. 14.2). Different STATs bind to specific cytokine receptors (see Table 14.1). STATs are themselves tyrosine phosphorylated by JAKs, which promote STAT dimerization. STATs then translocate to the nucleus, bind DNA, and regulate transcription.

There are seven mammalian STATs: STAT1, STAT2, STAT3, STAT4, STAT5A, STAT5B, and STAT6. *Stat* knockout mice document the essential and specific functions of these transcription factors in transmitting cytokine signals. *Stat1*$^{-/-}$ mice develop normally but have extreme susceptibility to viral and some bacterial infections, consistent with the defects seen in mice with mutations of IFNs and IFNRs. *STAT1* LOF mutations in humans result in susceptibility to *Salmonella* and mycobacterial infections. GOF mutations cause chronic mucocutaneous candidiasis, along with viral susceptibility. *Stat2*$^{-/-}$ mice have viral susceptibility, as do humans with *STAT2* LOF mutations. Patients with *STAT2* GOF mutations have autoinflammatory disease and interferonopathy.

Gene targeting of *Stat3* leads to early embryonic lethality, in part due to interference with leukemia inhibitory factor (LIF) function. Conditional knockouts of *Stat3* in myeloid cells display exaggerated inflammatory responses due to failure of IL-10 signaling. STAT3 is also essential for Th17 cells. LOF mutations of *STAT3* underlie Hyper-IgE syndrome or Job syndrome associated with impaired IL-17 production. *STAT3* GOF mutations cause systemic autoimmunity. *STAT3* polymorphisms are associated with IBD (Chapter 75).

STAT4 is activated by IL-12. *Stat4*$^{-/-}$ mice develop normally but have defective Th1 differentiation and IFN-γ production combined with augmented Th2 development. *STAT4* LOF mutations are associated with tuberculosis, atypical mycobacterial susceptibility, and fungal susceptibility. *STAT4* polymorphisms are associated with RA, Sjogren syndrome, and SLE.

STAT5A and STAT5B are highly homologous but nonetheless have different functions. *Stat5a*$^{-/-}$ mice have impaired mammary gland development and failure of lactation, whereas *Stat5b*$^{-/-}$ mice have defective sexually dimorphic growth and growth hormone-dependent regulation of liver gene expression. *Stat5a/5b* doubly-deficient mice manifest increased perinatal lethality, decreased size, female infertility, and impaired lymphocyte development. *Stat5*$^{-/-}$ mice develop lymphoproliferative disease reminiscent of IL-2- and IL-2R-deficient mice related to loss of Treg cells and expansion of Tfh and Th17 cells. Patients with *STAT5B* mutations have short stature and immune dysregulation. Somatic *STAT5* GOF mutations are associated with eosinophilia, urticaria, leukemia, and lymphoma.

STAT6 is activated by IL-4 and IL-13. *Stat6*$^{-/-}$ mice have defective Th2 development with defective IgE responses following parasitic infection. Lack of STAT6 dramatically attenuates allergic and asthmatic disease in animal models.

Attenuation of Type-I and Type-II Cytokine Signaling

Perhaps as important as the triggers that initiate signal transduction are the mechanisms that extinguish responses (see Fig. 14.2). There are several families of proteins involved in downregulating cytokine signaling. Among these are phosphatases, cytokine-inducible inhibitor molecules, and transcriptional repressors. The phosphatase SHP-1 interacts with cytokine receptors and downregulates signaling. Mice with the naturally occurring "motheaten" mutation in SHP-1 die at an early age from autoimmune disease.

Suppressors of cytokine signaling (SOCS) are SH2-containing proteins that bind to either cytokine receptors or JAKs to inhibit signaling. There are at least eight members of this family. Largely

due to systemic hyper-responsiveness to IFN-γ, *Socs-1*$^{-/-}$ mice die within a few weeks of birth. SOCS-2 has been shown to regulate Th2 differentiation and allergic responses and SOCS-3 regulates IL-6's actions including Th17 differentiation.

USP18 and ISG15 negatively regulate IFN signaling. LOF mutations of *USP18* and *ISG15*, along with mutations of *STAT2* that abrogate interaction with ISG15, cause interferonopathy.[30,31]

KEY CONCEPT

Properties of the Tumor Necrosis Factor Receptor Superfamily

- Activation of a TNF receptor can lead to a wide range of effects, from proliferation to apoptosis.
- Transduction of signals through TRAFs leads to the enhancement of survival.
- Signaling through death domains leads to the induction of apoptosis.

THE TNF CYTOKINE AND RECEPTOR SUPERFAMILY

This large family of structurally related ligands, receptors, and inhibitory decoy receptors has various roles in both the immune system and beyond. The first two members of this family discovered were TNF and lymphotoxin-α (LTα; formerly named TNF-β). They are principally secreted by activated myeloid and T cells. They share pro-inflammatory functions and belong to a large family of related molecules that includes CD30 ligand (CD30L), CD40L, FASL, and TRAIL. The TNF cytokine family contains 19 ligands and 29 receptors, each of which exhibits marked differences in tissue expression, ligand specificity, receptor binding, and biological function (Fig. 14.3). This section describes general aspects of TNF and TNFR (TNF receptor) biology, with examples from the best-studied TNF-family members and others that show promising potential in disease intervention and therapy. More complete listings of TNF-family cytokines and their receptors can be found in Tables 14.2 and 14.3. A listing of recombinant cytokines and biological agents currently in clinical use or being tested for clinical application at the time this chapter was written can be found in Table 14.4.

Ligand and Receptor Structure

Much of our understanding of the structural and functional characteristics of the TNF ligand and receptor superfamilies has come from the analysis of TNF (TNFSF2), LTα (TNFSF1), FASL (TNFSF6), and their receptors. All TNF ligands are synthetized as type II transmembrane glycoproteins. TNF and LTα are closely related homotrimeric proteins (32% identity). Human TNF contains an amino-terminal sequence that anchors as a type II cell membrane protein. A secreted form of TNF is generated through enzymatic cleavage of membrane-bound TNF by a metalloproteinase termed TNF-α–converting enzyme (TACE). Both soluble and membrane-bound forms of TNF are homotrimers held together by noncovalent interactions between each subunit's trimerization domain. Both TNF forms are biologically active, but have different affinities for the two TNF receptors and thus can exhibit different biological properties.

Unlike TNF, LTα is synthesized as a secreted glycoprotein homotrimer. It can bind both TNF receptors with affinities comparable to those of TNF and has similar biological effects. LTα1β2 is a heteromeric membrane-bound form of LT (mLT)

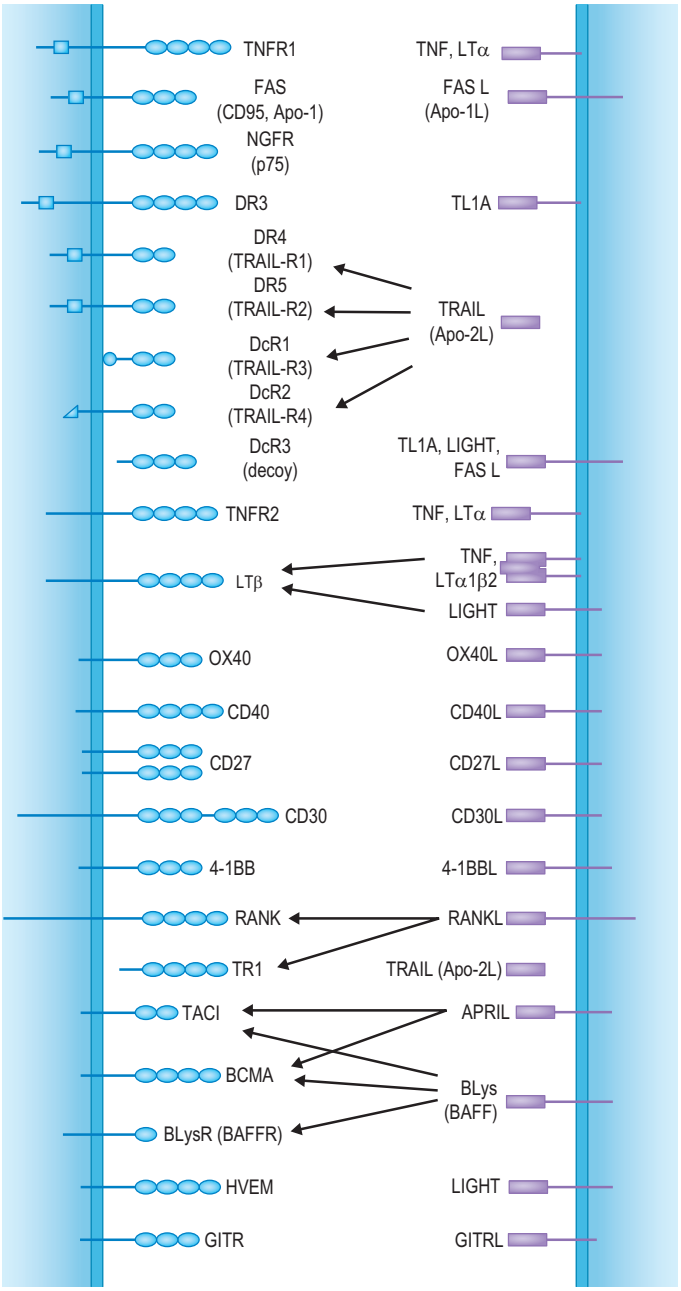


FIG 14.3 Schematic Representation of Tumor Necrosis Factor *(TNF)* Ligand and Receptor Superfamily Members.

composed of LTα subunit noncovalently linked to two molecules of LTβ. mLT is not cleaved by TACE and is thought to exist exclusively as a membrane-bound complex. mLT specifically interacts with another member of the TNF receptor superfamily, the lymphotoxin β receptor (LTβR). TNF and the two LT subunits are encoded by closely linked, single-copy genes situated in the class III major histocompatibility locus at chromosome 6p21.3 in humans (Chapter 5).

All TNF receptor family members are type-I transmembrane glycoproteins. The two receptors for TNF (and LTα) are designated TNFR1 (TNFRSF1A, also known as p60 in humans and p55 in mice) and TNFR2 (TNFRSF1B, also known as p80 in humans and p75 in mice). These receptors are characterized by cysteine repeat domains (CRD) of about 40 amino acids in their amino-terminal extracellular domains, each consisting of

TABLE 14.2 Tumor Necrosis Factor Superfamily Cytokines

Symbol	Common Name	Aliases	Binds to Receptor(s)	OMIM ID	Key Functions	Phenotype Associated With Overexpression	Phenotype Associated With Deficiency	Human Genetic Disease Associations
TNFSF1	Lymphotoxin alpha (Ltα)	LT, TNFB, TNFSF1	TNFR2 (1B), TNFR1 (1A), HVEM (14)	153440	Lymphoid organ formation		Absence of LN and PP, defective GC formation	
TNFSF2	Tumor necrosis factor (TNF)	DIF, TNFA, TNFSF2, CACHECTIN	TNFR2 (1B), TNFR1 (1A)	191160	Inflammation	Wasting syndrome, arthritis	Defective GC formation, resistance to endotoxic shock and experimental arthritis	TNF2 (G-308A) promotor polymorphism associated with increased susceptibility to septic shock, asthma, and RA severity
TNFSF3	Lymphotoxin beta (LTβ)	p33, TNFC, TNFSF3	As a $\beta_2\alpha_1$ heterotrimer with LTα binds to LTβ receptor (3)	600978	Lymphoid organ formation	Ectopic lymphoid organ formation		
TNFSF4	OX40 Ligand	GP34, OX4OL, TXGP1, CD134L, OX-40 L	OX40 (4)	603594	CD4 T-cell expansion, survival, and Th2 development	Increased Th2 responses	Th2 deficiency, blockade improves EAE	Associated with SLE in GWAS
TNFSF5	CD40 Ligand	IGM, IMD3, TRAP, gp39, CD154, CD40L, HIGM1, T-BAM	CD40 (5)	300386	Co-stimulation and differentiation of B cells and APCs	Constitutive expression in B cells or keratinocytes leads to SLE-like syndrome	Immunodeficiency due to defective Ig class switching and germinal center formation	X-linked hyper-IGM syndrome associated with CD40L LOF mutations
TNFSF6	Fas Ligand	FASL, CD178, CD95L, APT1LG1	Fas (6), DcR3 (6B)	134638	Mediator of CD4 T-cell apoptosis due to restimulation and apoptosis in other cell types		Lymphadenopathy and systemic autoimmunity	Autoimmune lymphoproliferative syndrome (ALPS) type Ib
TNFSF7	CD27 Ligand	CD70, CD27L, CD27LG	CD27 (7)	602840	T cell co-stimulation	T cell hyperactivation eventually leading to HIV-like immunodeficiency		
TNFSF8	CD30 Ligand	CD153, CD30L, CD30LG	CD30 (8)	603875				
TNFSF9	4-1-BB Ligand	4-1BB-L	4-1BB (9)	606182	T cell co-stimulation			
TNFSF10	TRAIL (TNF-like apoptosis inducing ligand)	TL2, APO2L, TRAIL, Apo-2 L	DR4 (10A), DR5 (10B), DcR1 (10C), DcR2 (10D)	603598	DC apoptosis, NK cell-mediated tumor cell killing		Defective NK-mediated tumor eradication	

TABLE 14.2 Tumor Necrosis Factor Superfamily Cytokines—Cont'd

Symbol	Common Name	Aliases	Binds to Receptor(s)	OMIM ID	Key Functions	Phenotype Associated With Overexpression	Phenotype Associated With Deficiency	Human Genetic Disease Associations
TNFSF11	RANK-L	ODF, OPGL, sOdf, RANKL, TRANCE, hRANKL2	RANK (11A)	602642	Mediates osteoclast formation and bone remodeling, stimulates APCs			
TNFSF12	TWEAK	APO3L, DR3LG, TWEAK, MGC20669	TWEAK-R (12A)	602695	Potential role in inflammation and lymphocyte function			
TNFSF13	APRIL	APRIL, TALL2, TWE-PRIL	TACI (13B), BCMA (17)	604472	Promotes T-cell independent type-2 responses via TACI interactions	Overexpression in T cells produces prolonged T-cell survival and enhanced TI-2 responses		
TNFS-F13B	BlyS, BAFF	BAFF, BLYS, TALL1, THANK, ZTNF4	TACI (3B), BAFF-R (13C), BCMA (17),	603969	Promotes B-cell maturation, plasmablast survival	SLE-like systemic autoimmunity and arthritis		
TNFSF14	LIGHT	LTg, TR2, HVEML, LIGHT	HVEM (14), LT-βR (3), DcR3 (6B)	604520	CD8 T-cell and APC costimulation	Inflammation, T-cell hyperactivation, Th1 bias	Defective CD8 T-cell costimulation	
TNFSF15	TL1A	TL1, TL1A, VEGI	DR3 (25)	604052	Ligand for DR3 (TNFRSF25) on lymphocytes	T-cell activation, IL-13 dependent small intestinal hyperplasia and inflammation	Reduced immunopathology in T-cell-dependent autoimmune diseases	Common variant associated with inflammatory bowel disease through GWAS
TNFSF18	GITR Ligand	TL6, AITRL, GITRL, hGITRL	GITR (18)	603898	T-cell costimulation, $CD25^+$ regulatory T cells			
ED1	ectodermal dysplasia 1, anhidrotic (EDA1)	EDA, HED, EDA1, XHED, XLHED	EDAR	305100	Tooth, hair, and sweat gland formation			X-linked ectodermal dysplasia

three or four cysteine-rich regions involved in intrachain disulfide bonds. The cytoplasmic domains of these receptors lack intrinsic enzymatic activity. However, the CRDs harbor a specific sequence to which adaptor molecules will bind and activate various signaling pathways that can lead to a remarkably diverse set of cellular responses. These include differentiation, activation, release of inflammatory mediators, and apoptosis.

Family Members and Their Actions

Tumor Necrosis Factor, Lymphotoxin-α, and Receptors

TNF is a major physiological mediator of inflammation and of the physiopathology of septic shock. It is one of the first cytokines made in response to TLR4 stimulation by bacterial (LPS) and other TLR ligands.[32] IFN-γ also induces TNF and augments its effects. TNF upregulates MHC class I and class II expression, activates phagocytes, and induces mononuclear phagocytes to produce IL-1, IL-6, chemokines, and TNF. TNF increases adhesion of cells to endothelium and can be cytotoxic, particularly to tumor cells.

TNF-deficient mice are resistant to septic shock induced by high doses of LPS but have increased susceptibility to bacterial infection. The dual role of TNF in controlling bacterial replication and in septic shock emphasizes the point that, although the goal of an immune response is to eliminate invading microorganisms, the response itself can be injurious to normal host tissues. Septic shock is an extreme example of this. Although the primary source of TNF is mononuclear

TABLE 14.3 Tumor Necrosis Factor Superfamily Receptors (Receptors in Italics Have a C-Terminal Death Domain)

Symbol	Common Name(s)	Aliases	Binds to Ligand(s)	OMIM ID	Key Functions	Phenotype Associated With Deficiency	Human Genetic Diseases
TNFRSF1A	Tumor necrosis factor receptor 1 (TNF-R1)	FPF, p55, p60, TBP1, TNF-R, TNFAR, TNFR1, p55-R, CD120a, TNFR55, TNFR60, TNF-R-I, TNF-R55, MGC19588	TNF-α (2), Lτα (1)	191190	Mediates TNF-induced inflammation (and apoptosis in some cells)	Resistance to TNF-induced arthritis models, resistant to endotoxic shock, increased sensitivity to bacterial pathogens	Periodic fever syndrome (TRAPS) associated with heterozygous extracellular mutations; locus associated with primary biliary cirrhosis and multiple sclerosis in GWAS
TNFRSF1B	Tumor necrosis factor receptor 2 (TNF-R2)	p75, TBPII, TNFBR, TNFR2, CD120b, TNFR80, TNF-R75, p75TNFR, TNF-R-II	TNF-α (2), Lτα (1)	191191	May enhance pro-apoptotic effect of TNFR1	Still susceptible to TNF-induced arthritis models, defective CD8T cell apoptosis after restimulation, increased sensitivity to bacterial pathogens	
TNFRSF3	Lymphotoxin β receptor	LTBR, CD18, TNFCR, TNFR-RP, TNFRSF3, TNFR2-RP, LT-BETA-R, TNF-R-III	LIGHT (14), LTβ (3)	600979	Lymphoid organ formation	No lymph nodes and Peyer patches, defective GC formation	
TNFRSF4	OX40	OX40, ACT35, CD134, TXGP1L	OX40L (4)	600315	T-cell co-stimulation	Defective CD4 T-cell responses	
TNFRSF5	CD40	p50, Bp50, CD40, CDW40, MGC9013	CD40L (5)	109535	Co-stimulation and differentiation of B cells and APCs	Defective Ig class switching and germinal center formation	Autosomal hyper-IgM syndrome associated with LOF mutations; locus associated with RA through GWAS
TNFRSF6	*Fas, CD95*	FAS, APT1, CD95, APO-1	FasL (6)	134637	Apoptosis of restimulated CD4 T cells, B cells, other cells	Defective apoptosis of restimulated CD4 T cells	Autoimmune lymphoproliferative syndrome (ALPS) associated with heterozygous dominant-interfering mutations
TNFRSF6B	Decoy receptor 3	M68, TR6, DCR3, DJ583P15.1.1	FasL (6), TL1A (15), LIGHT (14)	603361	Soluble decoy receptor for FasL, LIGHT, and TL1A; may have a role in tumor immune evasion		Associated with inflammatory bowel disease in GWAS
TNFRSF7	CD27	T14, CD27, S152, Tp55, MGC20393	CD27L (7)	186711	T-cell co-stimulation	Defective T-cell responses	
TNFRSF8	CD30	CD30, KI-1, D1S166E	CD30L (8)	153243	Inhibition of CD8 T-cell effector function		
TNFRSF9	4-1BB, CD137	ILA, 4-1BB, CD137, CDw137, MGC2172	4-1BBL (9)	602250	T-cell co-stimulation	Defective CD8 T-cell responses	Locus associated with ulcerative colitis in GWAS
TNFRSF10A	*Death Receptor 4 (DR4)*	DR4, APO2, MGC9365, TRAILR1, TRAILR-1	TRAIL (10)	603611	Mediates DC and tumor cell apoptosis		
TNFRFRS-F10B	*Death Receptor 5 (DR5)*	DR5, KILLER, TRICK2, TRICKB, ZTNFR9, TRAILR2, TRICK2A, TRICK2B, TRAIL-R2, KILLER/DR5	TRAIL (10)	603612	Mediates DC and tumor cell apoptosis		
TNFRSF10C	Decoy receptor 1	LIT, DCR1, TRID, TRAILR3	TRAIL (10)	603613	GPI-linked decoy receptor, interferes with TRAIL function		
TNFRSF10D	Decoy receptor 2	DCR2, TRUNDD, TRAILR4	TRAIL (10)	603614	Transmembrane decoy receptor, interferes with TRAIL function		

TABLE 14.3 Tumor Necrosis Factor Superfamily Receptors (Receptors in Italics Have a C-Terminal Death Domain)—Cont'd

TNFRSF11A	(Receptor activator of NF-κB) RANK	OFE, ODFR, PDB2, RANK, TRANCER	RANKL (11)	603499	Mediates DC co-stimulation and osteoclast maturation and activation	Osteopetrosis due to osteoclast deficiency, no lymph nodes, impaired B-cell development	
TNFRSF11B	Osteoprotegerin (OPG)	OPG, TR1, OCIF, MGC29565	TRAIL (10), RANKL (11)	602643	Soluble decoy receptor for RANK	Osteoporosis, arterial calcification	
TNFRSF12A	TWEAK-receptor	FN14, TWEAKR	TWEAK (12)	605914			
TNFRSF13B	TACI	TACI	APRIL (13), BAFF (13B)	604907	May inhibit some of the pro-survival effects of BAFF-R	Decreased TI-2 B-cell responses, but B-cell hyperplasia and autoimmunity	LOF mutations associated with familial CVID
TNFRSF13C	BAFF receptor (BAFF-R)	BAFF/BLyS receptor 3	BAFF (13B)	606269		Impaired survival of immature transitional B cells	
TNFRSF14	Herpes virus entry mediator (HVEM)	TR2, ATAR, HVEA, HVEM, LIGHTR	LIGHT (14), herpes viruses	602746			Locus associated with RA and ulcerative colitis in GWAS
TNFRSF16	NGF-R	TNFRSF16, p75(NTR)	NGF (not a TNF family member)	162010	NGF receptor (evolutionary outlier as NGF is not classic TNF family molecule)	Defective sensory neuron innervation, impaired heat sensitivity	
TNFRSF17	B cell maturation antigen (BCMA)	BCM	APRIL (13), BAFF (13B)	109545		Apparently, no B-cell phenotype	
TNFRSF18	Glucocorticoid-induced TNF receptor (GITR)	AITR, GITR, GITR-D	GITRL (18)	603905	T-cell co-stimulation, marker for $CD4^+CD25^+$ Treg cells, modulates Treg function	T-cell hyperactivation	
TNFRSF19	Toxicity and JNK inducer (TAJ)	TROY, TRADE, TAJ-alpha		606122	Similar to EDAR, expressed in skin and brain		
TNFRSF19L	RELT	RELT, FLJ14993			Possible T-cell co-stimulator		
TNFRSF21	*Death Receptor 6 (DR6)*	DR6, BM-018		605732	Negative regulator of B- and T-cell responses	Enhanced T- and B-cell activation	
TNFRSF25	*Death Receptor 3 (DR3)*	*DR3, TR3, DDR3, LARD, APO-3, TRAMP, WSL-1, WSL-LR, TNFRSF12*	*TL1A (15)*	*603366*		Impaired thymic negative selection, reduced immunopathology and T-cell accumulation at site of autoimmune disease	
EDAR	*Ectodysplasin 1, anhidrotic receptor*	DL, ED3, ED5, ED1R, EDA3, EDA-A1R	E1	604095	Tooth, hair, and sweat gland formation	Abnormal tooth, hair, and sweat gland formation	Ectodermal dysplasia
XEDAR	XEDAR: ectodysplasin A2 isoform receptor	EDAA2R, EDA-A2R	EDA-A2	300276			

Locuslink ID: Gene 'homepage' curated by NCBI. Go To and type the locuslink ID in the search window.
OMIM: ID in the Online Mendelian Inheritance in Man Database. Go to and type in the OMIM ID in the search window.
GC, Germinal center; *LN*, lymph node; *PP*, Peyer patch.

TABLE 14.4 Recombinant Cytokines and Biological Agents Currently in Clinical Use or Being Tested for Clinical Application

Name	Target	Type	Phase	Indications
Anakinra	IL-1 α and β	Human recombinant IL-1Ra	In clinic	Autoinflammatory syndromes, RA
Rilonacept	IL-1 α and β	Human IL-1R–Fc (IgG1) fusion protein	In clinic	Autoinflammatory syndromes, gout, JIA
Canakinumab	IL-1β	mAb	In clinic	Autoinflammatory syndromes, JIA
Bermekimab	IL-1α	mAb	2	AD, HS
Basiliximab	IL-2Rα	mAb	In clinic	Transplantation, uveitis, ulcerative colitis
Mepolizumab, Reslizumab	IL-5	mAb	In clinic	Eosinophilic asthma, EGPA
Benralizumab	IL-5Rα +ADCC	mAb	In clinic	Eosinophilic asthma
Tocilizumab, Sarilumab	IL-6R	mAb	In clinic	RA, JIA, GCA, cytokine release syndrome
Siltuximab	IL-6	mAb	In clinic	Castleman disease
Clazakizumab, Sirukumab	IL-6	mAb	2	RA, Antibody-mediated graft rejection
Ustekinumab	IL-12/23 p40	mAb	In clinic	Psoriasis, PsA IBD,
Tralokinumab, Lebrikizumab, Anrukinzumab	IL-13	mAb	2, 3	AD
Dupilumab	IL-13R/IL4R	mAb	In clinic	Asthma, AD, rhinosinusitis
Secukinumab, Ixekizumab	IL-17A	mAb	In clinic	Psoriasis, PsA, ankylosing spondylitis, HS
Bimekizumab	IL-17A, IL-17F	mAb	3	Psoriasis, PsA
Brodalumab	IL-17RA	mAb	In clinic	Psoriasis
GSK1070806	IL-18	mAb	2	IBD
Tadekinig Alfa	IL-18	IL-18 BP	2	Systemic-onset JIA, Still disease
Guselkumab, Risankizumab Tildrakizumab	IL-23	mAb	In clinic	Psoriasis, PsA, HS (Ph3), IBD (Ph3)
Nemolizumab	IL-31	mAb	2	Prurigo nodularis, atopic dermatitis
Etokimab	IL-33	mAb	2	Chronic rhinosinusitis, atopic dermatitis
SAR440340/REGN3500	IL-33	mAb	2	Asthma
GSK3772847	IL-33R	mAb	2	Asthma
Spesolimab (BI655130)	IL-36R	mAb	2	Pustular psoriasis, IBD
Imsidolimab	IL-36R	mAb	2	Pustular psoriasis
Anifrolumab	IFNAR1	mAb	3	SLE
Emapalumab	IFNγ	mAb	in clinic	Hemaphagocytic lymphohistocytosis
Mavrilimumab, Gimsilumab	GM-CSF	mAb	2	RA, giant cell arteritis, Covid-19
Fresolimumab	TGF-β 1, 2, and 3	mAb	2	Systemic sclerosis, pulmonary fibrosis
Tezepelumab	TSLP	mAb	2	Asthma
Etanercept	TNF	TNFR2-Fc (IgG1) fusion protein	In clinic	RA, JIA, PsA, plaque psoriasis, ankylosing spondylitis
Infliximab, Adalimumab, Golimumab	TNF	mAb	In clinic	RA, psoriasis, Crohn disease, ankylosing spondylitis, PsA, ulcerative colitis, HS
Certolizumab	TNF	PEGylated Fab	3	RA, JIA, PsA, plaque psoriasis, ankylosing spondylitis
Belimumab	BAFF/BLyS	mAb	In clinic	SLE
Ianalumab	BAFFR +ADCC	mAb	2	SLE, pSS, autoimmune hepatitis
Atacicept	APRIL/BAFF	TACI-Fc	2	SLE
Denosumab	RANKL	mAb	In clinic	Osteoporosis, hypercalcemia of malignancy, cancer
Iscalimab	CD40	mAb	2	Sjogren sundrome
VIB4920	CD40L	CD40-Fc	2	RA, pSS
Brentuximab-vedotin	CD30	mAb-drug conjugate	In clinic	Hodgkin lymphoma
Belantamab-mafodotin	BCMA	mAb-drug conjugate	In clinic	Multiple myeloma

Source: Pipeline (Drug pipeline information database by Citeline, Inc.).

phagocytes, TNF is also produced by T cells, NK cells, and mast cells.

Anti-TNF inhibitor mAbs and recombinant receptors are approved for multiple indications from arthritis to IBD. LTα shares many of the same biological effects as TNF, mainly because of its ability to bind the same receptors. However, LTβR plays a unique role in the development of secondary lymph nodes. As such, unlike with TNF, targeting of lymphotoxin signaling did not result in successful outcomes in clinical trials.

Fas Ligand (FasL) and Its Receptor, Fas/APO-1/CD95

Fas (also known as Apo-1, CD95, or TNFRSF6) is a type I integral membrane protein structurally related to TNFR1. Fas can trimerize and transduce proapoptotic signals upon binding of its ligand, FasL. Similar to TNF, FasL (CD95L) is synthesized as a type II membrane protein, expressed on activated B cells, T cells, and NK cells. Fas-induced apoptosis plays an essential role in the termination of T-cell responses, particularly in the peripheral immune system. Fas can also play a key role in the induction of cell death by cytotoxic T cells (CTLs) and NK cells (Chapter 12), where it functions in conjunction with perforin. Nonapoptotic functions of FasL include lymphocyte costimulation and T-cell differentiation into short-lived effector memory cells.[33,34]

CD40 Ligand and CD40

CD40 is expressed by a variety of cell types, including B cells, DCs, monocytes, macrophages, and endothelial cells. It plays a major

co-stimulatory role in B-cell differentiation, immunoglobulin recombination, and promotes cell survival through the induction of BCL-2 family members. Studies of both CD40-deficient mice and patients with hyper-IgM syndrome (Chapter 33) reveal that its function extends beyond the humoral immune response; CD40 signaling also plays a role in cell-mediated immunity. CD40L (CD154) is expressed by activated CD4 T cells that can bind to and activate CD40 by cell–cell contact.

CD40L on T cells triggers antigen-presenting cell (APC) activation, including the upregulation of the CD28 ligands: B7-1 and B7-2. This indirectly boosts co-stimulation of the T-cell response. Because of its critical role in mediating T-cell help in B-cell class switching and autoantibody formation, blocking CD40L/CD40 interactions has been a therapeutic goal in autoimmune diseases. Clinical trials of a blocking anti-CD40L antibody in SLE showed promising results, but were halted because of thrombotic events likely resulting from the off-target effects of these antibodies on CD40L expressed on platelets. Recent drugs targeting CD40L have been engineered to limit platelet activation. Anti-CD40 mAbs, including iscalimab, are being tested in Sjogren syndrome, SLE, and other indications.[35]

OX40 Ligand and OX40

OX40 is primarily upregulated by T cells upon antigen stimulation, but it is also present on other cell types including NK cells, NKT cells, and neutrophils. OX40L (OX40 ligand) is mainly expressed on APCs, including activated B cells, macrophages, and DCs. Additionally, OX40L has also been found on mast cells, endothelial cells, Langerhans cells, NK cells, and activated T cells. Co-stimulation through OX40 promotes survival, clonal expansion, and cytokine production. Human and mouse studies suggest that OX40 signaling is important in rheumatic diseases, and *OX40L* polymorphisms are associated with Sjögren syndrome. Several blocking antibodies have been generated with encouraging effects in early trials. Given their function, co-stimulatory receptors such as CD40 and OX40 are thought to be strong therapeutic candidates for enhancing the immune response against tumors.

TL1A and DR3

TL1A (TNFSF15) is mainly expressed by APCs upon stimulation by LPS or immune complexes. Its receptor—DR3—is found primarily on activated T cells and innate lymphocytes. This signaling pathway plays critical roles in diverse autoimmune disease models by facilitating T-cell accumulation at the site of inflammation. Both TL1A and DR3 have been reported to be upregulated in human biopsy samples and in animal models of IBD. A blocking TL1A antibody has been developed with a potential application in IBD. The TL1A-DR3 axis appears to play a role in the development of RA. A role for TL1A in host defense against infection has thus far been limited in controlling T-cell responses to *Salmonella* and select viral infections in various mouse models.

Other TNF-Family Cytokines

Other members of the TNF family play various roles in the development and function of the immune system. The vast majority of them participate in co-stimulation of lymphocyte activation. Many of those pathways are now being targeted as therapeutic candidates. The TNF family ligand BAFF (BlyS/TALL1/TNFSF13B) promotes B-cell maturation and antibody secretion and can bind three distinct receptors, TACI (TNFRSF13B), BADD-R (TNFRSF13C), and BCMA (TNFRSF17). BCMA is present on multiple myeloma cells but absent in normal tissue. Belantamab mafodotin—a BCMA mAb conjugated to monomethyl auristatin F—is approved for treatment of multiple myeloma. Brentuximab vedotin is a mAb-drug conjugate targeting CD30, which is approved for the treatment of Hodgkin lymphoma (Chapter 78).

Some TNF members are also involved in development and function outside the immune system. Ectodysplasin A (EDA) affects the formation of tooth germs and sweat glands, and receptor activator of NF-kB ligand (RANKL) regulates bone regeneration and remodeling. The anti-RANKL mAb, denosumab, is approved for treatment of osteoporosis.

Signaling

The TNF receptor superfamily can be divided into three subfamilies on the basis of the types of intracellular signaling molecules recruited (e.g., FADD, TRADD or TRAF) (Fig. 14.4).[36] The cytoplasmic domains of TNFR1, FAS, Death Receptor 3 (DR3), DR4, and DR5 contain a conserved motif termed the death domain (DD). This element is required for recruitment of DD-containing adaptor molecules involved in initiating apoptosis and thus are termed "death receptors" (Chapter 17). The function of some death receptors can be regulated by decoy receptors that bind ligands but lack functional intracellular domains. Other TNF receptors that lack DD (e.g., CD27, CD30, CD40, HVEM, TNFR2, LT-βR, OX-40, 4-1BB) associate with different types of adapter molecules, mostly members of the TRAF (TNFR-associated factor) family.

Death Domains: TRADD and FADD

FAS recruits a Fas-associated protein with death domain (FADD) to its cytoplasmic DD, leading to the rapid formation of the death-inducing signaling complex (DISC), which contains FADD and caspase-8, thereby permitting activation of downstream caspases. Caspase-8 is recruited through a structurally related domain termed the death-effector domain (DED). The FADD DED contains two hydrophobic patches, not present in the DD, that are necessary for binding to the DEDs in the pro-domain of caspase-8 and for apoptotic activity.[37] Under low FASL oligomerization, caspase-8 can participate in preventing necroptosis and can also lead to non-apoptotic signaling.

TNF-induced apoptosis upon TNFR1 ligation is mediated through recruitment of TRADD to the receptor via the DD motifs found in both molecules (Fig. 14.4). In turn, TRADD recruits FADD, also through their DD, leading to the formation of an intracellular DISC.[37] This initiates the apoptotic pathway. TRADD contains a TRAF-binding motif, which contributes to TRAF-dependent activation of inflammatory responses by inducing the NF-kB and mitogen-activated protein kinase (MAPK) pathways. Although cell death in tumor cells can be induced by TNF, the most common result of TNFR1 ligation in primary immune cells is inflammation and, sometimes, protection from TNF-induced apoptosis. Activation of proapoptotic and pro-inflammatory signaling by TNFR1 proceeds in sequential steps.

TNF receptor members lacking DDs contain short peptide consensus sequences that enable recruitment of the TRAF proteins. Structural studies of TRAFs have revealed a mushroom-like structure, with a trimer of the three TRAF subunits stabilized by a stalk-like coiled-coil domain. TRAF proteins recruit and activate protein complexes that ultimately induce

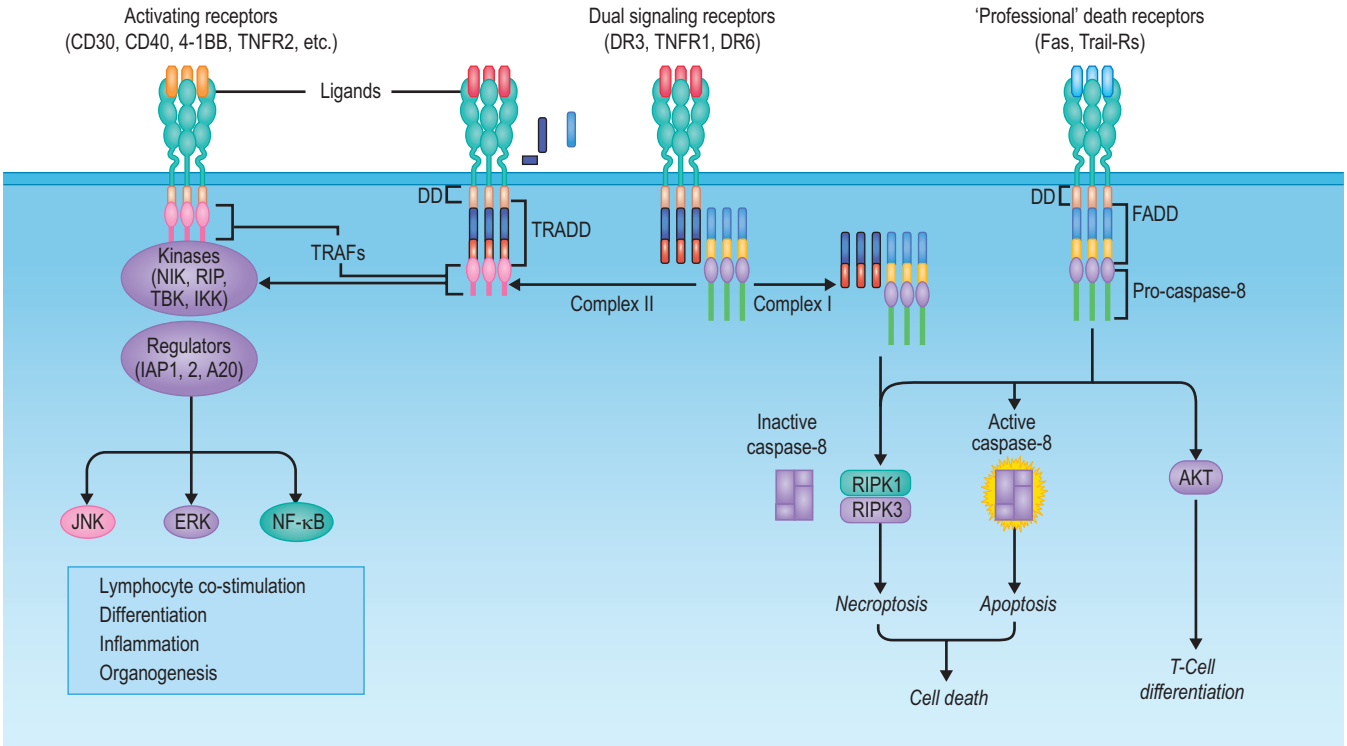


FIG. 14.4 Signal Transduction Cascade Downstream of the Tumor Necrosis Factor *(TNF)* Family of Cytokines and Their Receptors and the Resultant Effects on Immune Cells.

NF-κB and MAPK pathways. Most TRAFs can function as E3 ubiquitin ligases through their RING domain, but detailed mechanisms remain elusive. TRAF6 mediates NF-κB activation by a number of TNF-family receptors. TRAF6 also associates with a protein complex that mediates K63-linked ubiquitination and activation of the inhibitor of κB kinase (IKK) complex, which consists of two catalytic subunits, IKKα and IKKβ, and a regulatory protein IKKγ or nuclear factor κB (NF-κB) essential modulator (NEMO). K63-linked ubiquitination activates IKK, leading to the phosphorylation and degradation of IκB (an inhibitor of NF-κB) and the release of active NF-κB subunits, which translocate to the nucleus and regulate the expression of a wide variety of genes involved in the inflammatory response.

LTβ receptor activates the IKK complex via the serine–threonine kinase NF-κB -inducing kinase (NIK). A naturally occurring LOF NIK mutation in mice termed alymphoplasia (aly) leads to a lack of lymph nodes and Peyer patches and disorganized splenic and thymic structures. This mutation, and the phenotype of LTβR knock-out mice, revealed the critical role of this receptor in normal lymph node development and the formation of "tertiary" lymphoid tissue in inflammation.

When a single TNF-family ligand, such as TNF, binds both a death receptor (TNFR1) and a non–death receptor (TNFR2), a number of mechanisms regulate receptor signaling and the cellular outcome. Rather than functioning in cell death, the physiological function of TNFR2 may be as a co-stimulator of lymphocyte proliferation.[38] Accumulated evidence suggests that TNF-induced death is not only determined by the complex activity of TNFR1-TNFR2 signaling but also from its crosstalk with other, equally intricate signaling networks driven by pattern recognition receptors (PRRs), inflammasomes, and IFNs.[39]

CLINICAL RELEVANCE

TNFR Superfamily Cytokines and Receptors and Disease

- Dominant mutations of the gene encoding TNFR1 are associated with autosomal dominant periodic fever syndromes known as TNFR1-associated periodic syndromes (TRAPS).
- LOF mutations in the gene encoding CD40L are associated with X-linked hyper-IgM syndrome (X-HIM).
- Dominant interfering mutations in *TNFRSF6,* encoding the Fas receptor are associated with autoimmune lymphoproliferative syndrome (ALPS).
- RA often responds to therapeutic use of TNF antagonists.

Clinical Relevance

Mutations affecting *TNFR1* are associated with periodic fever syndromes (Chapter 37). Patients with the TNFR1-associated periodic syndrome (TRAPS) have missense mutations in exons encoding the extracellular regions of the receptor that cause intracellular accumulation and TNF-independent signaling that amplify inflammatory responses through the wild-type TNFR1.[40] Both blocking TNF with etanercept and blocking IL-1 have shown efficacy in reducing symptoms in TRAPS.[41]

The role of Fas signaling in the regulation of the immune system in vivo was confirmed when the naturally arising *lpr* and *gld* mouse strains were found to harbor homozygous mutations of Fas and Fas ligand, respectively. Both of these mouse strains are characterized by lymphadenopathy and splenomegaly due to the accumulation of unusual $CD4^{-}CD8^{-}$ T cells as well as the production of autoantibodies. Heterozygous dominant negative mutations of FAS cause autoimmune lymphoproliferative syndrome (ALPS) (Chapter 34).

The gene encoding CD40 ligand is defective in X-linked hyper-IgM syndrome (X-HIM) in which affected male children generate only IgM antibodies, many of which are autoantibodies (Chapter 33). Patients with X-HIM frequently suffer opportunistic infections, usually bacterial, and have increased susceptibility to cancer.

The physiological role of the BAFF receptor in mouse B-cell development is illustrated by BAFF-R mutations in the A/WySnJ mouse, which lacks peripheral B cells. TACI knockout mice have hyperactive B cells; but in humans, dominant negative TACI mutations have been found in patients with common variable immunodeficiency that affect B-cell numbers and function (Chapter 33), which indicates that in humans TACI serves as a positive modulator of B cells. Belimumab is an anti-BAFF mAb approved for the treatment of SLE.

Polymorphisms of TNFSF15 are associated with susceptibility to IBD. IBD patients have increased levels of TL1A and DR3 at the site of inflammation. In RA, TL1A is specifically elevated early in the disease course and in at-risk first-degree relatives. Interestingly, TL1A levels decline after TNF blockade, placing TL1A downstream of TNF.

The gene encoding EDA is defective in X-linked hypohidrotic ectodermal dysplasia (XLHED), a disorder leading to impaired development of sweat glands and teeth. The most severe clinical outcome for patient with XLHED is lethal hyperthermia occurring after birth. In a small clinical sample, prenatal treatment with protein-replacement therapy *in utero* was able to restore tooth and sweat gland development.[42]

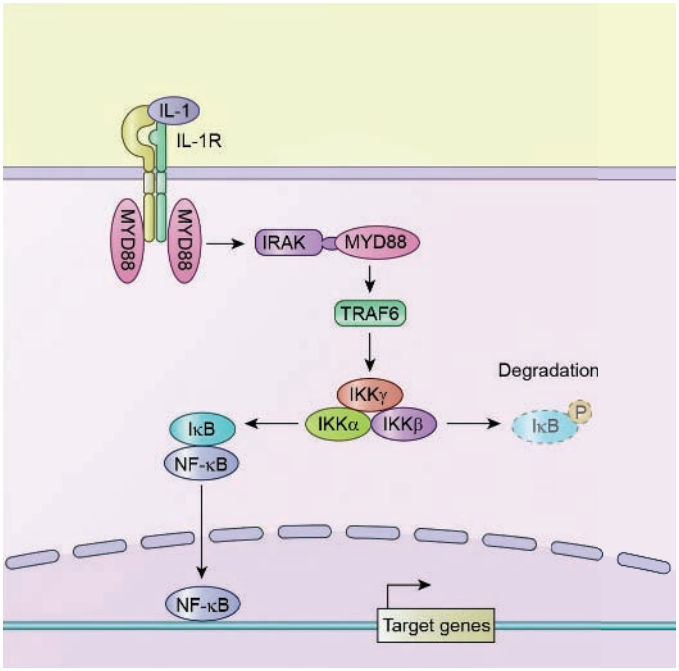


FIG. 14.5 Mechanism of Signal Transduction Downstream of Interleukin (IL)-1R and Related Receptors. IL-1 uses the adapter protein MyD88 to link to TRAF6 to activate inhibitor of kappa B kinases *(IKK)*. Phosphorylated inhibitor of kappa B (IκB) is degraded upon activation, allowing nuclear factor kappa B (NF-κB) to translocate to the nucleus, bind DNA, and regulate transcription.

KEY CONCEPT

Properties of the Interleukin-1R/Toll-Like Receptor Family

- IL-1 plays a key role in fever and acute-phase responses.
- IL-18 augments Th1 differentiation.
- TLRs modulate pro-inflammatory signals in response to bacterial proteins.

INTERLEUKIN-1/TOLL-LIKE RECEPTOR FAMILY

Ligand and Receptor Structure

The IL-1/Toll-like family of receptors comprises 11 members. IL-1 family members are subdivided into three subfamilies based on shared receptor or co-receptor binding. The IL-1 subfamily utilizes co-receptor binding to IL-1R3 and includes IL-1α, IL-1β, IL-33, and IL-1 receptor antagonist (IL-1Ra). The IL-18 subfamily consists of IL-18 and IL-37, both of which bind IL-1R5 (also known as IL-18Rα). The IL-36 subfamily includes IL-36α, IL-36β, IL-36γ, IL-36 receptor antagonist (IL-36Ra), and IL-38, which bind to IL-1R6 (IL-36R). All of these cytokines are produced as pro-peptides and are cleaved to generate the molecule, which binds to the appropriate receptor complex. This excludes IL-1Ra, which possesses a signal peptide and is secreted in a more classical fashion.

Family Members and Their Actions

Interleukin-1

There are two cell surface receptors for IL-1—type I (IL-1R1) and type II (IL-1R2). Both of these bind ligands (Fig. 14.5), but only IL-1R1 transduces signals. Upon ligand binding, IL-1R1 associates with IL-1R accessory protein (IL-1RAcP), which is critical for the initiation of signaling. The IL-1R2 cytoplasmic domain is extremely short and has been suggested to be a "decoy" receptor, competing with IL-1-RI for ligand binding, thus attenuating signaling. Both IL-1Rs are susceptible to proteolytic cleavage near the membrane surface. Thus, they can be found as soluble proteins that can "buffer" IL-1 signaling. These soluble receptors are readily detectable in circulation. IL-1R also associates with a second subunit, IL-1R associated protein (IL-1Rap).

Both IL-1α and IL-1β are synthesized as precursor proteins. IL-1α and IL-1β are structurally similar and have similar actions, but they are regulated differently. IL-1α is processed by a calpain-like converting enzyme but can also be cleaved by granzyme B, neutrophil proteases, and mast cells proteases. The pro-form of IL-1α has biological activity, but cleavage results in increased biological activity.

IL-1β is regulated at the level of mRNA stabilization and translation and it requires proteolytic activation. Pro-IL-1β remains in the cytoplasm until it is cleaved by proteases, such as caspase-1, and then transported out of the cell.

The cleavage of IL-1β occurs in a multi-protein complex called the inflammasome (Chapter 3). The key components of the inflammasome are caspase-1 and a recognition/assembly component NOD-like receptor (NLR). The adapter protein ASC, which contains pyrin and CARD domains, is required to facilitate inflammasome assembly. NLR proteins are intracellular pattern-recognition receptors that contain three domains: a segment with multiple leucine-rich repeats that recognizes the trigger for activation (it remains unclear whether this occurs directly or indirectly), a NACHT domain that leads to ATP-dependent dimerization of the NLR after trigger recognition,

and a protein-protein interaction domain—most commonly a pyrin or a CARD domain—that recruits caspase-1.

The best-studied inflammasome is the NLRP3 inflammasome. Trigger recognition by NLRP3 leads to its dimerization, the recruitment of ASC via interaction of the pyrin domains of NLRP3 and ASC, and the subsequent recruitment of caspase-1 via the CARD domains present in both ASC and caspase-1. Dimerization of caspase-1 upon inflammasome assembly allows for auto-activation by cleavage of its pro-form to generate the active enzyme. The NLRP3 inflammasome can be triggered by ATP, gram-positive bacterial cell wall components, intracytoplasmic DNA, molecules resulting from tissue damage, uric acid crystals, alum, silica, asbestos, and amyloid-β, among others. Cigarette smoke and cholesterol crystals also activate caspase-1.[43] IL-1β secretion can also occur via gasdermin N channels and caspase-8 and caspase-1 cleavage of the precursor gasdermin D. Secretion of IL-1β via gasdermin channels results in cell death via a mechanism known as pyroptosis (Chapter 17).[44]

Similar to IL-1α, IL-β can be cleaved and activated by other proteases, including neutrophil elastase, cathepsin, and proteinase-3 (PR-3). Mast cell proteases, as well as caspase-8, have also been shown to cleave the IL-1β precursor. The mechanism of release of IL-1 and related family members from cells is somewhat mysterious because they lack a classic signal peptide and don't enter the secretory pathway. Although there is some evidence that IL-1β can be cleaved and released from cells without cell death, a large body of evidence suggests that IL-1 family proteins are released from cells upon lysis. Because of the large numbers of cellular insults that result in release of IL-1 and the related proteins that trigger inflammation, it has been proposed that IL-1 family proteins function as "alarmins," acting as sentinels of cellular damage.

Principal functions of IL-1 include the induction of acute-phase protein synthesis, cachexia, and fever. In fact, it was the first endogenous pyrogen to be identified. IL-1 induces the production of IL-6 and chemokines, promotes hematopoiesis, stimulates adhesion of vascular leukocytes to endothelium, and has pro-coagulant effects. Importantly, IL-1 is a critical differentiation factor for Th17 cells, which underscores the significant role of this cytokine in inflammation and inflammatory diseases. IL-1 can activate ILC2s, inducing proliferation and cytokine expression.[45] Unlike Fas and some of the other TNF-family cytokines, IL-1 does not directly induce cell death. Mononuclear phagocytes are the main, but not exclusive, source of IL-1. *IL-1RI*$^{-/-}$ and *IL-1*$\beta^{-/-}$ mice have blunted fever responses to some—but not all—stimuli. Thus, despite the impressive actions of IL-1, it is somewhat redundant in febrile responses.

Interleukin-18

IL-18R is expressed predominantly on T cells, B cells, and NK cells associated with IL-18RacP. IL-18 is primarily produced by DC and macrophages and induces IFN-γ and activates NK cells, acting synergistically with IL-12. These functions are important for its anti-tumor activity, but IL-18 can also promote angiogenesis and tumor progression.[46] IL-18 can induce IL-4 and IL-13 production, indicating a somewhat broader range of action. The IL-18 precursor can be cleaved by the NLR family pyrin domain containing 3 (NRLP3) inflammasome as well as the NLR family CARD domain-containing protein 4 (NRLC4). IL-18-binding protein (IL-18bp) interacts with IL-18 to act as a decoy receptor that attenuates the actions of IL-18.

Interleukin-33

IL-33 (IL-1 F11) binds to the IL-1 receptor-related protein ST2, also designated IL-33R. IL-33 additionally acts as a transcriptional repressor by associating with chromatin. Because of this dual effect, and because of the expression of ST2 on different cell types, IL-33 acts on both immune and non-immune cells. It acts on T and B cells to promote Th2-associated cytokines, including IL-4, IL-5, and IL-13.[47] IL-33 enhances cell survival and promotes degranulation of mast cells, basophils, and granulocytes. Upon cleavage by mast cell proteases, IL-33 can also activate ILC2-inducing cytokine production and eosinophil recruitment.[48,49] IL-33-treated basophils have been shown to suppress arthritic inflammation. IL-33 correlates with RA disease severity, although it is unclear whether IL-33 substantially contributes to RA pathogenesis. IL-33 can act also on endothelial and epithelial cells to induce angiogenesis and production of other cytokines and chemokines.

Interleukin-36

The three members of the IL-36 subfamily—IL-36α, β, and γ—are encoded by separate genes, but all bind to a receptor composed of IL-36R and IL-1RacP. These cytokines are produced by both innate and adaptive immune cells. In turn, IL36s then induce the secretion of other pro-inflammatory cytokines. IL-36-producing cells are highly expressed in the skin and airways and are involved in diseases such as psoriasis and aspergillosis. Anti-IL-36R mAbs are being tested in pustular psoriasis (Chapter 64).

IL-36 receptor antagonist (IL-36Ra, also known as IL-1 F5) antagonizes the three IL-36 family cytokines. LOF *IL-36RA* mutations (deficiency of IL-36 receptor antagonist or DITRA) are associated with generalized pustular psoriasis. Keratinocytes from patients with deficiency of IL-36 receptor antagonist (DITRA) have elevated levels of multiple inflammatory cytokines in lesioned skin.

Interleukin-37

IL-37 (also known as IL-1 F7) broadly and negatively regulates excessive inflammatory responses by suppressing the innate and acquired immune response. IL-37 exists in several splice variants, of which IL-1F 7b has been the most studied. IL-37 binds to IL-18Rα chain, although with a lower affinity than IL-18. IL-37 uses TIR8 (also known as SIGIRR) as a second chain of the receptor. Although IL-18 and IL-37 bind to the same receptor and have a similar capacity to complex with IL-18Acp, IL-37 does not seem to act as a receptor antagonist for IL-18. IL-37 translocates to the nucleus and binds Smad3 (small mothers against dodecaplegia 3), enabling regulation of gene transcription. Cleavage of IL-37 appears to be dependent on caspase-1 and -4. Its anti-inflammatory activity is dependent on mTOR suppression and increased phosphorylation of AMP kinase, thus limiting glycolysis and ATP production. IL-37 tolerizes the DC and macrophages that express IL-37 such that they no longer secrete pro-inflammatory cytokines. IL-37 transgenic mice are resistant to LPS-induced shock.[50]

Interleukin-38

IL-38 (IL-1 F10) has homology with IL-36RA and, similarly, has anti-inflammatory activity. It is released by apoptotic cells to limit macrophage activation and it inhibits IL-17 and IL-22 production. *IL-1F10* polymorphisms are associated with RA, spondyloarthritis and psoriatic arthritis.

Other Members of the Interleukin-1 Family

The remaining members of the IL-1 and IL-1R families are the receptor homologues IL1RAPL2 and TIGIRR. These receptors have limited tissue distribution. TIGIRR is found almost exclusively in the brain, and IL1RAPL2 is found in the brain and a small number of other tissues.[51] Relatively little is known about the function of either TIGIRR or IL1RAPL2; although *IL1RAPL2* mutations cause mental retardation.

Signaling

Ligand binding to IL-1R, IL-18R, and TLRs results in NF-κB activation (see Fig. 14.5). These receptors all associate with the adapter protein MyD88 (Chapter 3). MyD88 has a C-terminal TIR domain and an N-terminal death domain. MyD88 allows the recruitment of IL-1 receptor-associated kinase (IRAK), which also has an N-terminal death domain. In turn, IRAK permits the recruitment and activation of a member of the TNF receptor-associated factor (TRAF) family, TRAF6. This leads to the activation of the serine kinases TAB2, TAK1, and inhibitors of κB kinases, IKKα, and IKKβ. With IKKγ or NEMO, these kinases phosphorylate IκB. The phosphorylated IκB is degraded within proteasomes, thus freeing bound NF-κB for nuclear translocation. Mice deficient in MyD88, IRAK, or TRAF6 have diminished responses to IL-1R/TLR family ligands. Other adapter molecules, including Mal and TRIF, are involved in TLR signaling.

CLINICAL RELEVANCE

Diseases Associated With the Interleukin-1 Family of Cytokines

- Mutations in genes involved in the formation of the inflammasome complexes result in increased secretion of IL-1 and other IL-1-family members.
- Mutations in the gene coding for IL-1 receptor antagonist cause a systemic autoinflammatory disease.

Clinical Relevance

The actions of IL-1 are tempered by the actions of a critical natural cytokine antagonist, IL-1 receptor antagonist (IL-1Ra), which is encoded by the *IL1RN* gene. Mutations in *IL1RN* can cause a systemic autoinflammatory disease, which is denoted Deficiency of IL-1Ra or DIRA.[52] Mutations in *NLRP3* contribute to several hereditary periodic fever syndromes (Chapter 37), including familial cold autoinflammatory syndrome (FCAS), Muckle-Wells syndrome, and neonatal onset multi-organ inflammatory disease (NOMID; in Europe called chronic infantile, neurological, cutaneous, articular syndrome or CINCA). NLRP3 is also called cryopyrin, thus, these disorders have been collectively referred to as cryopyrinopathies.[53] DITRA and other autoinflammatory diseases including NOMID are treated with recombinant IL1Ra, anakinra, and rilonacept, a fusion protein comprising the extracellular domains of IL-1Ra and IL-1RacP linked to the Fc portion of IgG1.

IL-1α, IL-1β, and other members of the IL-1 family are present in the synovial fluid of patients with osteoarthritis. Patients with RA treated with recombinant, human IL-1Ra have reduced joint space narrowing. The anti-IL-1β mAb, canakinumab, is used in the treatment of systemic juvenile idiopathic arthritis. Canakinumab has also been shown to reduce rates of recurrent major cardiovascular events, mortality, hospitalization, and new onset diabetes. The incidence of gout, osteoarthritis, and cancer also appear to be reduced, confirming a role for IL-1β in these pathologies.[54]

Anti-IL-33 mAbs (etokimab, SAR440340) are being tested in chronic rhinosinusitis, atopic dermatitis, asthma, and chronic obstructive pulmonary disease.

Mutations in the alternative inflammasome component, NRLC4, result in autoinflammation due to excessive IL-18 production and, ultimately, macrophage activation syndrome.[55] IL-18bp circulates at serum levels typically 20 to 30 times greater than those of IL-18; thus, IL-18bp is in sufficient excess to block any IL-18 activity. The amount of free IL-18 in the serum correlates with disease severity in several pathologies, including sepsis, SLE, Crohn disease, adult-onset Still disease, and macrophage activation syndrome. Accordingly, neutralization of IL-18 with rIL-18BP can result in the resolution of the hyper-inflammation.

High concentrations of circulating IL-37 have been reported in patients with RA, systemic juvenile idiopathic arthritis, and Still disease. Increased IL-37 and T-cell activation appear to correlate in patients with RA.

IL-38 has been detected in RA synovium and correlates with the expression of the other members of the IL-36 family. Serum concentrations of IL-38 correlate with reduced disease activity in SLE.

INTERLEUKIN-17 RECEPTORS

IL-17 and related cytokines are major inducers of inflammation, serve to recruit inflammatory cells, and provide protection against extracellular fungal and bacterial pathogenic species.

Ligand and Receptor Structure

The IL-17 receptor family comprises five, ubiquitously expressed receptors: IL-17AR, IL-17BR (IL-17RH1), IL-17RL (receptor like), IL-17RD, and IL-17RE.[56] These receptors have a single transmembrane domain and exceptionally large cytoplasmic tails. The ligands in the IL-17 family include IL-17A-F. Structurally, the IL-17 family members contain cystine knots and are structurally related to the nerve growth factor and platelet-derived growth factor.

IL-17A, commonly known as *IL-17*, is the founding member of the *IL-17* family. It is located on human chromosome 6 (mouse chromosome 1)[57] adjacent to the *IL17F* gene. IL-17A and IL-17F are produced by activated CD4, γ/δ, NKT, CD8 T cells, ILC3, and lymphoid tissue-inducer (LTI)-like cells and induced by IL-23 (Chapter 14). Beyond IL-23, Th17 cells are generated by IL-1β, IL-6, IL-21, and TGF-β.

Both IL-17A and IL-17F evoke inflammation largely by inducing the production of chemokines, G-CSF, and GM-CSF, leading to the subsequent recruitment of polymorphonuclear leukocytes. IL-17 also induces production of matrix metalloproteinase by epithelial cells, which may be an important aspect of its pro-inflammatory effects. IL-17 family cytokines are important in host defense against fungi and extracellular bacteria. Abundant data also point to the potential pathogenic roles of IL-17A in models of immune-mediated disease and in human autoimmune disorders.

Less well studied are IL-17B, IL-17C, and IL-17D, which are expressed in a variety of non-hematopoietic tissues, although IL-17D is reportedly produced by CD4 T cells. All of these cytokines are involved in anti-microbial immunity and barrier maintenance, as well as in inflammation promotion and neutrophil recruitment.

IL-17E, also known as IL-25, is constitutively produced by a population of intestinal epithelial cells, termed tuft cells.[58] IL-25 induces the expression of the type 2 cytokines (e.g., IL-4, IL-5, IL-13) in Th2 and ILC2, thereby enhancing mucus production,

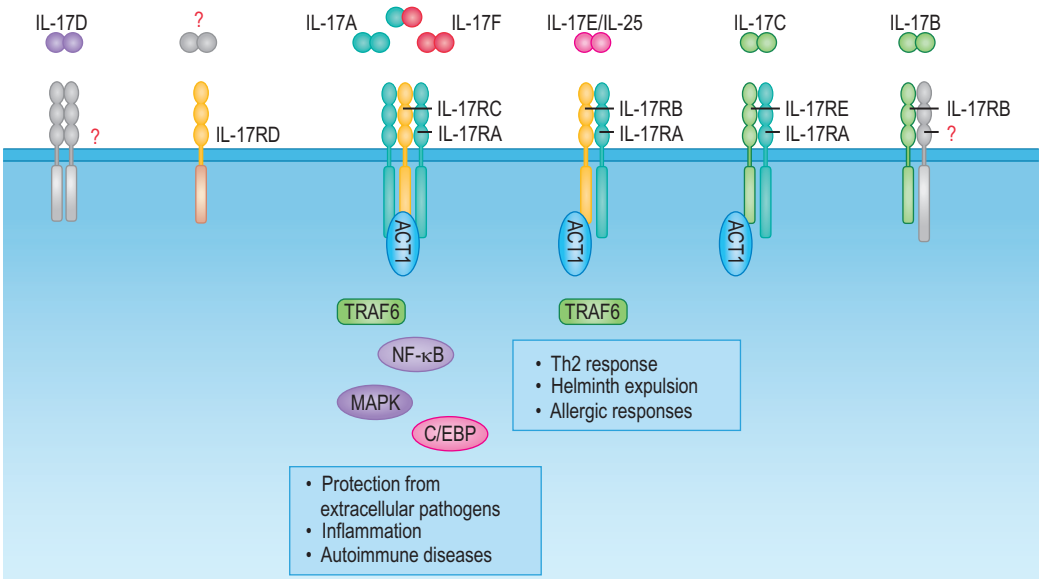


FIG. 14.6 Signal Transduction Cascade Downstream of Interleukin (*IL*)-17 Family of Cytokines and Their Receptors. IL-17 receptor family members recruit the adapter molecule ACT1 to couple to downstream signaling by nuclear factor kappa B (*NF-κB*), MAP kinases, and the transcription factor C/EBP.

epithelial cell hyperplasia, and eosinophilia. IL-25 is essential for the elimination of helminthic parasites.[59]

Signaling

Engagement of the IL-17 receptor activates MAP kinases, the PI3 kinase pathway, and NF-κB via Act-1 and TRAF6 (Fig. 14.6). IL-17 acts synergistically with TNF.[60] The IL-17R associates with an adapter molecule called Act through the SEFIR domains, which are part of both the receptor and the adaptor.[61]

CLINICAL RELEVANCE

Therapeutic Application of Interleukin-17 Blockade

- mAbs neutralizing IL-17 or antagonizing the IL-17 receptor are effective in treating psoriasis and psoriatic arthritis.

Clinical Relevance

Many human diseases and animal models of autoimmune disease have been associated with increased levels of IL-17. Anti-IL-17 and anti-IL-17 receptor mAbs are approved for the treatment of psoriasis, psoriatic arthritis, and ankylosing spondylitis and could represent a useful strategy to limit malignant transformation.[62] IL-17A and IL-17F are essential for mucocutaneous immunity against *Candida albicans*, and mutations of *IL17F*, *IL17RA*, or *IL17RC* result in chronic mucocutaneous candidiasis.[63]

CYTOKINES ACTIVATING RECEPTOR TYROSINE KINASES

Ligand and Receptor Structure

Many growth factors, such as insulin and epidermal growth factor, utilize receptor tyrosine kinases (RTKs). Some—but not all—of these factors can be classified as cytokines, including CSF-1 (colony-stimulating factor-1 also known as monocyte–macrophage-CSF or macrophage-CSF (M-CSF), stem cell factor (SCF, c-KIT ligand, or steel factor), platelet-derived growth factor (PDGF), and FLT3 ligand (FMS-like tyrosine kinase 3 ligand, FLT3-L). All of these have important hematological effects and tend to be included in discussions of cytokines. The structures of SCF and CSF-1 are similar to that of the cytokines that bind type I receptors as they too form four α-helical bundles, even though their receptors are entirely distinct. The similarities in their three-dimensional structures point to a common evolutionary ancestor. The receptors in this subfamily typically have five immunoglobulin-like loops in their ligand-binding extracellular domains. The cytoplasmic domain contains a tyrosine kinase catalytic domain interrupted by an "insert region" that does not share homology with other tyrosine kinases. This region is used to recruit various signaling molecules.

Family Members and Their Actions

KIT and Stem Cell Factor

Bone marrow stromal cells synthesize SCF as a soluble form (sSCF) and membrane-bound form (mSCF) resulting from alternative splicing and proteolytic cleavage. SCF is widely expressed during embryogenesis. Detectable in the circulation of normal adults, it is required to make stem cells responsive to other CSFs. sSCF signaling via c-KIT results in proliferation, differentiation, survival, and migration of stem cells; whereas mSCF prevents endocytosis of the Kit receptor. SCF has effects on germ cells, melanocytes, and hematopoietic precursors, as well as important effects on the mast cell differentiation. Naturally occurring mouse mutations of SCF (Steel) or its receptor (W) result in defects in hematopoiesis and fertility, loss of mast cells, and absent coat pigmentation.

Colony Stimulating Factor Receptor 1 (CSF-1R) and CSF1

CSF-1 is a hematopoietic growth factor that supports the survival and differentiation of monocytic cells. It is produced by a wide variety of cells, including monocytes, smooth muscle cells, endothelial cells,

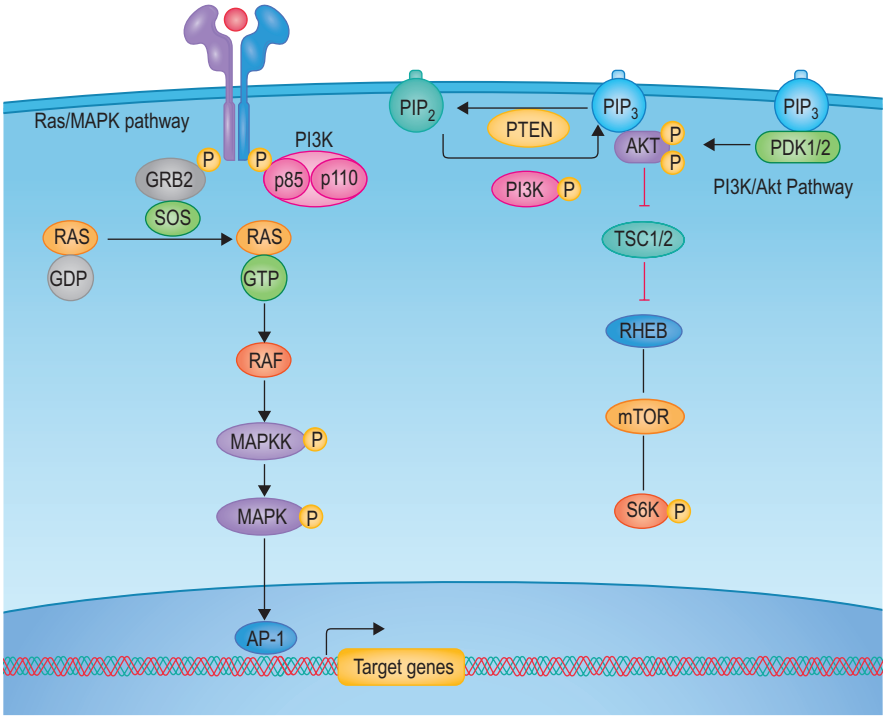


FIG. 14.7 Mechanism of Signal Transduction by Receptor Tyrosine Kinases. Ligand binding to receptor tyrosine kinases activates their enzymatic function leading to autophosphorylation, forming docking sites for adapter proteins that couple to activation of the RAS-Raf mitogen activated protein *(MAP)* kinase pathway. RTKs also activate phosphoinositide 3′ kinase (PI3K) leading to activation mTOR (mammalian target of rapamycin).

and fibroblasts. In humans, overexpression of CSF-1 leads to the recruitment of colony-stimulating factor-1 receptor (CSF-1R) expressing macrophages to the primary cell type within tenosynovial giant cell tumors (TGCTs). TGCT is a rare benign tumor that involves the synovium, bursa, and tendon sheath and results in the reduced mobility of the affected joint or limb. Kinase inhibitors that target CSF-1/CSF-1R have emerged as potential systemic agents for the treatment of TGCT.[64]

FLT and FLT3-L

FLT3-L synergizes with other cytokines, including SCF, to induce proliferation of hematopoietic precursors and is an important regulator of DC.

Signaling

The first step in signaling by the RTKs is ligand-induced receptor dimerization (Fig. 14.7). Dimerization brings the two kinase domains into proximity and results in the activation of phosphotransferase activity. This leads to autophosphorylation of the receptor subunits on the tyrosine residues, which are then bound by a variety of signaling molecules to initiate signal transduction. During this step, the signaling and adapter molecules recognize phosphotyrosine residues on the RTKs through either their SH2 (src homology 2) or phosphotyrosine binding (PTB) domains.

RTKs activate the RAS/RAF/ERK pathway. In some cases, GRB2 binds directly to phosphotyrosine residues on the cytoplasmic tail of the receptor via its SH2 domain. Alternatively, SHC can bind first and then recruit GRB2. In addition to an SH2 domain, GRB2 has two SH3 domains that bind proline-rich segments of the guanine nucleotide exchange factor, son of sevenless protein (SOS), recruiting it to the membrane and allowing it to activate RAS, a small G-protein. Activated RAS binds and activates the serine/threonine kinase RAF, which in turn phosphorylates the dual-specificity kinase MEK. Activated MEK phosphorylates and activates ERK (extracellular signal-regulated kinase), which then translocates to the nucleus where it phosphorylates and modulates the activity of various transcription factors, including ELK-1. GOF *RAS* mutations of RAS are common in a wide variety of human cancers.

RTKs also act through the phosphatidylinositol 3′-OH kinase (PI3-kinase) pathway.[65] PI3-kinase catalyzes the formation of phosphatidylinositol-3,4,5-trisphosphate (PtdIns(3,4,5)P3), and PtdIns(3,4)P2. These phospholipids are recognized by proteins with pleckstrin homology (PH) domains. One such protein is protein kinase B (PKB or AKT), which has been implicated in the regulation of apoptosis. The PI3-kinase pathway is inhibited by the lipid phosphatase PTEN, which dephosphorylates PI3-kinase-generated phosphatidylinositides. Deletion of PTEN has been found in numerous tumor types, demonstrating its role as a tumor suppressor.

Clinical Relevance

GOF c-KIT mutations result in systemic mastocytosis (Chapter 44). Mutations resulting in a fusion between the *PDGFRA* and *FIP1L1* genes underlie hypereosinophilic syndrome.[66,67]

KEY CONCEPT

Properties of the Transforming Growth Factor-β Receptor Family

- TGF-β receptors play a key role in lymphoid homeostasis, with pro- and anti-inflammatory actions.
- TGF-β promotes the differentiation of regulatory T cells and Th17 cells.
- TGF-β receptors transduce signals through SMAD proteins.
- TGF-β receptor function is dysregulated in many forms of human cancer.

TRANSFORMING GROWTH FACTOR-β LIGAND AND RECEPTOR FAMILIES

The transforming growth factor-βs (TGF-β) are a family of over 40 cytokines involved in processes, including tissue remodeling, wound repair, development, and hematopoiesis. Mutations of the elements in this pathway also contribute to malignant transformation and connective tissue and bone disorders. The mammalian ligands that belong to this family include TGF-β1, -β2, and -β3, bone morphogenic proteins (BMPs), growth and differentiation factors (GDFs), activins, inhibins, nodal, leftys, and Müllerian-inhibiting substance. TGF-βs induce collagen and fibronectin production by fibroblasts, which are thought to be responsible—at least in part—for diseases characterized by fibrosis (e.g., systemic sclerosis, pulmonary fibrosis). Functionally, TGF-β inhibits many aspects of lymphocyte function, including T-cell proliferation and CTL maturation. In mice, disruption of TGF-β receptor signaling results in massive expansion of lymphoid organs and development of T-cell lymphoproliferative disorders, indicative of a critical role for TGF-β in T-cell homeostasis.[68]

Ligand and Receptor Structure

The TGF-βs are expressed as biologically inactive disulphide-linked dimers that are cleaved to form active dimers. On translocation into the endoplasmic reticulum, the N-terminal leader peptide is cleaved, and the mature protein is subsequently generated by a second cleavage event that releases an N-terminal pro-region. The pro-region can remain associated with the biologically active C-terminal region, inhibiting its activity.

The biological effects of the TGF-βs and their related ligands are mediated by two classes of receptor, designated type I (RI) and type II (RII).[69] A third group of receptors, denoted type III, also exists (i.e., TGF-βRIII for TGF-β); however, the latter group function as co-receptors.

Similar to the RTKs, the cytoplasmic domains of TGF-β receptors possess intrinsic kinase activity; however, TGF-βRI and TGF-βRII encode serine/threonine kinases. The signaling cascade is initiated by the binding of TGF-β to the type II receptor, inducing the assembly of a ternary complex containing TGF-β, TGF-βRII, and TGF-βRI. The principal type I receptor in the TGF-β pathway is the ~55-kDa activin-like kinase-5 (ALK-5). ALK-1 can also be recruited into the complex and can transduce TGF-β-mediated signals.

TGF-β Family Members and Their Actions

The three known human TGF-βs (TGF-β1, TGF-β2, and TGF-β3) are closely related and have very similar biological functions. TGF-β1, the most abundant form, is the only isoform found in platelets. T cells and monocytes mainly synthesize TGF-β1, a critical function of which is to antagonize lymphocyte responses.

Approximately half of TGF-β1$^{-/-}$ mice survive until birth, but then succumb 3 to 4 weeks later to an overwhelming autoimmune state characterized by lymphoid and mononuclear infiltration of the heart, lung, and other tissues and by autoantibody production. These studies, together with selective inhibition of TGF-β function in T cells, indicate that TGF-β plays a crucial role in T-cell homeostasis and the prevention of spontaneous T-cell differentiation. TGF-β1 induces FoxP3, promotes adaptive Treg cell differentiation, and inhibits IFN-γ production. TGF-β can also regulate T-cell tolerance independently of FoxP3. Conversely, TGF-β1 with IL-6 induces IL-17. TGF-β also promotes IL-22 secretion by Th17 cell through Ahr activation and promoting tumors. TGF-β with IL-4 induces IL-9, a cytokine with pro- and anti-inflammatory actions. Thus, TGF-β1 has both pro-inflammatory and anti-inflammatory activities.

Of the three isoforms, TGF-β2 is the most abundant and TGF-β3 is the least abundant isoform in body fluids. TGF-β2 and TGF-β3-null mice exhibit defects distinct from those observed in TGF-β1 knockouts, particularly in bone and internal organ formation. Their deficiency is embryonically lethal. Thus, although the three isoforms function similarly in vitro, they play distinct roles in vivo.

Signaling

The human type II receptor is an 80-kDa glycoprotein and is the principal receptor for TGF-β. Once TGF-β binds to type II receptor, the type I receptor is recruited into the complex and activated through phosphorylation of its GS domain, a juxtamembrane domain preceding the kinase domain (Fig. 14.8).[69] In turn, the type I receptor is responsible for phosphorylating key signaling intermediates, SMADs, the primary substrates of the receptor.

Small mothers against dodecaplegias (SMADs) are a group of highly conserved proteins that play a critical role in TGF-β signal transduction.[70] Eight SMADs have been identified in mammals. All exhibit a high degree of specificity for conserved motifs in the cytoplasmic tail of type I receptors with no known structural or enzymatic motifs. They are subdivided into three classes based on functional distinctions.

The receptor-regulated SMADs (R-Smads) directly interact with the type I receptor. These include SMAD1, -2, -3, -5, and -8. Smad2 and Smad3 are phosphorylated in response to TGF-β, whereas Smad1, -5, and -8 are primarily activated in response to BMP activation. The interaction of R-SMADs with TGF-β

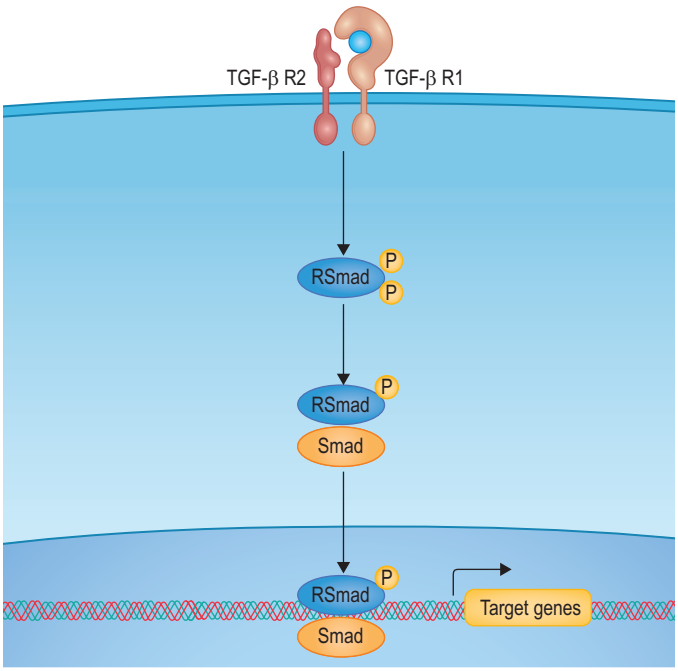


FIG. 14.8 Signaling byTransforming Growth factor-β (*TGF-β*) Family Receptor Serine Kinases. Ligand binding activates enzymatic activity leading to phosphorylation of SMADs, which translocate to the nucleus, bind DNA, and regulate gene expression.

receptors can also be regulated by another molecule termed SARA (SMAD anchor for receptor activation). SARA binds unphosphorylated SMAD2 and SMAD3. Through its lipid-binding domain (FYVE), SARA relocates SMADs to the plasma membrane, facilitating receptor binding.

Once phosphorylated, R-SMADs dissociate from SARA and from the activated type I receptor and associate with SMAD4 in the cytoplasm. This is followed by nuclear translocation of the heteromeric SMAD complex with binding to cognate DNA motifs in the promoters of TGF-β-responsive genes and to concomitant induction of transcription. SMAD2 null mice lack anterior/posterior specification and fail to develop mesoderm, causing embryonic lethality. SMAD3-deficient mice have limb malformations and defective immune function.

Upon recruitment to their cognate activated type I receptor, R-SMADs are phosphorylated on C-terminal serine residues, triggering homodimerization of R-SMADs or heterodimerization of the common SMAD—or C-SMAD. SMAD4 is the only known C-SMAD in vertebrates, functioning as the central and essential downstream mediator of other SMADs in all TGF-β/BMP pathways. SMAD4-deficient mice exhibit severe defects in gastrulation and die early in embryogenesis. SMAD4 specific deletion provided insight into its role in bone development and pathogenesis of chondrodysplasia[71]

SMAD6 and SMAD7 comprise the third subfamily of SMADs in mammals, the inhibitory SMADs (I-SMADs). SMAD6 competes with SMAD4 for binding to an R-SMAD. Studies with SMAD6-deficient mice indicate a role in the development and homeostasis of the cardiovascular system. SMAD7 is induced by TGF-β and binds to TGF-β receptors, inhibiting the phosphorylation of R-SMADs, thus serving as a classic feedback inhibitor.

Although SMADs bind DNA, they also interact with a variety of other transcription factors, transcriptional co-activators, and transcriptional co-repressors to coordinately regulate transcription of a select subset of complex promoters. For example, SMADs can bind ATF2—the vitamin D receptor—and other transcription factors and can recruit the co-activators CBP/p300. SMAD4 can also regulate FZD4, a component of the Wnt network, therefore establishing a bridge between TGF-β and Wnt.[72]

A number of members of the mitogen-activated protein kinase (MAPK) family are activated in response to TGF-β, including ERK, JNK, and p38 MAPK. Another MAPK member, TAK1 (TGF-β-associated kinase-1) can be recruited upon TGF-β stimulation, forming a complex with TAB1 (TAK1-asociated binding protein), leading to downstream activation of p38/MPK2 and c-JUN N-terminal kinase (JNK) through MKK6 and MKK4, respectively. TGF-β can also interact with the PI3K-AKT-mTOR signaling pathway. Thus, TGF-β signaling influences many cellular functions by its ability to crosstalk with other paths.

Clinical Relevance

Loeys-Dietz syndrome is a connective tissue disorder that results from mutations in *TGFBR1*, *TGFBR2*, *TGFB2*, *TGFB3*, *or SMAD3*. GOF mutations of activin receptor-like kinase 2 (ALK-2, encoded by *ACVR1*) underlie the disorder fibrodysplasia ossificans progressive, which is characterized by widespread heterotopic bone formation. SMAD4 is deleted in half of all human pancreatic carcinomas. Mutations in SMAD2 have been identified in patients with colon cancer, and somatic mutations in TGF-β receptors have been identified in colon and gastric cancers. Loss of SMAD3 is associated with leukemia. Oncogenic RAS can repress SMAD signaling by negatively regulating SMAD2 and SMAD3. Variants in SMAD6 are associated with complex cardiovascular pathology. Somatic mutations of *SMAD3* are associated with melorheostosis, a rare genetic bone disease.[73] Activation of the TGF-β network can contribute to pathological fibrosis in most organs and has been associated with multiple skeletal muscle myopathies, including Duchenne muscular dystrophy.

OTHER CYTOKINES

Interleukin-14

IL-14 was identified as a high molecular weight B-cell growth factor produced by T cells and some B-cell tumors. The precise nature of this putative cytokine is still uncertain, although it has been implicated in the pathogenesis of autoimmune diseases, SLE, Sjögren syndrome (Chapter 55), and Graves disease or Hashimoto thyroiditis (Chapter 70).[74]

Interleukin-16

IL-16 was formerly termed lymphocyte chemoattractant factor because of its ability to recruit CD4 T cells.[75] Unrelated to other cytokines, its only known receptor is CD4. Originally identified as a product of CD8 T cells, its message is widely expressed. CD4 T cells, eosinophils, and mast cells as well as epithelial cells, fibroblasts, and monocytes can all secrete IL-16. It is present in bronchoalveolar lavage fluids from asthmatics and sarcoid patients. IL-16 has also been detected in blister fluid from bullous pemphigoid lesions (Chapter 63) and overexpressed in RA, IBD, SLE, and autoimmune thyroid disease.

Interleukin-32

IL-32 is structurally distinct from other families.[76] It is inducible by the combination of IL-12 and IL-18, and nine different isoforms are generated by mRNA splicing. IL-32 induces the expression of various cytokines, including TNF, IL-1, IL-6, and chemokines. It can synergize with muramyl dipeptides and signals via NF-κB and p38. IL-32 is present in rheumatoid synovium and is found at increased serum concentrations in patients with Behçet disease, SLE, and atopic dermatitis (Chapter 48). IL-32 may be a promising approach to the evaluation of the presence of pulmonary arterial hypertension in systemic sclerosis.[77]

Interleukin-34

IL-34 is the second ligand for CSF-1R. Its functions include the differentiation, proliferation, and survival of monocytes, macrophages, and osteoclasts. IL-34 has been associated with the inflammatory process seen in RA, IBD, and Sjogren syndrome. IL-34 levels appear to be elevated in serum and synovial fluid in patients with RA.[78]

CONCLUSIONS AND SUMMARY

Cytokines encompass a wide range of molecules that are essential for communication between cells of the immune system and other non-immune cells. Although the number of cytokines already seems vast, it is likely that more will be discovered in the future. Considerable progress has been made in defining the in vivo functions of various cytokines. Equally impressive have been the advances in our understanding of how the dysregulation of cytokines and cytokine signaling contributes to human disease. While there is no doubt that innumerable advances,

including direct therapeutic triumphs (Table 14.4), have come from understanding the basic science of cytokine biology, there is still much to do. An exciting possibility for the future is engineering synthetic cytokines. Equally, there are many possible bispecific antibodies that target more than one cytokine or cytokine receptor that can be generated. Improved understanding of regulators like SOCS proteins should also offer new therapeutic opportunities. As imaging of cells and macromolecular complexes improve, it will be particularly exciting to really see how cytokines signal. The best advice is to "stay tuned to this station."

REFERENCES

1. Dumonde DC, Wolstencroft RA, Panayi GS, Matthew M, Morley J, Howson WT. "Lymphokines": non-antibody mediators of cellular immunity generated by lymphocyte activation. *Nature.* 1969;224(5214):38–42.
2. Revised nomenclature for antigen-nonspecific T cell proliferation and helper factors. J Immunol. 1979;123(6):2928-9.
3. Cohen S, Bigazzi PE, Yoshida T. Commentary. Similarities of T cell function in cell-mediated immunity and antibody production. *Cell Immunol.* 1974;12(1):150–159.
4. Vilcek J. The cytokines: an overview. In: LM Thompson AW, ed. *The cytokine handbook*. San Diego, CA: Academic Press; 2003:3–18.
5. Boulay JL, O'Shea JJ, Paul WE. Molecular phylogeny within type I cytokines and their cognate receptors. *Immunity.* 2003;19(2):159–163.
6. Murakami M, Kamimura D, Hirano T. Pleiotropy and Specificity: Insights from the Interleukin 6 Family of Cytokines. *Immunity.* 2019;50(4):812–831.
7. Garbers C, Rose-John S. Dissecting Interleukin-6 Classic- and Trans-Signaling in Inflammation and Cancer. *Methods Mol Biol.* 2018;1725:127–140.
8. Tait Wojno ED, Hunter CA, Stumhofer JS. The Immunobiology of the Interleukin-12 Family: Room for Discovery. *Immunity.* 2019;50(4):851–870.
9. Dougan M, Dranoff G, Dougan SK. GM-CSF, IL-3, and IL-5 Family of Cytokines: Regulators of Inflammation. *Immunity.* 2019;50(4):796–811.
10. Leonard WJ, Lin JX, O'Shea JJ. The gammac Family of Cytokines: Basic Biology to Therapeutic Ramifications. *Immunity.* 2019;50(4):832–850.
11. Abbas AK, Trotta E, RS D, Marson A, Bluestone JA. Revisiting IL-2: Biology and therapeutic prospects. *Sci Immunol*. 2018;3(25).
12. DiToro D, Winstead CJ, Pham D, Witte S, Andargachew R, Singer JR, et al. Differential IL-2 expression defines developmental fates of follicular versus nonfollicular helper T cells. *Science*. 2018;361:6407.
13. Ghelani A, Bates D, Conner K, Wu MZ, Lu J, Hu YL, et al. Defining the Threshold IL-2 Signal Required for Induction of Selective Treg Cell Responses Using Engineered IL-2 Muteins. *Front Immunol*. 2020;11:1106.
14. Gieseck 3rd RL, Wilson MS, Wynn TA. Type 2 immunity in tissue repair and fibrosis. *Nat Rev Immunol*. 2018;18(1):62–76.
15. O'Shea JJ, Paul WE. Mechanisms underlying lineage commitment and plasticity of helper CD4+ T cells. *Science*. 2010;327(5969):1098–1102.
16. Trinchieri G, Pflanz S, Kastelein RA. The IL-12 family of heterodimeric cytokines: new players in the regulation of T cell responses. *Immunity.* 2003;19(5):641–644.
17. Watford WT, Hissong BD, Bream JH, Kanno Y, Muul L, O'Shea JJ. Signaling by IL-12 and IL-23 and the immunoregulatory roles of STAT4. *Immunol Rev*. 2004;202:139–156.
18. Filipe-Santos O, Bustamante J, Chapgier A, Vogt G, de Beaucoudrey L, Feinberg J, et al. Inborn errors of IL-12/23- and IFN-gamma-mediated immunity: molecular, cellular, and clinical features. *Semin Immunol*. 2006;18(6):347–361.
19. Teng MW, Bowman EP, McElwee JJ, Smyth MJ, Casanova JL, Cooper AM, et al. IL-12 and IL-23 cytokines: from discovery to targeted therapies for immune-mediated inflammatory diseases. *Nat Med*. 2015;21(7):719–729.
20. Klose CSN, Mahlakoiv T, Moeller JB, Rankin LC, Flamar AL, Kabata H, et al. The neuropeptide neuromedin U stimulates innate lymphoid cells and type 2 inflammation. *Nature.* 2017;549(7671):282–286.
21. Nagashima H, Mahlakoiv T, Shih HY, Davis FP, Meylan F, Huang Y, et al. Neuropeptide CGRP Limits Group 2 Innate Lymphoid Cell Responses and Constrains Type 2 Inflammation. *Immunity.* 2019;51(4): 682-95 e6.
22. Silverberg JI, Pinter A, Pulka G, Poulin Y, Bouaziz JD, Wollenberg A, et al. Phase 2B randomized study of nemolizumab in adults with moderate-to-severe atopic dermatitis and severe pruritus. *J Allergy Clin Immunol*. 2020;145(1):173–182.
23. Corren J, Parnes JR, Wang L, Mo M, Roseti SL, Griffiths JM, et al. Tezepelumab in Adults with Uncontrolled Asthma. *N Engl J Med*. 2017;377(10):936–946.
24. Lazear HM, Schoggins JW, Diamond MS. Shared and Distinct Functions of Type I and Type III Interferons. *Immunity*. 2019;50(4):907–923.
25. Uggenti C, Lepelley A, Crow YJ. Self-Awareness: Nucleic Acid-Driven Inflammation and the Type I Interferonopathies. *Annu Rev Immunol*. 2019;37:247–267.
26. Morand EF, Furie R, Tanaka Y, Bruce IN, Askanase AD, Richez C, et al. Trial of Anifrolumab in Active Systemic Lupus Erythematosus. *N Engl J Med*. 2020;382(3):211–221.
27. Ouyang W, O'Garra A. IL-10 Family Cytokines IL-10 and IL-22: from Basic Science to Clinical Translation. *Immunity*. 2019;50(4):871–891.
28. Villarino AV, Gadina M, O'Shea JJ, Kanno Y. SnapShot: Jak-STAT Signaling II. *Cell.* 2020;181(7): 1696-e1.
29. Gadina M, Chisolm DA, Philips RL, McInness IB, Changelian PS, O'Shea JJ. Translating JAKs to Jakinibs. *J Immunol.* 2020;204(8):2011–2020.
30. Gruber C, Martin-Fernandez M, Ailal F, Qiu X, Taft J, Altman J, et al. Homozygous STAT2 gain-of-function mutation by loss of USP18 activity in a patient with type I interferonopathy. *J Exp Med*. 2020;217(5).
31. Zhang X, Bogunovic D, Payelle-Brogard B, Francois-Newton V, Speer SD, Yuan C, et al. Human intracellular ISG15 prevents interferon-alpha/beta over-amplification and auto-inflammation. *Nature*. 2015;517(7532):89–93.
32. Beutler BA. The role of tumor necrosis factor in health and disease. *J Rheumatol Suppl*. 1999;57:16–21.
33. Klebanoff CA, Scott CD, Leonardi AJ, Yamamoto TN, Cruz AC, Ouyang C, et al. Memory T cell-driven differentiation of naive cells impairs adoptive immunotherapy. *J Clin Invest*. 2016;126(1):318–334.
34. Park SM, Schickel R, Peter ME. Nonapoptotic functions of FADD-binding death receptors and their signaling molecules. *Curr Opin Cell Biol*. 2005;17(6):610–616.
35. Croft M, Siegel RM. Beyond TNF: TNF superfamily cytokines as targets for the treatment of rheumatic diseases. *Nat Rev Rheumatol*. 2017;13(4):217–233.
36. Siegel RM, Muppidi J, Roberts M, Porter M, Wu Z. Death receptor signaling and autoimmunity. *Immunol Res*. 2003;27(2-3):499–512.
37. Siegel RM. Caspases at the crossroads of immune-cell life and death. *Nat Rev Immunol*. 2006;6(4):308–317.
38. Kim EY, Priatel JJ, Teh SJ, Teh HS. TNF receptor type 2 (p75) functions as a costimulator for antigen-driven T cell responses in vivo. *J Immunol*. 2006;176(2):1026–1035.
39. Wajant H, Siegmund D. TNFR1 and TNFR2 in the Control of the Life and Death Balance of Macrophages. *Front Cell Dev Biol*. 2019;7:91.
40. Simon A, Park H, Maddipati R, Lobito AA, Bulua AC, Jackson AJ, et al. Concerted action of wild-type and mutant TNF receptors enhances inflammation in TNF receptor 1-associated periodic fever syndrome. *Proc Natl Acad Sci U S A*. 2010;107(21):9801–9806.
41. Bulua AC, Mogul DB, Aksentijevich I, Singh H, He DY, Muenz LR, et al. Efficacy of etanercept in the tumor necrosis factor receptor-associated periodic syndrome: a prospective, open-label, dose-escalation study. *Arthritis Rheum*. 2012;64(3):908–913.
42. Schneider H, Faschingbauer F, Schuepbach-Mallepell S, Korber I, Wohlfart S, Dick A, et al. Prenatal Correction of X-Linked Hypohidrotic Ectodermal Dysplasia. *N Engl J Med*. 2018;378(17):1604–1610.
43. Martinon F, Mayor A, Tschopp J. The inflammasomes: guardians of the body. *Annu Rev Immunol*. 2009;27:229–265.
44. Dinarello CA. The IL-1 family of cytokines and receptors in rheumatic diseases. *Nat Rev Rheumatol*. 2019;15(10):612–632.
45. Golebski K, Ros XR, Nagasawa M, van Tol S, Heesters BA, Aglmous H, et al. IL-1beta, IL-23, and TGF-beta drive plasticity of human ILC2s towards IL-17-producing ILCs in nasal inflammation. *Nat Commun*. 2019;10(1):2162.

46. Fabbi M, Carbotti G, Ferrini S. Context-dependent role of IL-18 in cancer biology and counter-regulation by IL-18BP. *J Leukoc Biol*. 2015;97(4):665–675.
47. Schmitz J, Owyang A, Oldham E, Song Y, Murphy E, McClanahan TK, et al. IL-33, an interleukin-1-like cytokine that signals via the IL-1 receptor-related protein ST2 and induces T helper type 2-associated cytokines. *Immunity.* 2005;23(5):479–490.
48. Halim TY, Krauss RH, Sun AC, Takei F. Lung natural helper cells are a critical source of Th2 cell-type cytokines in protease allergen-induced airway inflammation. *Immunity.* 2012;36(3):451–463.
49. Lefrancais E, Duval A, Mirey E, Roga S, Espinosa E, Cayrol C, et al. Central domain of IL-33 is cleaved by mast cell proteases for potent activation of group-2 innate lymphoid cells. *Proc Natl Acad Sci U S A*. 2014;111(43):15502–15507.
50. Sims J, Towne J, Blumberg H. 11 IL-1 family members in inflammatory skin disease. *Ernst Schering Res Found Workshop*. 2006;56:187–191.
51. Carrie A, Jun L, Bienvenu T, Vinet MC, McDonell N, Couvert P, et al. A new member of the IL-1 receptor family highly expressed in hippocampus and involved in X-linked mental retardation. *Nat Genet*. 1999;23(1):25–31.
52. Aksentijevich I, Masters SL, Ferguson PJ, Dancey P, Frenkel J, van Royen-Kerkhoff A, et al. An autoinflammatory disease with deficiency of the interleukin-1-receptor antagonist. *N Engl J Med*. 2009;360(23):2426–2437.
53. Neven B, Prieur AM. Quartier dit Maire P. Cryopyrinopathies: update on pathogenesis and treatment. *Nat Clin Pract Rheumatol*. 2008;4(9):481–489.
54. Ridker PM, Everett BM, Thuren T, MacFadyen JG, Chang WH, Ballantyne C, et al. Antiinflammatory Therapy with Canakinumab for Atherosclerotic Disease. *N Engl J Med*. 2017;377(12):1119–1131.
55. Canna SW, de Jesus AA, Gouni S, Brooks SR, Marrero B, Liu Y, et al. An activating NLRC4 inflammasome mutation causes autoinflammation with recurrent macrophage activation syndrome. *Nat Genet*. 2014;46(10):1140–1146.
56. Kawaguchi M, Adachi M, Oda N, Kokubu F, Huang SK. IL-17 cytokine family. *J Allergy Clin Immunol*. 2004;114(6):1265–1273. quiz 74.
57. McGeachy MJ, Cua DJ, Gaffen SL. The IL-17 Family of Cytokines in Health and Disease. *Immunity.* 2019;50(4):892–906.
58. O'Leary CE, Schneider C, Locksley RM. Tuft Cells-Systemically Dispersed Sensory Epithelia Integrating Immune and Neural Circuitry. *Annu Rev Immunol*. 2019;37:47–72.
59. Fort MM, Cheung J, Yen D, Li J, Zurawski SM, Lo S, et al. IL-25 induces IL-4, IL-5, and IL-13 and Th2-associated pathologies in vivo. *Immunity*. 2001;15(6):985–995.
60. Maezawa Y, Nakajima H, Suzuki K, Tamachi T, Ikeda K, Inoue J, et al. Involvement of TNF receptor-associated factor 6 in IL-25 receptor signaling. *J Immunol*. 2006;176(2):1013–1018.
61. Li X, Bechara R, Zhao J, McGeachy MJ, Gaffen SL. IL-17 receptor-based signaling and implications for disease. *Nat Immunol*. 2019;20(12):1594–1602.
62. Gaffen SL, Jain R, Garg AV, Cua DJ. The IL-23-IL-17 immune axis: from mechanisms to therapeutic testing. *Nat Rev Immunol*. 2014;14(9):585–600.
63. Puel A, Cypowyj S, Bustamante J, Wright JF, Liu L, Lim HK, et al. Chronic mucocutaneous candidiasis in humans with inborn errors of interleukin-17 immunity. *Science.* 2011;332(6025):65–68.
64. Benner B, Good L, Quiroga D, Schultz TE, Kassem M, Carson WE, et al. Pexidartinib, a Novel Small Molecule CSF-1R Inhibitor in Use for Tenosynovial Giant Cell Tumor: A Systematic Review of Pre-Clinical and Clinical Development. *Drug Des Devel Ther*. 2020;14:1693–1704.
65. Cantley LC. The phosphoinositide 3-kinase pathway. *Science.* 2002;296(5573):1655–1657.
66. Longley BJ, Reguera MJ, Ma Y. Classes of c-KIT activating mutations: proposed mechanisms of action and implications for disease classification and therapy. *Leuk Res*. 2001;25(7):571–576.
67. Cools J, DeAngelo DJ, Gotlib J, Stover EH, Legare RD, Cortes J, et al. A tyrosine kinase created by fusion of the PDGFRA and FIP1L1 genes as a therapeutic target of imatinib in idiopathic hypereosinophilic syndrome. *N Engl J Med*. 2003;348(13):1201–1214.
68. Ghoreschi K, Laurence A, Yang XP, Hirahara K, O'Shea JJ. T helper 17 cell heterogeneity and pathogenicity in autoimmune disease. *Trends Immunol*. 2011;32(9):395–401.
69. Heldin CH, Miyazono K, ten Dijke P. TGF-beta signalling from cell membrane to nucleus through SMAD proteins. *Nature.* 1997;390(6659):465–471.
70. Massague J, Gomis RR. The logic of TGFbeta signaling. *FEBS Lett*. 2006;580(12):2811–2820.
71. Yan J, Li J, Hu J, Zhang L, Wei C, Sultana N, et al. Smad4 deficiency impairs chondrocyte hypertrophy via the Runx2 transcription factor in mouse skeletal development. *J Biol Chem*. 2018;293(24):9162–9175.
72. Du X, Li Q, Yang L, Liu L, Cao Q, Li Q. SMAD4 activates Wnt signaling pathway to inhibit granulosa cell apoptosis. *Cell Death Dis*. 2020;11(5):373.
73. Kang H, Jha S, Ivovic A, Fratzl-Zelman N, Deng Z, Mitra A, et al. Somatic SMAD3-activating mutations cause melorheostosis by up-regulating the TGF-beta/SMAD pathway. *J Exp Med*. 2020;217(5).
74. Kemp EH, Ajjan RA, Metcalfe RA, Watson PF, Weetman AP. IL-14 and IL-16 are expressed in the thyroid of patients with either Graves' disease or Hashimoto's thyroiditis. *Clin Endocrinol (Oxf)*. 2015;83(5):726–732.
75. Wilson KC, Center DM, Cruikshank WW. The effect of interleukin-16 and its precursor on T lymphocyte activation and growth. *Growth Factors*. 2004;22(2):97–104.
76. Heinhuis B, Netea MG, van den Berg WB, Dinarello CA, Joosten LA. Interleukin-32: a predominantly intracellular proinflammatory mediator that controls cell activation and cell death. *Cytokine.* 2012;60(2):321–327.
77. Di Benedetto P, Guggino G, Manzi G, Ruscitti P, Berardicurti O, Panzera N, et al. Interleukin-32 in systemic sclerosis, a potential new biomarker for pulmonary arterial hypertension. *Arthritis Res Ther*. 2020;22(1):127.
78. Lin W, Xu D, Austin CD, Caplazi P, Senger K, Sun Y, et al. Function of CSF1 and IL34 in Macrophage Homeostasis, Inflammation, and Cancer. *Front Immunol*. 2019;10:2019.

15

Chemokines and Chemokine Receptors

Philip M. Murphy

The large number of leukocyte subsets that populate the immune system, each with distinct and complex trafficking itineraries, places great demands on an operating system. Orderly trafficking (Chapter 16) depends on the specificity of chemokine receptors for chemokines and leukocytes, dynamic temporospatial regulation of chemokine expression in source cells and chemokine receptor expression in responding cells, integration of multiple signal transduction pathways, and signal-modifying factors in individual cells (Fig. 15.1).[1] Accordingly, the chemokines are the largest family of cytokines (Chapter 14), and chemokine receptors are one of the largest families of 7-transmembrane-domain receptors. In order to fine-tune immune responses, combinatorial complexity permits multitasking and back-up by individual system components in a single cell type or among different cell types.

CHEMOKINES

Chemokines are defined by their structure, not their function.[2] Conserved and uniformly spaced disulfide-bonded cysteines help create a highly conserved tertiary structure (Figs. 15.1 and 15.2A). Most chemokines have at least four conserved cysteines. The first two are either adjacent (CC motif, $n = 24$ in human) or separated by one (CXC motif, $n = 16$ in human) or three (CX3C motif, $n = 1$) amino acids. Two chemokines have only two cysteines (XC motif), which correspond to C-2 and C-4 in the other groups. Disulfide bonds link C-1 to C-3 and C-2 to C-4. The chemokine core domain contains 3 β-sheets connected by short loops and arranged in the shape of a Greek key. The core is overlaid by a C-terminal α-helical domain, and flanked by an N-terminal domain that lacks order. The four chemokine cysteine motifs are followed by the letter "L" for *ligand* and a number (e.g., CXCL1) in a systematic nomenclature.[3]

Chemokines may form multiple quaternary structures, including monomers, homodimers, homotetramers, and heterodimers. Complex multimers noncovalently bound to glycosaminoglycans (GAGs) on the endothelial cell surfaces may be important for presentation to leukocytes in vivo. However, structural studies indicate that chemokines probably bind to receptors as monomers, potentially released as a cloud from multimerized deposits at GAG binding sites.[4] CXCL16 and CX3CL1 are exceptional in having a multimodular structure that includes a classic chemokine domain, a mucin-like stalk, a transmembrane domain, and a C-terminal cytoplasmic module. Each can exist as either a membrane-bound form or shed form, enabling either direct cell–cell adhesion or chemotaxis, respectively. CXCL16 also multitasks as a scavenger receptor, binding to phosphatidylserine and oxidized low-density lipoprotein.

KEY CONCEPTS

Chemokine and Chemokine Receptors

Definition
- Chemokines are defined by a common structure, the chemokine fold.
- Chemokine receptors are defined by a common biochemical function: chemokine binding-dependent cell signaling.
  - Most chemokine receptors catalyze guanine nucleotide-exchange on Gi-type G proteins.
  - A small group of so-called atypical receptors do not signal through G proteins; some signal through an arrestin-dependent pathway.

Classification
- Chemokines form four main structural subclasses (C, CC, CXC, and CX3C) and two main immunological subclasses (inflammatory and homeostatic).

Evolution
- Chemokines and chemokine receptors arose in vertebrates and have been copied or mimicked by many viruses.
- Chemokines and chemokine receptor repertoires can differ among species and among individuals of the same species.

Ligand–Receptor Promiscuity
- The majority of chemokine receptors pair promiscuously with chemokine ligands, usually restricted to a single chemokine subclass; these typically mediate inflammatory responses.

Cell Biology
- Chemokines coordinate leukocyte trafficking.
- They can have prominent nontrafficking functions:
  - Immunologic effects: lymphocyte proliferation/apoptosis/differentiation/activation, granulocyte degranulation/superoxide production/direct antimicrobial activity.
  - Nonimmunologic effects: development, cancer, angiogenesis.

Biology
- Chemokines act redundantly or nonredundantly, depending on context.
- Host chemokine receptors mediate antimicrobial defense.
- Some pathogens (e.g., HIV and *Plasmodium vivax*) can exploit chemokine receptors to infect the host.
- Excessive or inappropriate chemokine expression may pathologically amplify immunologically mediated disease.

CHEMOKINE RECEPTORS

Chemokine receptors are defined as signal transducers that trigger cellular responses upon binding chemokines. In humans, 23 subtypes of chemokine receptors have been divided into two main groups, the conventional G protein-coupled chemotactic chemokine receptors (GPCRs; $n = 19$) and the

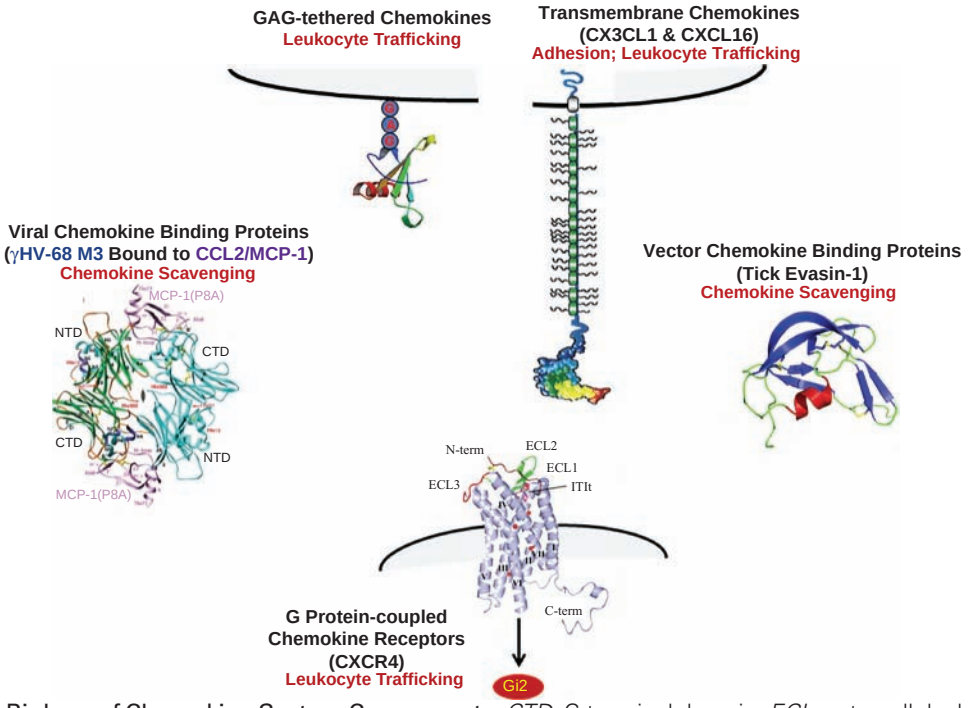


FIG. 15.1 Structural Biology of Chemokine System Components. *CTD*, C-terminal domain; *ECL*, extracellular loop; *GAG*, glycosaminoglycan; *ICL*, intracellular loop; *MCP-1*, monocyte chemoattractant protein-1; *NTD*, N-terminal domain; *N-term*, N-terminus. *P8A*, proline for alanine substitution at position 8. IT1t is a small molecule antagonist bound to CXCR4. (The figure is modified from Bachelerie F, Ben-Baruch A, Burkhardt AM, et al. International Union of Basic and Clinical Pharmacology. [corrected]. LXXXIX. Update on the extended family of chemokine receptors and introducing a new nomenclature for atypical chemokine receptors. *Pharmacol Rev*. 2014;66(1):1–79.)

atypical G protein-uncoupled chemokine receptors (ACKRs; $n = 4$) (see Figs. 15.1 and 15.2B). Chemokine binding, membrane anchoring, and signaling domains for receptors from both groups come from a single polypeptide chain, although evidence exists for both homo- and heterodimers.[2,5,6]

Each chemokine has a unique receptor specificity profile, and vice versa. Five chemokine GPCRs (CXCR4, CXCR5, CCR6, CCR7 and CCR9) pair monogamously with a chemokine ligand, all of which support homeostatic trafficking into, within and out of primary and secondary immune organs. The other 14 chemokine GPCRs bind multiple chemokines restricted mainly to one structural group that support inflammatory signaling.

The chemokine GPCRs are systematically named based on ligand group specificity. The distinction between inflammatory and homeostatic chemokines and chemokine receptors can be vague since the same molecule may multitask. For example, CXCR2 is important for both acute tissue neutrophilic inflammation and homeostatic release of newly produced bone marrow neutrophils to the blood.

Almost all chemokines are chemotactic agonists. A few may be agonists at one GPCR, antagonists at another (e.g., CCL11 agonism at CCR3 and antagonism at CXCR3), and also bind to ACKRs. For chemokines acting at the same GPCR, differential receptor usage, differential regulation of expression, and biased agonism all contribute to nonredundant function in vivo.[7] The sheer size and high degree of promiscuity of the chemokine ligand-receptor repertoire help explain why, in complex chronic inflammatory diseases, the response to monospecific drugs that target the system one component at a time has been underwhelming, given the many types of leukocytes, chemokines, and chemokine receptors that can be involved in pathogenesis.[8]

ATYPICAL CHEMOKINE SYSTEM COMPONENTS

There are three main classes of atypical chemokine system components. The first includes the four atypical chemokine receptors (ACKRs). These bind promiscuously to multiple chemokines either without inducing a signal or with signaling occurring through a β-arrestin-dependent G protein-independent pathway. ACKRs are subclassified by the degree of chemokine promiscuity.[9] ACKR1 and 2 bind overlapping sets of many proinflammatory CC chemokines, and ACKR1 also binds multiple proinflammatory CXC chemokines. Prominent expression of both receptors on tissue barrier cells, including vascular and lymphatic endothelial cells and placental trophoblasts, and of ACKR1 on mature erythrocytes, which hugely outnumber leukocytes in blood, has suggested they may function as chemokine sinks to shape more effective chemokine gradients or to dampen excessive inflammatory responses. For example, inflammatory chemokine scavenging by placental trophoblast-expressed ACKRs may contribute to immunologic tolerance during pregnancy. ACKR1 on endothelial cells may also present chemokines to leukocytes by mediating transcytosis of bound chemokine to the abluminal side of the cell. Internalization of bound chemokine is a general property of all chemokine receptors; however, ACKR2 is specialized for chemokine scavenging by extremely rapid internalization of degradation bound ligands and rapid receptor recycling.

In contrast to ACKR1 and 2, ACKR3 and 4 bind only 2 and 3 chemokines, respectively. They coordinate local chemokine gradient shaping with specific GPCRs for which they share ligand specificity (CXCR4 and CCR7, respectively). For example, gradients of CXCL12 shaped by ACKR3 that then signal chemotactically through CXCR4 mediate zebrafish germ cell

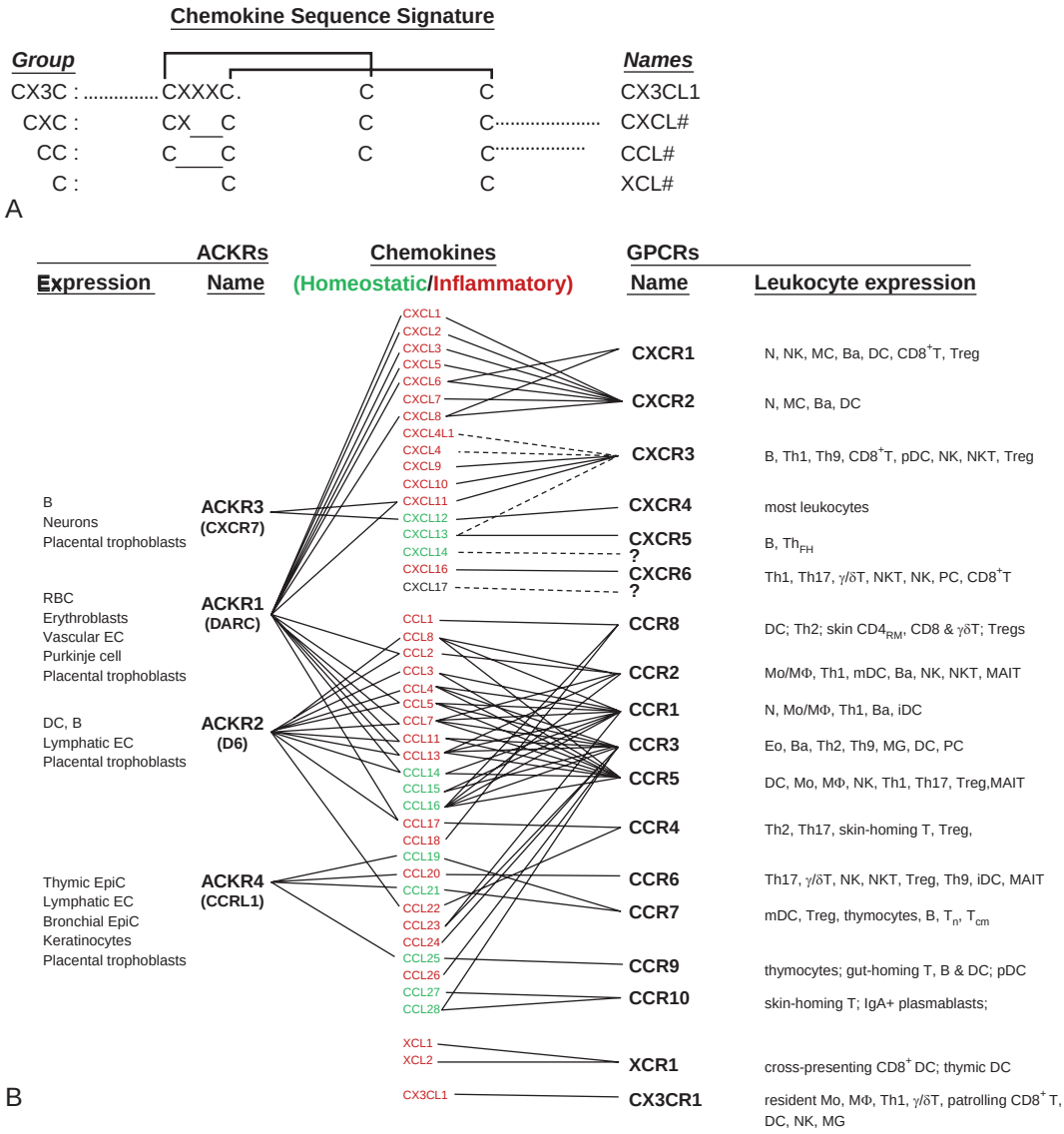


FIG. 15.2 Molecular and Cellular Organization of the Chemokine System. (A) Chemokine classification. Chemokines are defined by the number and arrangement of conserved cysteines, as shown. *Brackets* link cysteines that form disulfide bonds. *X* refers to an amino acid other than cysteine. The *underscore* is a spacer used to optimize the alignment. The N- and C-termini can vary considerably in length (not illustrated). For molecules with four cysteines, there are approximately 24 amino acids between Cys-2 and Cys-3, and 15 amino acids between Cys-3 and Cys-4. At *right* are listed the nomenclature system and the number of human chemokines known in each class (N). (B) Chemokine receptor specificity for chemokine ligands and cells. *ACKR*, Atypical chemokine receptor; *B*, B lymphocyte; *Ba*, basophil; *DC*, dendritic cell; *EC*, endothelial cell; *Eo*, eosinophil; *EpiC*, epithelial cell; *GPCR*, G protein-coupled receptor; *iDC*, immature dendritic cells; *MAIT*, mucosal-associated invariant T cell; *MG*, microglial cells; *Mo*, monocyte; *MΦ*, macrophage; *N*, neutrophil; *NK*, natural killer cell; *NKT*, NKT cell; *PC*, plasma cells; *pDC*, plasmacytoid dendritic cells; *RBC*, red blood cell; T_{cm}, central memory T cells; *Tfh*, follicular helper T cells; *Tmem*, memory T cells; *Treg*, regulatory T cells. *Dashed lines* indicate low affinity and unknown interactions.

migration during development. ACKR4 shaping of gradients for shared ligands signaling through CCR7 may similarly facilitate B, T, and dendritic cell trafficking in mammalian lymph node (Chapter 2). ACKR3 may also function independently of CXCR4: for example, to position marginal zone B cells (Chapter 7) in mouse spleen.

The second class of atypical chemokine system components includes endogenous nonchemokine agonists that act at chemokine receptors (e.g., MIF at CXCR2 and β-defensin-2 at CCR6). The third class includes virally encoded chemokine system mimics, including pirated bona fide chemokines and 7TM chemokine receptors, structurally unique secreted chemokine scavengers, and nonchemokine chemokine receptor agonists or antagonists (see Fig. 15.1).[10] Viral chemokine elements can function to evade the immune system, recruit new target cells, reprogram gene expression for cell proliferation and angiogenesis, and mediate target cell entry. Secreted and structurally unique broadly specific chemokine-binding proteins known as evasins have been identified in tick saliva, which could help explain the lack of inflammation associated with tick bites.

KEY CONCEPTS

Immunologic Classification of the Chemokine System

Homeostatic System

- Constitutively expressed ligands and receptors.
- Important in hematopoiesis and immune surveillance.
- Key receptors: CXCR4 on all leukocytes, especially hematopoietic stem and progenitor cells; CXCR5 on B cells; CCR7 on mature dendritic cells and both naïve and central memory T cells; and gut and skin-specific T-cell homing receptors (CCR9 and CCR10, respectively).

Inflammatory System

- In innate immunity, inducible ligands, and constitutively expressed receptors (e.g., neutrophil CXCR1 and CXCR2, monocyte CCR2 and CX3CR1, eosinophil CCR3, and NK cell CX3CR1).
- In adaptive immunity, inducible ligands and inducible receptors (e.g., CXCR3, CCR4, and CCR6 on Th1, Th2, and Th17 subsets of CD4 T cells, respectively).

Immunologic Classification

Mature hematopoietic cell types express a variably sized signature subset of multiple chemokine receptors. The sole exception is the erythrocyte, which expresses only ACKR1. Several chemokine receptors are consistently and highly expressed on almost all cells of a given subset, acquiring the status of a reliable subset marker (e.g., CXCR1 and CXCR2 for neutrophils, CCR3 for eosinophils, CCR2 and CX3CR1 for different subsets of monocytes, CXCR3 for Th1 cells, CCR4 for Th2 cells, CCR6 for Th17 cells, CXCR5 for B cells, and CCR7 for naïve and central memory T cells and mature dendritic cells).

Homeostatic chemokines are differentially and constitutively expressed in specific microenvironments of primary and secondary immune organs. They direct trafficking of both mature and immature leukocytes (Chapter 16) via constitutively expressed receptors. Noxious stimuli induce inflammatory chemokines in diverse tissue cells and leukocytes. Inflammatory chemokine receptors are constitutively expressed on myeloid (Chapter 39) and NK cells (Chapter 12), and are induced upon activation of effector lymphocytes (Chapter 10). Dynamic shifts in receptor expression occur during dendritic cell (Chapter 6) and NK cell maturation, and during lymphocyte maturation, activation, and differentiation.

The chemokine and chemokine receptor repertoires vary between and within species. For example, the major human CXC chemokine CXCL8 (interleukin-8) is not found in mice and chemokine gene copy number variation and sequence polymorphism in humans affects the risk of acquiring certain diseases (e.g., *CCL3* copy number variation and the *CCR5Δ32* mutation in HIV/AIDS (Chapter 41), and the ACKR1 Duffy mutation in *Plasmodium vivax* malaria (Chapter 29)).

CHEMOKINE PRESENTATION MECHANISMS

Chemokines typically act at the local site of production. They may be presynthesized and stored in granules (e.g., CXCL4 and CXCL7 in platelets) awaiting a proinflammatory release signal, or else be constitutively released, sometimes achieving high concentrations in the blood (e.g., CCL14–16). Others are newly synthesized in response to activation signals (e.g., inflammatory chemokines). Chemokines may be presented to leukocytes as ligands tethered to either extracellular matrix proteins and, as mentioned previously, to endothelial cells either by binding to GAGs or by transmembrane domains in the case of CX3CL1 and CXCL16. The tethering cell may have produced the chemokine or else imported it by transcytosis from neighbors. The ligand binding site includes the receptor N-terminus and one or more extracellular loops, which allow docking of the chemokine core domains, and multiple seven-transmembrane (7TM) domains, which accept the chemokine's N-terminus and are critical for triggering. Additional sites of interaction have been identified by direct structural studies.[2,5,6]

CHEMOKINE ACTIVATION OF CHEMOTACTIC SIGNALING PATHWAYS

Chemokines trigger G protein–coupled chemokine receptors to act as guanine nucleotide exchange factors (GEFs) for heterotrimeric G_i-type G proteins, releasing GDP from and binding GTP to the $G_i\alpha$ subunit (Fig. 15.3).[11,12] The G protein dissociates into α and βγ subunits, which in turn activates diverse G protein–dependent effectors, including phospholipases A2, C (subtypes β2 and β3) and D, phosphatidylinositol-3-kinase γ (PI3Kγ), protein tyrosine kinases (PTK) and phosphatases, low-molecular-weight GTPases, and mitogen-activated protein kinases.

Phospholipase C (PLC) hydrolyzes phosphatidylinositol bisphosphate (PIP_2) to form 1,2-diacylglycerol (DAG) and inositol-1,4,5-trisphosphate (IP_3). IP_3 induces Ca^{2+} release from intracellular stores, which acts with DAG to activate protein kinase C (PKC). PI3Kγ phosphorylates PIP_2 to form PIP_3, which recruits proteins containing pleckstrin homology (PH) or PHOX (PX) domains to lamellipodium, thereby converting shallow analog extracellular chemokine gradients into steep digital intracellular effector gradients. Four PH domain-containing targets—Akt, and GEFs for Rac, Rho, and Cdc42—modulate distinct phases of cell movement in various model systems. Rho regulates cell adhesion and chemotaxis, and myosin contraction. Rac and Cdc42 control lamellipodia and filipodia formation, respectively. Downstream targets of Rac include Pak1, which also regulates myosin contraction.

Chemokine activation of G_i may also inhibit adenylyl cyclase activity via the dissociated $G\alpha_i$ subunit. Chemokine agonists induce phosphorylation of the *C*-tail domain of their receptors to which β arrestin is recruited for alternative signaling and downregulation of the receptor by internalization through a clathrin-dependent mechanism. Chemokines also activate MAP kinases, and PTK with downstream effects on cell migration and gene regulation.

REGULATION OF CHEMOKINE ACTION

Chemokine and chemokine receptor expression can be positively or negatively regulated at the transcriptional level by factors that include proinflammatory cytokines, oxidant stress, hypoxia, viruses, bacterial products such as lipopolysaccharide and N-formylpeptides, cell adhesion, antigen uptake, and T-cell costimulation. Multiple transcription factors regulate chemokine and chemokine receptor expression, including nuclear factor-κB (NF-κB) and cEBP-δ. In innate immunity, proinflammatory cytokines such as interleukin (IL)-1, tumor necrosis factor (TNF), IL-15, and IL-17 (Chapter 14) induce expression of inflammatory chemokines important for recruitment of myeloid and natural killer (NK) cells. In adaptive immunity, signature cytokines of polarized helper T cells (Chapter 11) establish

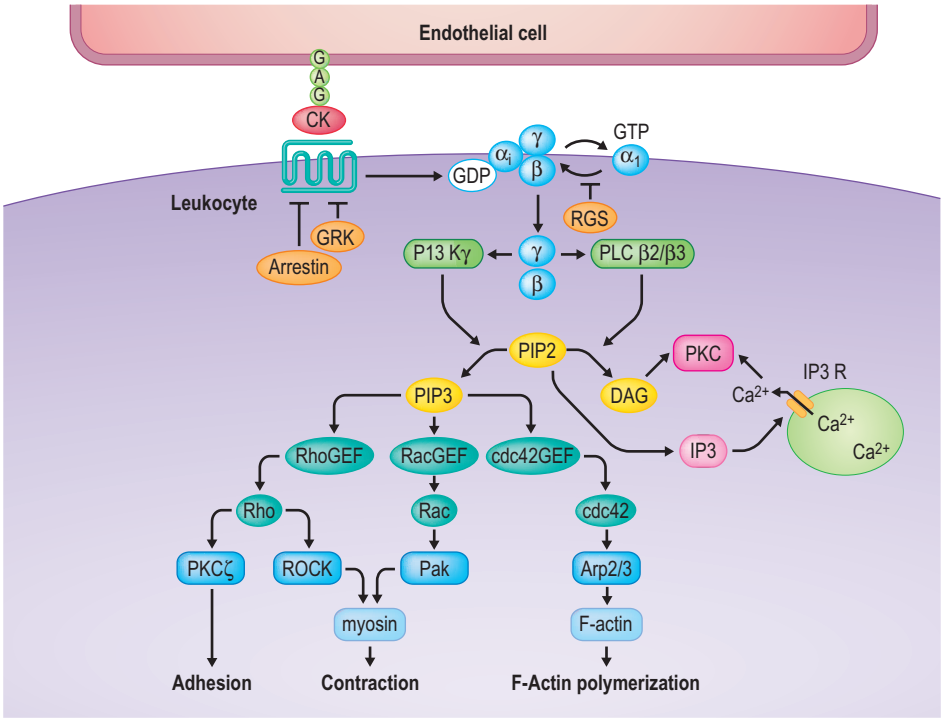


FIG. 15.3 Chemokine Signal Transduction in Chemotaxis. Depicted are key steps in two of the main pathways induced by most chemokines. The PI3Kγ pathway is particularly important for cell migration. Chemokines are able to activate other pathways as well, including non-Gi-type G proteins, protein tyrosine kinases, and MAP kinases. These pathways influence cell proliferation and activation. The model is modified from the Alliance for Cell Signaling (http://www.signaling-gateway.org). *CK*, Chemokine; *DAG*, diacylglycerol; *GAG*, glycosaminoglycan; *GEF*, guanine nucleotide exchange factor *GRK*, G protein-coupled receptor kinase; *IP3*, inositol trisphosphate; *PI3K*, phosphatidylinositol-3-kinase; *PIP*, phosphatidylinositol phosphate; *PKC*, protein kinase C; *PLC*, phospholipase C; *RGS*, regulator of G protein signaling.

positive feedback loops for tissue cell production of signature chemokines able to specifically recruit additional helper T cells, thereby reinforcing polarization of the response. Interferons, glucocorticoids, and antiinflammatory cytokines (e.g., IL-10, transforming growth factor-β [TGF-β]) can inhibit inflammatory chemokine gene expression. Chemokine system components can also be regulated at the level of mRNA stability by *cis* motifs, microRNAs and long-noncoding RNAs (Chapter 19).

Chemokine genes may generate functional variants by alternative splicing and posttranslational modification, especially N- and C-terminal proteolytic trimming.[13] Proteases can target many chemokines (e.g., CD26 [dipeptidyl peptidase IV] and matrix metalloproteinases [MMP]), or few or only one (e.g., TACE [the TNF-α converting enzyme], plasmin, urokinase plasminogen activator and cathepsin G). Chemokines can be scavenged not only by ACKRs, but also by receptor decoys and autoantibodies, or they can be blocked by endogenous receptor antagonists. In addition, cytokines may convert a signaling receptor into a decoy (e.g., IL-10 uncouples CCR2 signaling in monocytes). Chemokine receptors may exist in different functionally distinct conformational states and may be modulated in signalosomes, which may include physical and/or functional interactions with other chemokine receptors and accessory molecules. Likewise, chemokine action may be modulated by binding to other chemokines and accessory molecules: for example, the ability of the chromatin-binding protein HMGB1 released from dying cells to bind CXCL12, thereby enhancing its signaling potential at CXCR4.

LEUKOCYTE RESPONSES TO CHEMOKINES

Whereas all leukocyte subtypes migrate in response to one or more chemokines, each subtype can also respond in additional stereotypical ways. For example, lymphocytes may proliferate or undergo apoptosis (Chapter 17), and strengthen immune synapses (Chapter 10) or release immunoregulatory (Chapter 13) and cytotoxic factors (Chapter 12); and granulocytes (Chapter 39) can release antimicrobial and inflammatory mediators (e.g., superoxide, defensins, proteases, histamine, eicosanoids).

The mechanism of leukocyte migration can vary depending on the leukocyte subtype and the environment. In the classic multistep model of leukocyte transendothelial migration (Chapter 16),[14] an initial chemokine-independent step involves leukocytes rolling on inflamed endothelium by reversible interactions between endothelial selectins (e.g., L-selectin) and leukocyte selectin ligands (e.g., sialyl-Lewis^x). Next, chemokines presented by endothelium via GAGs stimulate cognate chemokine receptors on rolling leukocytes to induce expression of activated β_2 integrins, which mediate firm adhesion of leukocytes to endothelium by binding to endothelial intercellular adhesion molecule (ICAMs).

Leukocytes sense chemokine gradients, polarize, and become poised to crawl. Motion involves shear-dependent coordinated cytoskeletal remodeling, involving expansion of the leading edge (lamellipodium), myosin-based contraction at the trailing edge (uropod), release of the uropod

from substrate, and membrane lipid movement. Navigation across endothelium and through tissue may involve additional steps in which specific chemoattractants and adhesion molecules play differential roles and the steps may vary for different barriers in tissue.[15] For example, some chemokines are specialized in arrest of leukocytes on endothelium and others are critical for terminal transendothelial migration (Fig. 15.4A).[16]

CHEMOKINE REGULATION OF HEMATOPOIESIS

Chemokines play major roles in hematopoiesis (see Fig. 15.4B).[17] Both CXCR2 and CXCR4 are expressed on hematopoietic stem cells (HSCs) and neutrophils, and the balance of CXCR2 and CXCR4 signaling influences homing to and egress from bone marrow niches. CXCL12 is produced by CXCL12-associated reticular (CAR) cells in hematopoietic stem cell (HSC) niches whereas bone marrow microvascular endothelial cells produce oppositely oriented gradients of the CXCR2 agonists CXCL1 and CXCL2.[18] HSCs and neutrophils can both be rapidly mobilized pharmacologically by granulocyte colony-stimulating factor (G-CSF), which induces CXCR4 degradation by neutrophil elastase, or by CXCR4 antagonists and CXCR2 agonists. Mice lacking either CXCL12 or its receptor CXCR4 display defective bone marrow myelopoiesis, whereas CXCR2 knockout mice demonstrate resting peripheral blood and bone marrow neutrophilia and impaired acute neutrophil migration into tissue under inflammatory conditions. The atypical CXCL12 receptor ACKR3 does not regulate hematopoiesis, but instead regulates marginal zone B-cell positioning in spleen (Chapters 2 and 7), among other functions. CCR2 is important for monocyte release from bone marrow.

Chemokines and chemokine receptors are differentially expressed in thymus and coordinate thymocyte migration from cortex to medulla.[19] CCR9 and its sole ligand CCL25 are particularly important in this process. CCL25 is expressed by medullary dendritic cells and both cortical and medullary epithelial cells. CCR9 is expressed on the majority of immature $CD4^+CD8^+$ thymocytes but is downregulated during transition to the $CD4^+$ or $CD8^+$ single-positive stage. Just before thymic egress, thymocytes become CCR9 negative and upregulate L-selectin. $CD4^+CD8^+$ thymocyte transition in the cortex to $CD4^+$ or $CD8^+$ single-positive thymocytes in the medulla involves upregulation of CCR4 and CCR7, receptors for CCL22, and CCL19 and CCL21, respectively, which are expressed in the medullary stroma. Accordingly, these chemokines attract thymocytes between the late cortical and medullary stages of development in vitro. Neutralization studies suggest that egress of newly formed T cells from fetal thymus to the circulation is mediated by CCL19, which localizes to endothelial cells of medullary venules and acts at CCR7 on mature thymocytes.

CHEMOKINE REGULATION OF THE IMMUNE RESPONSE

Innate Immunity

Platelet-Derived Chemokines

Made primarily during platelet development, stored in platelet α granules, and rapidly released during platelet degranulation, CXCL4 and CXCL7 are among the first chemokines to appear at sites of tissue injury and infection, particularly when there is hemorrhage and vascular damage, and can reach high concentrations.[20] CXCL7 can function as an immediate-early mediator of neutrophil recruitment released from platelets at sites of inflammation. Although it is not a prominent leukocyte chemoattractant, CXCL4 complements CXCL7 by inducing neutrophil secondary granule exocytosis and release of matrix-degrading enzymes, which may facilitate neutrophil penetration of infected or injured tissues.

Myeloid Cell Trafficking

Once myeloid cells are released from the bone marrow, they undergo specific trafficking itineraries and, in some cases, become resident in tissue. CXCL14 is important for macrophage (Chapter 3) positioning in lung by a still undefined receptor, and CCL11 and its receptor CCR3 are important for positioning of eosinophils (Chapter 45) in spleen and the gastrointestinal tract (Chapter 24). CX3CR1 regulates localization of myeloid DCs in Peyer patches and may be important for antigen sampling from the intestine.

All seven CXC chemokines containing an N-terminal ELR motif preferentially recruit neutrophils in vitro by binding to CXCR2. CXCL6 and CXCL8 are also potent agonists at the closely related neutrophil receptor CXCR1.[21] CXCR2 chemokines are rapidly inducible in most tissue cells. Temporospatial differences in expression provide a mechanism for graded navigation of neutrophils from blood into tissue. Blocking studies have demonstrated the importance of CXCL8 and CXCR2 in neutrophil accumulation in response to infectious and noninfectious stimuli. CXCR1 does not appear to mediate neutrophil recruitment into infected sites but is instead involved in activation of antimicrobial effector mechanisms.

Intradermal injection of CXCL8 in man causes rapid (<30 minutes) and selective accumulation of large numbers of neutrophils in perivascular regions of the skin (Chapter 23) without causing edema. Tissue-specific transgenic overexpression of mouse CXCL8 paralogues KC and MIP-2 suggests that these factors may recruit cells, but may not independently activate cytotoxic mechanisms. In a human blister model, endogenous CXCL8 peaks at approximately 24 hours, whereas C5a (Chapters 3 and 39) and leukotriene B4, which also recruit neutrophils, appear earlier. Thus, the primary role of CXCL8 may be to amplify early inflammatory responses initiated by other types of chemoattractants. CXCL1, 2, 3, 7, and 8 also induce basophil chemotaxis and histamine release in vitro, which together with other factors, including complement-derived anaphylatoxins C3a and C5a, promote vasodilatation early during innate immune responses.

Monocyte recruitment typically follows neutrophil accumulation with delayed kinetics and can be mediated by multiple inflammatory CC receptors and CX3CR1. CCR2 and CX3CR1 are particularly important and define two monocyte subsets, $CX3CR1^{hi}CCR2^-$ and $CX3CR1^{lo}CCR2^+$, which are referred to as "resident" and "inflammatory" monocytes because of their distinct trafficking characteristics.[22]

NK Cells

Human NK cell subsets (Chapter 12) express unique repertoires of chemokine receptors. The $CD56^{dim}CD16^+$ subset, which is associated with high cytotoxic capacity and low cytokine production, expresses primarily CXCR1 and CX3CR1. The minor subset of $CD56^{bright}CD16^{dim}$ cells,

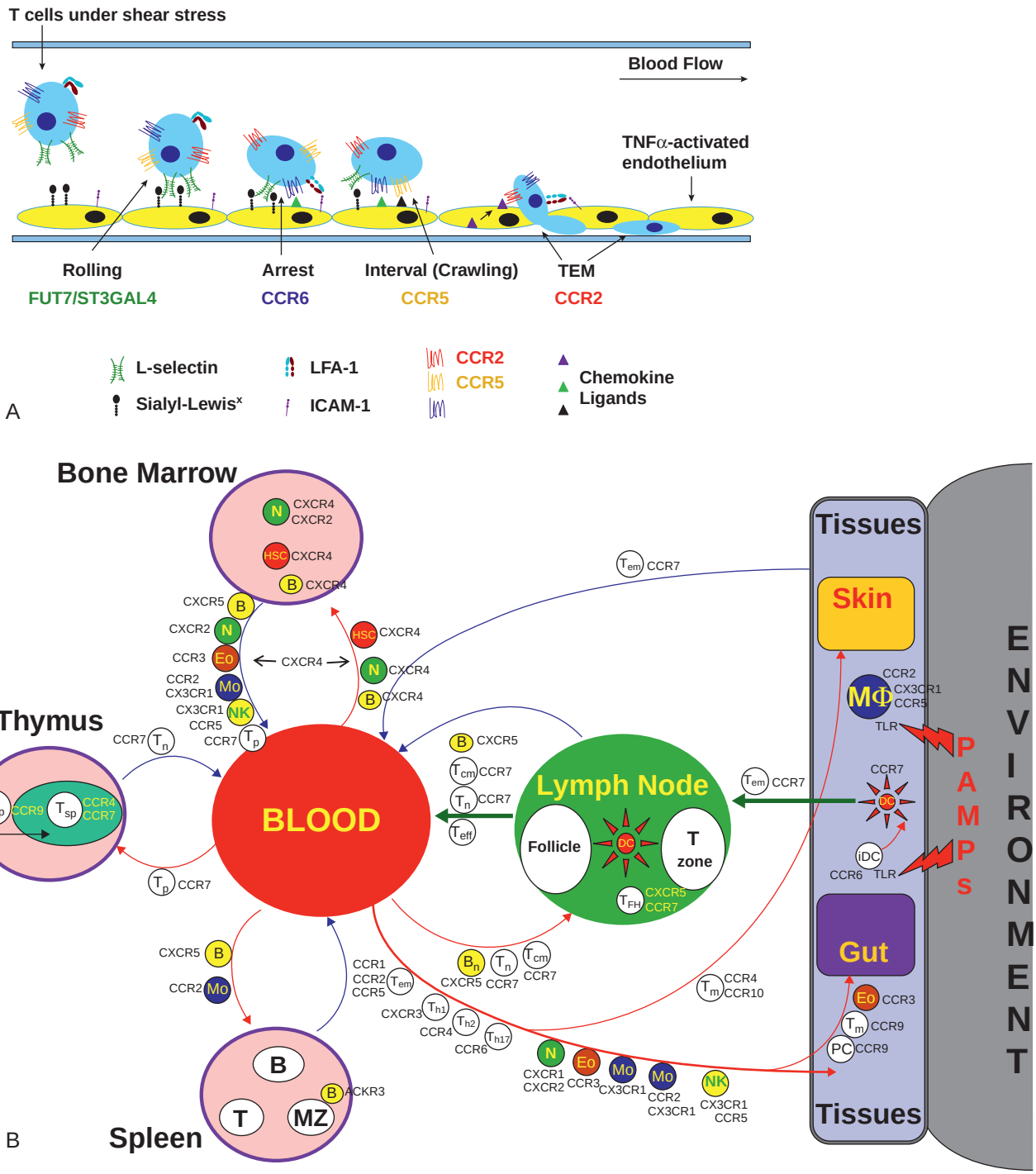


FIG. 15.4 Chemokine Receptor Control of Leukocyte Trafficking. (A) Multistep model of leukocyte transendothelial migration *(TEM)*. An example of MAIT cell trafficking across TNF-activated human umbilical vein endothelial cells *(HUVEC)* is summarized.[16] FUT7/ST3GAL4 is a fucosyltransferase important for biosynthesis of L-selectin ligand. (B) Chemokine regulation of leukocyte trafficking itineraries under homeostatic and inflammatory conditions. Shown are key receptors on the indicated leukocyte subtypes that control trafficking into, within, and out of primary and secondary immune organs and the periphery. *Eo*, Eosinophil; *GC*, germinal center *HSC*, hematopoietic stem cell; *iDC*, immature dendritic cells; *Mo*, monocyte; *Mφ*, macrophage; *N*, neutrophil; *NK*, natural killer cell; *PC*, plasma cell; T_{cm}, central memory T cells; T_{eff}, effector memory T cells; T_{FH}, follicular helper T cells; T_m, memory T cells; T_n, naïve T cells; T_p, precursor T cells. The model is based primarily on studies of mice where the relevant gene has been inactivated by gene targeting. (Panel A was created by Joshua Farber, NIAID.)

which produce large amounts of cytokines but have low killing capacity, preferentially express CCR7. The exact profile of chemokine receptor expression can be modulated by adherence and stimulation ex vivo with IL-2. Cognate chemokines chemoattract NK cells and promote degranulation and killing.

Dendritic Cells and Transition to the Adaptive Immune Response

Transition from innate to adaptive phases of the immune response involves antigen uptake by antigen-presenting cells (Chapter 6), especially dendritic cells, mediated by Fc and complement phagocytic receptors; as well as pattern recognition receptors (PRRs) (Chapter 3), including DC-SIGN and Toll-like receptors (TLRs), which differentially induce inflammatory chemokine expression. Pathogens can skew the nature and magnitude of the immune response in a specific direction by means of specific PRR ligands.

The chemokine receptors expressed on DCs vary depending on the nature of the inflammatory stimulus and type. For example, blood-derived plasmacytoid and myeloid DCs express a similar repertoire of inflammatory chemoattractant receptors, but they are functional only on myeloid DCs. CCL3, CCL4, and CCL5 may be particularly important for recruiting additional mononuclear phagocytes and DCs to sites of infection. This can amplify the late stage of the innate immune response, and, in the extreme, can progress to endotoxic shock. Genetic disruption of the CCL3/CCL4/CCL5 receptor CCR5 renders mice relatively resistant to LPS-induced endotoxemia.

Adaptive Immunity

Afferent Trafficking to Secondary Lymphoid Tissue

The homeostatic receptors CXCR5 and CCR7 and their ligands are major regulators of the immune response, acting at the level of B and T lymphocyte and DC trafficking to and within secondary lymphoid tissue (Chapter 16).[19,23,24] DC maturation in peripheral tissues is associated with downregulation of inflammatory receptors such as CCR6, which is important for recruitment, migration and retention in the periphery, and reciprocal upregulation of CCR7, which mediates mature DC migration to draining lymph nodes. Inflammatory receptors, such as CCR2, may also contribute to afferent lymphatic trafficking. In addition to these general DC receptors, XCR1 is expressed on a niche subset of cross-presenting CD8 dendritic cells. CCR7 is also a major lymph node trafficking receptor for naïve T cells and can mediate activated T-cell exit from inflamed tissue.

The CCR7 ligand CCL21 is constitutively expressed on afferent lymphatic endothelium, high endothelial venules (HEVs), stromal cells and interdigitating dendritic cells in T-cell zones of lymph node, Peyer patch, mucosa-associated lymphoid tissue, and spleen. It is not expressed in B-cell zones or sinuses. CCL19, another CCR7 ligand, is also restricted to the T-cell zone and is expressed on interdigitating DCs.

CCR7$^{-/-}$ mice and the *plt* mouse, which is naturally deficient in CCL19 and a CCL21 isoform expressed in secondary lymphoid organs, have similar phenotypes: atrophic T-cell zones populated by a paucity of naïve T cells. This and the failure of activated DCs to migrate to lymph node from the skin of these mice explain why contact sensitivity (Chapter 48), delayed-type hypersensitivity, and antibody production are severely impaired. However, lymph node trafficking is not completely abolished in these mice, which may develop autoimmune phenomena.

CXCR5 is expressed on all peripheral blood and lymph node B cells as well as on some T cells. Its ligand CXCL13 is expressed constitutively on follicular HEV and controls trafficking of CXCR5 positive B and T cells from the blood into follicles (Chapter 16). In *Cxcr5*$^{-/-}$ mice, B cells do not migrate to lymph node, Peyer patches are abnormal, and inguinal lymph nodes are absent. CXCL13 is also required for B-1 cell homing, natural antibody production, and body cavity immunity (Chapter 7). *Cxcr5*$^{-/-}$ mice still can produce antibody, perhaps in part because B cells and follicular DC, by an unknown mechanism, are able to form ectopic germinal centers within T-cell zones of the periarteriolar lymphocyte sheath of spleen.

Migration Within Lymph Node Microenvironments

CXCR5 is expressed on a majority of memory CD4 T cells (Chapter 11) in the follicles of inflamed tonsils. Follicular helper T cells (T_{FH}), a CD57$^+$ subset of CXCR5$^+$ T cells, lack CCR7, which licenses them to move from the T-cell zone following activation to the follicles where they provide help for B-cell maturation and antibody production. Reciprocally, B cells activated by antigen in the follicles upregulate CCR7 and move towards the T-cell zone. Thus, B-T interaction can be facilitated by reciprocal movement of these cells, which may be influenced by the balance of chemokines made in adjacent lymphoid zones. CXCR4 signaling is also important in naïve and memory B-cell trafficking to germinal centers, and CCR5 ligands can guide antigen-specific CD8 T cells to sites of helper T cell–dendritic cell conjugation for activation in lymph nodes.

Efferent Trafficking

Naïve T cells that do not encounter antigen continue to recirculate between the blood and secondary lymphoid tissue in a CCR7-dependent manner. Most antigen-activated T cells die by apoptosis (Chapter 17). The survivors can be divided into functionally distinct memory and effector cell subsets with characteristic patterns of chemokine receptor expression. The CD4 memory subsets (Chapter 11) include effector memory cells (T_{EM}) and central memory cells (T_{CM}). T_{EM} do not express L-selectin or CCR7 and surveil peripheral tissues, rapidly releasing cytokines in response to activation by recall antigens. T_{CM} express CCR7 and L-selectin and recirculate between blood and secondary lymphoid organs, where they efficiently interact with cognate antigen-presenting DCs and differentiate into effector cells upon costimulation.

Upon antigen activation, effector CD4 T-cell subsets downregulate CCR7 and upregulate inflammatory chemokine receptors that mediate trafficking to tissue sites. Exit from lymph node via efferent lymphatics is mediated by additional mechanisms including S1PR (sphingosine-1-phosphate receptor) signaling, which can be blocked by the immunosuppressive drug FTY720 (fingolimod). CXCR3 is the signature trafficking receptor for Th1 cells, reinforced by CXCR6, CCR2, and CCR5; CCR4 is the signature chemokine receptor expressed on Th2 cells, reinforced by CCR3 and CCR8; and CCR6 is the signature chemokine receptor for Th17 cells.

Polarization is established in each case by positive feedback loops involving subset-specific signature cytokine induction in tissue cells of the chemokine ligands for the corresponding signature chemokine receptor. For example, the Th1 cytokine IFN-γ induces production in tissue cells of the CXCR3 agonists CXCL9–11, which maintain Th1 polarization by recruiting Th1 cells and by blocking the Th2 chemokine receptor CCR3. CCR3 is also expressed on eosinophils and basophils, which participate in Th2-type allergic inflammation. Each co-expressed receptor may have nonredundant function. For example, CCL8 signaling via CCR8 is a nonredundant pathway for recruiting a population of highly differentiated IL-5$^+$ Th2 cells to skin that is important in a mouse model of atopic dermatitis (Chapter 48). Tregs (Chapter 13) also

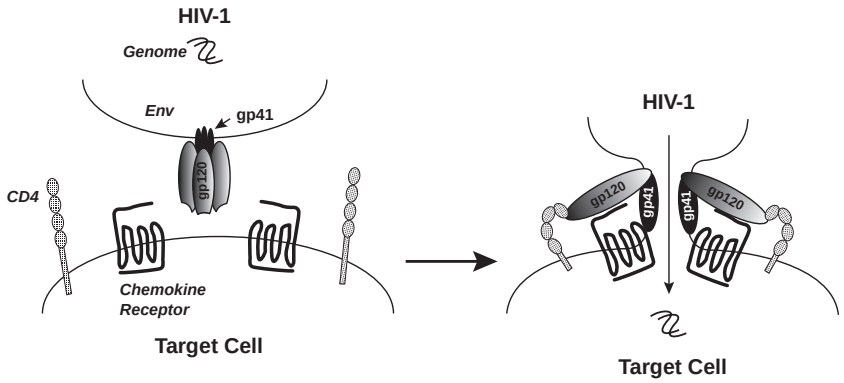


FIG. 15.5 Human Immunodeficiency Virus 1 (HIV-1) Usage of Chemokine Receptors for Cell Entry. See text for details.

express a distinct subset of chemokine receptors. CCR4 and CCR5 are expressed on most freshly isolated natural Tregs, with less consistent expression of CXCR3 and CXCR4.

Mucosa-associated invariant T (MAIT) cells (Chapter 3) are a prominent subset of human CD8 TCR αβ effector/memory cells in the blood and mucosa that express the MHC (Chapter 5) class 1-like molecule MR1 and a semi-invariant TCRα chain specific for riboflavin metabolites. They are thought to be important in innate-like T-cell responses to bacteria, and have been shown to migrate using CCL20/CCR6 as an endothelial arrest system and CCR2 for transendothelial migration (see Fig. 15.4A).[16] Conventional TCR $\alpha\beta^+$ CD8 cytotoxic T cells express CXCR3, CCR2, CCR5 and other inflammatory chemokine receptors. Optimal development of resident memory CD8 T cells in the skin involves signaling by CXCR6 and CCR10.

Tissue-Specific Lymphocyte Homing

Cutaneous lymphocyte-associated antigen (CLA^+) T lymphocytes, which home to skin, preferentially express CCR4 and CCR10. The CCR4 ligand CCL22 is made by resident dermal macrophages and DCs, whereas the CCR10 ligand CCL27 is made by keratinocytes. Blocking both pathways, but not either one alone, has been reported to inhibit lymphocyte recruitment to the skin in a delayed-type hypersensitivity model.

Homing to small intestine is determined in part by T-lymphocyte expression of the integrin $\alpha_4\beta_7$ and CCR9. The $\alpha_4\beta_7$ ligand MADCAM-1 and the CCR9 ligand CCL25 co-localize on normal and inflamed small intestine endothelium. Most T cells in the intraepithelial and lamina propria zones of the small intestine express CCR9. These cells, which are mainly $TCR\gamma\delta^+$ or $TCR\alpha\beta^+$ $CD8\alpha\beta^+$, are reduced in small intestine from $CCR9^{-/-}$ mice.

As B cells differentiate into plasma cells, they downregulate CXCR5 and CCR7 and exit the lymph node. B immunoblasts expressing IgG coordinately upregulate CXCR4, which promotes homing to the bone marrow, whereas B immunoblasts expressing IgA specifically migrate to mucosal sites. Like gut-homing T cells, B immunoblasts that home to small intestine express $\alpha_4\beta_7$ integrin and CCR9 and respond to CCL25.

CLINICAL CORRELATES

Chemokines and Disease

There is a vast literature addressing the presence and potential clinical relevance of chemokines in human disease. This section provides only a sampling of this work, highlighting those diseases with the strongest evidence for a role in pathogenesis in humans, in some cases leading to clinical trials and new treatments.

CLINICAL RELEVANCE

Examples of Chemokine and Chemokine Receptor Determinants of Human Disease

- WHIM syndrome: Mendelian combined immunodeficiency due to gain-of-function mutations in *CXCR4*
- *Plasmodium vivax malaria*: Protection due to a nonfunctional promoter variant in ACKR1/Duffy. Sub-Saharan Africa variant abrogates expression on erythrocytes, preventing parasite cell entry
- HIV: Protection from cell entry by the virus conferred by homozygous loss-of-function mutation *CCR5Δ32* and by FDA-approved CCR5 antagonist maraviroc/Selzentry (Pfizer)
- AIDS disease progression rate: Slowed by heterozygous *CCR5Δ32*
- West Nile virus disease: Increased risk with homozygous *CCR5Δ32*
- Kaposi sarcoma: Human herpesvirus 8 vGPCR
- Age-related macular degeneration: Increased risk with *CX3CR1 M280* allele
- Cardiovascular disease: Decreased risk with *CX3CR1 M280* allele
- Autoimmune heparin-induced thrombocytopenia: Caused by autoantibodies to CXCL4-heparin complexes
- Chronic renal allograft rejection: Reduced risk with homozygous *CCR5Δ32*
- Rheumatoid arthritis: Reduced risk with *CCR5Δ32*
- Eosinophilic esophagitis: Associated with CCL26 variant

HIV/AIDS

Human immunodeficiency virus (HIV) (Chapter 41) envelope glycoprotein gp120 mediates fusion of the viral envelope with the target cell membrane by binding to both CD4 and a specific chemokine receptor, which is referred to in this context as an HIV coreceptor (Fig. 15.5).[25] CCR5 and CXCR4, the most important HIV coreceptors, have been shown to physically associate with both CD4 and gp120, and HIV viruses are classified into three main subtypes (X4, R5, and R5/X4) by their specific usage of these receptors.

Individuals homozygous for *CCR5Δ32*, a nonfunctional allele that occurs in approximately 20% of North American Caucasians, appear healthy and are highly resistant to R5 HIV, the main transmitting strain. HIV-infected heterozygotes experience slower disease progression. The CCR5 antagonist maraviroc (Selzentry; Pfizer, New York, USA) is approved for treatment of patients with HIV/acquired immunodeficiency syndrome (AIDS). Remarkably, two HIV^+ patients (the "London patient" and the "Berlin patient") both developed hematologic malignancies that were fortuitously treated after cytoreductive

CLINICAL RELEVANCE

Examples of Approved Drugs Targeting the Chemokine System

- Maraviroc (Selzentry; Pfizer): Small-molecule oral CCR5 antagonist FDA approved in 2007 that blocked target cell entry by R5 strains of HIV.
- Plerixafor (Mozobil, AMD3100; Sanofi): Small-molecule parenteral CXCR4 antagonist FDA approved in 2008 as combined treatment with G-CSF for HSC mobilization for collection and autologous transplantation in patients with multiple myeloma and non-Hodgkin lymphoma receiving cytoreductive therapy.
- Mogamulizumab-kpkc (Poteligeo; Kyowa-Kirin): Humanized afucosylated monoclonal antibody that blocks CCR4, FDA approved in 2018 for treatment refractory patients with mycosis fungoides and Sezary syndrome.

therapy by rescue bone marrow transplant from *CCR5Δ32* homozygote donors, resulting in sustained undetectable viral load off antiretroviral therapy, suggesting a functional "cure." This has motivated gene editing cure strategies to inactivate *CCR5* in T cells in HIV/AIDS.[26] The creation of germline CRISPR-mutated *CCR5* humans has also been reported and broadly criticized on ethical grounds. CCR5 appears to be important in the pathogenesis of West Nile virus (WNV) infection, but instead of facilitating infection, it plays a protective role by mediating antiviral T-cell and monocyte trafficking to infected brain.

Malaria

The *Plasmodium vivax* Duffy binding protein (PvDBP) is a protein expressed in micronemes of the merozoite form of the parasite that facilitates infection of erythrocytes by binding via a cysteine-rich domain to the N-terminal domain of ACKR1 (originally known as Duffy and later as DARC [Duffy Antigen Receptor for Chemokines]. This promotes junction formation during invasion. *P. vivax*-malaria is rare in sub-Saharan Africa due to genetic deficiency in ACKR1 caused by a single nucleotide substitution in the gene promoter (−46 C) affecting an erythroid-specific GATA-1 site.[27] ACKR1 deficiency is associated with benign ethnic neutropenia.

WHIM syndrome

Autosomal dominant truncating mutations in CXCR4 that inhibit CXCL12-dependent receptor desensitization cause WHIM syndrome, a rare Mendelian disease characterized by the acronymic tetrad of *w*arts, *h*ypogammaglobulinemia, *i*nfection, and *m*yelokathexis (neutropenia without maturation arrest).[28] Myelokathexis and infection reflect the bone marrow retention function of CXCR4 for myeloid cells, inhibiting their egress to blood. Most patients are also lymphopenic. Mechanism-based treatment with the selective CXCR4 antagonists plerixafor (Mozobil, AMD3100; Sanofi, Massachusetts, USA) and mavorixafor (X4-Pharma, Massachusetts, USA) have been advanced to phase 3 clinical trials in WHIM syndrome. Plerixafor is a short-acting parenteral bicyclam FDA-approved drug for use with G-CSF to mobilize HSCs for autologous transplantation in patients with multiple myeloma or lymphoma undergoing cytoreductive therapy. Mavorixafor is a new oral agent.

A patient has been identified who was spontaneously cured of WHIM syndrome as an adult by an acquired second complex mutational event known as chromothripsis (chromosome shattering), which deleted the disease copy of *CXCR4* in a single HSC. The affected $CXCR4^{+/o}$ HSC acquired a selective advantage and repopulated the bone marrow, resulting in durable correction of neutropenia and monocytopenia, clearance of warts, and cessation of recurrent bacterial infections. $CXCR4^{+/o}$ mouse HSCs have enhanced proliferative capacity in vivo and phenocopy the engraftment advantage in the patient, suggesting specific *WHIM* allele deletion as a gene therapy cure strategy.

ON THE HORIZON

Basic Principles

- Elucidation of the stoichiometry, presentation requirements, and mechanisms of receptor activation by chemokines through structural studies of chemokine receptors complexed to chemokines, G proteins, and other effectors.
- A more comprehensive understanding of chemokine immunoregulation by means of systems immunology using proteomics and single cell technologies.
- More precise targeting of the chemokine system in disease through biased agonism.
- Detailed delineation of mechanisms of microbial pathogenesis, including immunomodulation by pathogens, by using microbial chemokine mimics.
- A more comprehensive understanding of chemokine multitasking and for mining natural products as therapeutics through the study of the evolution of the chemokine system.

Clinical Correlates

- More rapid development of novel therapeutics by development of disease validation and targeting pipelines.
- Unbiased systematic surveys of chemokine and chemokine receptor expression in the context of disease models and human disease samples from centralized Biobanks.
  - Targeted interrogation by CRISPR-mediated gene knockout technology in mice.
  - Drug discovery by target structure-based design.
  - Clinical trials drawing patients from large disease registries.
- Cure strategies in HIV/AIDS, WHIM syndrome and potentially other diseases based on chemokine system gene editing.
- Development of a next generation of multi-chemokine receptor blocking agents, based on the recognition that chemokines may drive chronic inflammation in humans in a combinatorial manner.
- Define the contribution of chemokines and chemokine receptors to Covid-19 pathogenesis.

Atherosclerosis

Macrophages are the dominant leukocyte present in atherosclerotic lesions and are associated with the presence of macrophage-targeted chemokines such as CCL2, CCL5, and CX3CL1.[29] $CCL2^{-/-}$, $CCR2^{-/-}$, $CX3CL1^{-/-}$, and $CX3CR1^{-/-}$ mice on the atherogenic $ApoE^{-/-}$ genetic background demonstrate smaller lesions and reduced accumulation of macrophages in the vessel wall, and mice lacking both *Ccl2* and *Cx3cr1* treated with Met-CCL5, a chemically modified variant of CCL5 that blocks CCR1, CCR3, and CCR5, were almost completely protected. Adoptive transfer studies with bone marrow from $Cxcr2^{-/-}$ mice have also revealed a role for CXCR2 in promoting atherosclerosis in mouse models, apparently by promoting monocyte adhesion to early atherosclerotic endothelium through interaction with its mouse ligand KC and activation of the VLA-4/VCAM-1 adhesion system. The CX3CR1 genetic variant CX3CR1-M280, which lacks normal CX3CL1-dependent adhesive function, is associated with reduced risk of atherosclerotic vascular disease. Mechanistic studies suggest that CX3CL1 on coronary artery smooth muscle cells anchors macrophages via CX3CR1.

Kaposi Sarcoma

HHV8 (also called Kaposi sarcoma [KS]–associated herpesvirus) is an example of a virus laden with genes encoding pirated

chemokines and chemokine receptors. It encodes three CC chemokines, vMIP-I, II, and III, as well as a constitutively active CC/CXC chemokine receptor named vGPCR, encoded by *ORF74*. All these factors are angiogenic and may contribute to the pathogenesis of KS, a highly vascular multicentric nonclonal tumor caused by HHV8, typically in the setting of immunosuppression such as in HIV/AIDS. Consistent with this, vGPCR induces KS-like tumors when expressed in transgenic mice. The mechanism may involve activation of NF-κB and induction of angiogenic factors and proinflammatory cytokines. This virus appears to have converted a hijacked receptor, probably CXCR2, into a regulator of gene expression.[10]

Autoimmunity

An established risk factor for thromboembolic complications of heparin therapy, heparin-induced thrombocytopenia occurs in 1% to 5% of patients treated with heparin and is the result of autoantibodies that bind specifically to CXCL4-heparin complexes in plasma. In general, T cell–dependent autoimmune diseases in humans, such as psoriasis (Chapter 64), multiple sclerosis (MS) (Chapter 66), rheumatoid arthritis (RA) (Chapter 53), and type I diabetes mellitus (Chapter 71), are associated with inflammatory chemokines and tissue infiltration by T lymphocytes and monocytes expressing inflammatory chemokine receptors.[30,31] In a mouse model of immune complex–induced arthritis, the specific contributions of Ccr1, Cxcr2, Blt1 (a leukotriene B4 receptor) and the C5a receptor have been dissected in detail in joint venules at the level of adhesion and transendothelial migration. A dominant negative antagonist of CCL2 inhibits arthritis in the MRL-*lpr* mouse model of RA, suggesting a potential role for CCL2 and CCR2. Met-CCL5 treatment was beneficial in a collagen-induced arthritis model in DBA/I mice. The importance of the CCL20/CCR6 axis has been validated in multiple mouse models of psoriasis, working in part through IL-17 produced by recruited γδ T cells in the dermis.

CCR9 is an attractive drug target in Crohn disease (Chapter 75) due its important role in T-cell homing to the gut and proof-of-principle preclinical evidence in mouse models. However, the specific allosteric CCR9 antagonist CCX282-B (ChemoCentryx), also known as vercirnon, has demonstrated mixed results in phase 2/3 clinical trials.

Acute Neutrophil-Mediated Inflammatory Disorders

Many neutrophil-mediated human diseases have been associated with the presence of CXCL8, including psoriasis, gout, acute glomerulonephritis, acute respiratory distress syndrome (ARDS), RA, and ischemia-reperfusion injury.[21] Over 25 years, 8 small-molecule CXCR2- and CXCR1/2-specific antagonists have been tested in 54 registered clinical trials for diverse disease indications, including chronic obstructive pulmonary disease (COPD), inflammation damage after islet cell transplantation, bullous pemphigoid, asthma, respiratory syncytial virus (RSV) infection, psoriasis, cancer, myelodysplastic syndrome, cystic fibrosis, and ulcerative colitis. Most of the inflammation trials were terminated. The CXCR1/2 antagonist reparixin (Dompe) was advanced into phase 3 for islet cell transplantation. The cancer trials opened recently and are enrolling.

Transplant Rejection

The most extensive analysis of the role of chemokines in transplant rejection (Chapter 89) has been carried out in an MHC class I/II-mismatched cardiac allograft rejection model in the mouse, which is mediated by a Th1 immune response. Sets of inflammatory chemokines are found in the mouse model that resemble those in the human disease; these appear in a strict temporal sequence. Analysis of knockout mice has demonstrated that while multiple chemokine receptors contribute to rejection in this model, there is a marked rank order: Cxcr3 > > Ccr5 > Ccr1=Cx3cr1=Ccr2. Most impressively, rejection and graft arteriosclerosis do not occur if the recipient mouse, treated with a brief, subtherapeutic course of cyclosporine, is *Cxcr3*$^{-/-}$ or if the donor heart is *Cxcl10*$^{-/-}$. In humans, CCR5 may be important in chronic kidney allograft rejection, since individuals homozygous for *CCR5Δ32* are underrepresented among patients with this outcome in a large German kidney transplantation cohort.

Allergic Airway Disease

Chemokine receptors associated with asthma (Chapter 43) include CXCR2, CCR3, CCR4, and CCR8. CCR3 is present on eosinophils, basophils, mast cells, and some Th2 T cells. CCR4 and CCR8 identify airway T cells of allergen-challenged atopic asthmatics. Despite extensive preclinical proof-of-principle, clinical trials of 3 CCR3 antagonists, 2 CCR4 antagonists, and 1 CXCR1/2 antagonist in asthma have been disappointing.

Eosinophilic esophagitis has been associated with a CCL26 variant. Although there is no CCL26 homologue in mouse, other mouse CCR3 ligands have been implicated in a mouse model of this disease.

Cancer

Many chemokines and leukocyte subtypes have been detected in situ in tumors, and cancer cells have been shown to produce chemokines and express chemokine receptors.[32] Despite extensive evidence for functional roles in mouse cancer models, the role played by endogenous tumor-associated chemokines in recruiting tumor-infiltrating lymphocytes and tumor-associated macrophages, and in promoting an antitumor immune response, has not been clearly delineated in humans. On the contrary, there are data from mouse models suggesting that the overall effect may be to promote tumorigenesis through additional effects on cell growth, angiogenesis, apoptosis, immune evasion, and metastasis. Controlling the balance of angiogenic and angiostatic chemokines may be particularly important. Chemokine receptors on tumor cells have been shown to directly mediate chemokine-dependent proliferation.

The most extensively studied chemokine receptor in cancer is CXCR4,[33] which is expressed ectopically or normally by multiple cancer types and has been implicated in multiple functions, including proliferation, anchorage, quiescence, and radiation and chemotherapy resistance of leukemic stem cells in bone marrow and breast cancer metastasis. It has been used successfully as a Cu^{64}-AMD3100 positron emission tomography (PET) imaging target for adrenocortical carcinoma metastasis in a patient. Acquired *CXCR4* mutations, including several gain-of-function mutations that cause WHIM syndrome, have been found in most patients with the plasma cell cancer Waldenstrom macroglobulinemia (Chapter 79), and have been associated with poor prognosis and poor treatment responses. A large number of clinical trials are testing CXCR4 inhibitors in various cancers, and seven have been advanced to phase 3.

CCR4 and CCR10 have been associated with several cancers, including melanoma. A humanized afucosylated cell-depleting anti-CCR4 monoclonal antibody named mogamulizumab-kpkc

(Poteligeo; Kyowa-Kirin) was FDA-approved in 2018 for the treatment of treatment-experienced refractory adult T-cell leukemia/lymphoma (Sezary syndrome and mycosis fungoides), which express CCR4 at high levels.

THERAPEUTIC APPLICATIONS

Chemokines and Chemokine Receptors as Targets for Drug Development

Chemokine receptors are the first cytokine receptors for which potent, selective nonpeptide small-molecule antagonists have been identified that work in vivo, including antagonists at CXCR2, CXCR3, CXCR4, CCR1, CCR2, CCR3, CCR5, and CCR9. Many of the compounds share a nitrogen-rich core and may block ligand binding by acting at either orthosteric or allosteric sites. As of 2020, approximately 700 clinical trials of drugs targeting the chemokine system have been registered and/or published, approximately `32 targeting chemokines, the rest targeting chemokine GPCRs. Approximately 70% of the trials targeted just three receptors, CCR4, CCR5, and CXCR4, the only three for which targeted drugs have received FDA approval: maraviroc targeting CCR5 in HIV/AIDS; plerixafor targeting CXCR4 in HSC mobilization; and mogamulizumab-kpkc targeting CCR4 in adult T-cell leukemia/lymphoma. Sixty-five of the approximately 700 trials have been advanced to phase 3, of which 59 were for the three approved drugs. Thus, in contrast to these three niche areas, the enormous effort expended toward targeting the chemokine system one component at a time in acute and chronic disease states has not yet resulted in approved drugs, and many failures have been reported. The fact that viral and tic antichemokines typically block multiple chemokines acting at multiple receptors hints that the most clinically effective chemokine-targeted antiinflammatory strategy may need to provide broad-spectrum coverage or define the key timepoint when intervention is most likely to be successful.[34]

Although small molecules taken as pills are the main goal, other blocking strategies are also under consideration, such as (1) ribozymes, (2) modified chemokines (e.g., amino-terminally-modified versions of CCL5), (3) intrakines, which are modified forms of chemokines delivered by gene therapy that remain in the endoplasmic reticulum and block surface expression of newly synthesized receptors, (4) monoclonal antibodies and llama nanobodies,[35] and (5) specific gene editing using zing finger nucleases, TALENs, and CRISPR/CAS9.

Chemokines as Biological Response Modifiers

Both inflammatory and homeostatic chemokines are being evaluated for therapeutic potential as biological response modifiers, acting mainly as immunomodulators or as regulators of angiogenesis. Studies to date have not revealed major problems with toxicity, and efficacy has been noted in models of cancer, inflammation, and infection. Clinical trials in cancer and stem cell protection have been disappointing. Chemokines are also being developed as vaccine adjuvants. Chemokine gene administration has also been shown to induce neutralizing antibody against the encoded chemokine, which is able to block immune responses and to ameliorate EAE and arthritis in rodent models.

Many chemokines, when delivered pharmacologically as recombinant proteins or by plasmid DNA or in transfected tumor cells, are able to induce immunologically mediated antitumor effects in mouse models and could be clinically useful. Chemokines could also be useful as adjuvants in tumor antigen vaccines and to potentiate leukocyte trafficking to tumor and immune responses that potentiate immune checkpoint inhibitor therapy. Chemokine-tumor antigen fusion proteins represent a novel twist on this approach that facilitates uptake of tumor antigens by antigen-presenting cells via the normal process of ligand–receptor internalization. Non-ELR CXC chemokines such as CXCL4 also exert antitumor effects through angiostatic mechanisms. Lastly, one can conceive of using chemokine-based immunotoxins to target disease cell–associated chemokine receptors, including cells infected with viruses such as human cytomegalovirus (HCMV) that encode chemokine receptors.

CONCLUSION

The chemokine system occupies a central place in immunoregulation and is an attractive source of potential drug targets for any disease with an innate or adaptive immune component. A basic outline for how the system works has been established using mouse models, and there has now been tangible progress translating this knowledge to the clinic. Three major successes, CCR5 inhibition in HIV/AIDS, CXCR4 inhibition for HSC mobilization, and CCR4 targeting in adult T-cell leukemia/lymphoma, are pioneers demonstrating the feasibility of targeting the chemokine system therapeutically in humans. However, all are eccentric indications that address niche roles of chemokine receptors. Chronic immune-mediated diseases remain a major unmet medical need, where the chemokine system offers many targets and difficult challenges for the future that may require targeting of multiple chemokine receptors simultaneously.

ACKNOWLEDGMENTS

This review was supported with funding from the Division of Intramural Research, National Institute of Allergy and Infectious Diseases, NIH.

REFERENCES

1. Bachelerie F, Ben-Baruch A, Burkhardt AM, et al. International Union of Basic and Clinical Pharmacology. [corrected]. LXXXIX. Update on the extended family of chemokine receptors and introducing a new nomenclature for atypical chemokine receptors. *Pharmacol Rev*. 2014;66(1):1–79.
2. Kufareva I, Salanga CL, Handel TM. Chemokine and chemokine receptor structure and interactions: implications for therapeutic strategies. *Immunol Cell Biol*. 2015;93(4):372–383.
3. Zlotnik A, Yoshie O, Nomiyama H. The chemokine and chemokine receptor superfamilies and their molecular evolution. *Genome Biol*. 2006;7(12):243.
4. Graham GJ, Handel TM, Proudfoot AEI. Leukocyte adhesion: reconceptualizing chemokine presentation by glycosaminoglycans. *Trends Immunol*. 2019;40(6):472–481.
5. Handel TM. The structure of a CXCR4:chemokine complex. *Front Immunol*. 2015;6:282.
6. Kufareva I, Gustavsson M, Zheng Y, et al. What do structures tell us about chemokine receptor function and antagonism? *Annu Rev Biophys*. 2017;46:175–198.
7. Jorgensen AS, Rosenkilde MM, Hjorto GM. Biased signaling of G protein-coupled receptors—From a chemokine receptor CCR7 perspective. *Gen Comp Endocrinol*. 2018;258:4–14.
8. Schall TJ, Proudfoot AE. Overcoming hurdles in developing successful drugs targeting chemokine receptors. *Nat Rev Immunol*. 2011;11(5):355–363.

9. Bachelerie F, Graham GJ, Locati M, et al. An atypical addition to the chemokine receptor nomenclature: IUPHAR Review 15. *Br J Pharmacol*. 2015;172(16):3945–3949.
10. Pontejo SM, Murphy PM, Pease JE. Chemokine subversion by human herpesviruses. *J Innate Immun*. 2018;10(5–6):465–478.
11. Legler DF, Thelen M. New insights in chemokine signaling. *F1000Res*. 2018;7:95.
12. Thelen M, Stein JV. How chemokines invite leukocytes to dance. *Nat Immunol*. 2008;9(9):953–959.
13. Proost P, Struyf S, Van Damme J, et al. Chemokine isoforms and processing in inflammation and immunity. *J Autoimmun*. 2017;85:45–57.
14. Butcher EC. Leukocyte-endothelial cell recognition: three (or more) steps to specificity and diversity. *Cell*. 1991;67(6):1033–1036.
15. Voisin MB, Nourshargh S. Neutrophil trafficking to lymphoid tissues: physiological and pathological implications. *J Pathol*. 2019;247(5):662–671.
16. Lee CH, Zhang HH, Singh SP, et al. C/EBPdelta drives interactions between human MAIT cells and endothelial cells that are important for extravasation. *Elife*. 2018;7
17. Bonavita O, Mollica Poeta V, Massara M, et al. Regulation of hematopoiesis by the chemokine system. *Cytokine*. 2018;109:76–80.
18. Murphy PM, Heusinkveld L. Multisystem multitasking by CXCL12 and its receptors CXCR4 and ACKR3. *Cytokine*. 2018;109:2–10.
19. Schulz O, Hammerschmidt SI, Moschovakis GL, Forster R. Chemokines and chemokine receptors in lymphoid tissue dynamics. *Annu Rev Immunol*. 2016;34:203–242.
20. Vandercappellen J, Van Damme J, Struyf S. The role of the CXC chemokines platelet factor-4 (CXCL4/PF-4) and its variant (CXCL4L1/PF-4var) in inflammation, angiogenesis and cancer. *Cytokine Growth Factor Rev*. 2011;22(1):1–18.
21. Stillie R, Farooq SM, Gordon JR, Stadnyk AW. The functional significance behind expressing two IL-8 receptor types on PMN. *J Leukoc Biol*. 2009;86(3):529–543.
22. Geissmann F, Mass E. A stratified myeloid system, the challenge of understanding macrophage diversity. *Semin Immunol*. 2015;27(6):353–356.
23. Lu E, Cyster JG. G-protein coupled receptors and ligands that organize humoral immune responses. *Immunol Rev*. 2019;289(1):158–172.
24. Rahimi RA, Luster AD. Chemokines: critical regulators of memory T cell development, maintenance, and function. *Adv Immunol*. 2018;138:71–98.
25. Berger EA, Murphy PM, Farber JM. Chemokine receptors as HIV-1 coreceptors: roles in viral entry, tropism, and disease. *Annu Rev Immunol*. 1999;17:657–700.
26. Hutter G, Bodor J, Ledger S, et al. CCR5 targeted cell therapy for HIV and prevention of viral escape. *Viruses*. 2015;7(8):4186–4203.
27. Zimmerman PA, Ferreira MU, Howes RE, Mercereau-Puijalon O. Red blood cell polymorphism and susceptibility to *Plasmodium vivax*. *Adv Parasitol*. 2013;81:27–76.
28. Heusinkveld LE, Majumdar S, Gao JL, et al. WHIM syndrome: from pathogenesis towards personalized medicine and cure. *J Clin Immunol*. 2019;39(6):532–556.
29. Noels H, Weber C, Koenen RR. Chemokines as therapeutic targets in cardiovascular disease. *Arterioscler Thromb Vasc Biol*. 2019;39(4):583–592.
30. Miyabe Y, Lian J, Miyabe C, Luster AD. Chemokines in rheumatic diseases: pathogenic role and therapeutic implications. *Nat Rev Rheumatol*. 2019;15(12):731–746.
31. Szekanecz Z, Koch AE. Successes and failures of chemokine-pathway targeting in rheumatoid arthritis. *Nat Rev Rheumatol*. 2016;12(1):5–13.
32. Allavena P, Germano G, Marchesi F, Mantovani A. Chemokines in cancer related inflammation. *Exp Cell Res*. 2011;317(5):664–673.
33. Scala S, D'Alterio C, Milanesi S, et al. New insights on the emerging genomic landscape of CXCR4 in cancer: a lesson from WHIM. *Vaccines (Basel)*. 2020;8(2):164.
34. Proudfoot AE, Bonvin P, Power CA. Targeting chemokines: pathogens can, why can't we? *Cytokine*. 2015;74(2):259–267.
35. Bobkov V, Arimont M, Zarca A, et al. Antibodies targeting chemokine receptors CXCR4 and ACKR3. *Mol Pharmacol*. 2019;96(6):753–764.

16

Lymphocyte Adhesion and Trafficking

Sirpa Jalkanen and Marko Salmi

EARLY LYMPHOCYTE PRECURSOR TRAFFICKING TO THE PRIMARY LYMPHOID ORGANS

Lymphocyte trafficking begins at an early stage of human ontogeny when lymphocyte precursor cells first appear and migrate into the primary lymphoid organs. The multipotent hematopoietic progenitor and stem cells from the yolk sac and from the aorta–gonad–mesonephros migrate via the circulation to the liver and the spleen, which are important organs that support lymphocyte production in the embryo (Chapter 2), and then into bone marrow.[1] Thereafter, the developmental maturation of B cells takes place solely in bone marrow (Chapter 7). T cells, in contrast, require an additional migratory event, which involves the entry of marrow-derived T-cell progenitors into the thymus (Chapter 9).[2] These early T-cell progenitors enter the thymus via the vessels in the cortical region. Concomitant with their differentiation and maturation via positive and negative selection, they pass from the cortex into the medulla.

MIGRATION OF NAIVE MATURE LYMPHOCYTES FROM BLOOD TO THE SECONDARY LYMPHOID ORGANS

After completing their initial course of development, newly arisen naïve B and T cells exit the primary lymphoid organs, travel through blood, and extravasate selectively to the secondary lymphoid organs.[3–5] These organs include peripheral lymph nodes, organized lymphoid tissues of the gut (e.g., Peyer patches and the appendix), and the spleen (Chapter 2). Interestingly, in mice lymphocyte homing to lymph nodes displays strong circadian regulation, with a peak at the onset of darkness.[6]

In lymph nodes, most lymphocyte trafficking from blood to tissues takes place in specialized postcapillary venules. The endothelial cells of the venule exhibit a characteristic high cuboidal morphology that has given them their name: high endothelial venules (HEVs).[4,7] The protrusion of the surface of these endothelial cells into the vascular lumen promotes the interaction of leukocytes with the endothelial surface membrane in the relatively low-shear venular part of the circulatory system (Fig. 16.1). HEVs carry many unique adhesion molecules that enable the capture of passing lymphocytes. They also have special intercellular connections that facilitate penetration of the vessel walls by these emigrating lymphocytes. It has been estimated that more than 50% of incoming lymphocytes make transient contacts with the vascular lining in the lymph nodes and that as many as one passing cell in four adheres to the endothelium and then extravasates into the tissue.

KEY CONCEPTS

Lymphocyte Recirculation

- Lymphocytes recirculate continuously between blood and lymphoid organs.
- Approximately 80% of lymphocytes enter the lymph nodes via specialized vessels called high endothelial venules (HEVs).
- The remaining lymphocytes enter the lymph nodes together with dendritic cells (DCs) and antigens via afferent lymphatics.
- Lymphocytes leave the lymph nodes via efferent lymphatics.
- Lymphocyte recirculation allows the lymphocytes to meet their cognate antigens and other leukocyte subsets to evoke an efficient immune response.

Antigens gather into these secondary lymphoid organs by a different route. Most antigens in the periphery can be taken up by dendritic cells (DCs) (Chapter 6), which subsequently migrate into the secondary lymphoid organs via the afferent lymphatics.[4] These afferent lymph channels open into the subcapsular sinus of the lymph node. Individual DCs subsequently penetrate the lymphatic endothelium and migrate into the stroma. Unbound, or free, antigens that are being carried via the afferent lymphatics through the body can diffuse into these secondary lymphoid organs, and then they can be captured by the professional antigen-presenting cells (APCs) of the lymph nodes.[8] Lymph nodes thus serve as traps for the immune system, collecting lymphocytes from blood and antigens from lymph (Fig. 16.2). In these organs, lymphocytes percolate through the tissue in search of their cognate antigens.[4,9] If a given lymphocyte does not find its antigen, it will leave the organ by entering the efferent lymphatics that drain the medullary sinuses and is then transported via a major lymphatic trunk, such as the thoracic duct, back into the large systemic veins. After reentering the circulation, the cell can randomly gain access to another lymph node, where it has another chance to extravasate into the tissue and find its cognate antigen. One round of recirculation from blood to the lymph node stroma, to a lymphatic vessel, and then back to blood takes about 1 day. Naïve lymphocytes continue recirculating until they either find their cognate antigen or die.

ACTIVATED LYMPHOCYTES DISPLAY SELECTIVE TISSUE HOMING PATTERNS

In a secondary lymphoid organ, a successful encounter between a lymphocyte and its cognate antigen leads to the proliferation of the cell within the organ and the maturation of its progeny.[4,5,9] Simultaneously, a process called *imprinting* leads to a profound

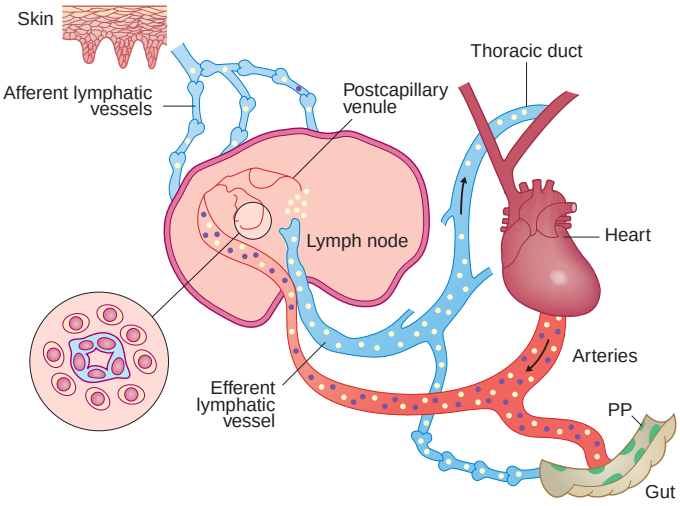


FIG. 16.1 Lymphocyte Recirculation Routes Under Physiological Conditions. A low level of continuous antigenic transport into lymphoid organs takes place via the afferent lymphatics draining the skin and epithelium of the gut. Blood-borne lymphocytes are carried to the organized lymphatic tissues (lymph nodes and Peyer patches *[PP]*) via the arterial tree, flow through the capillary bed, and then extravasate in the postcapillary high endothelial venules (HEVs). The extravasated lymphocytes percolate through the tissue parenchyma, enter the lymphatic vessels, and are then carried via the efferent lymphatics back to the systemic circulation. (Most of the venous circulation has been omitted from the figure.) *Inset:* a cross section of HEV (blue). (From Salmi M, Jalkanen S. How do lymphocytes know where to go: current concepts and enigmas of lymphocyte homing. *Adv Immunol.* 1997;64:139, with permission from Elsevier.)

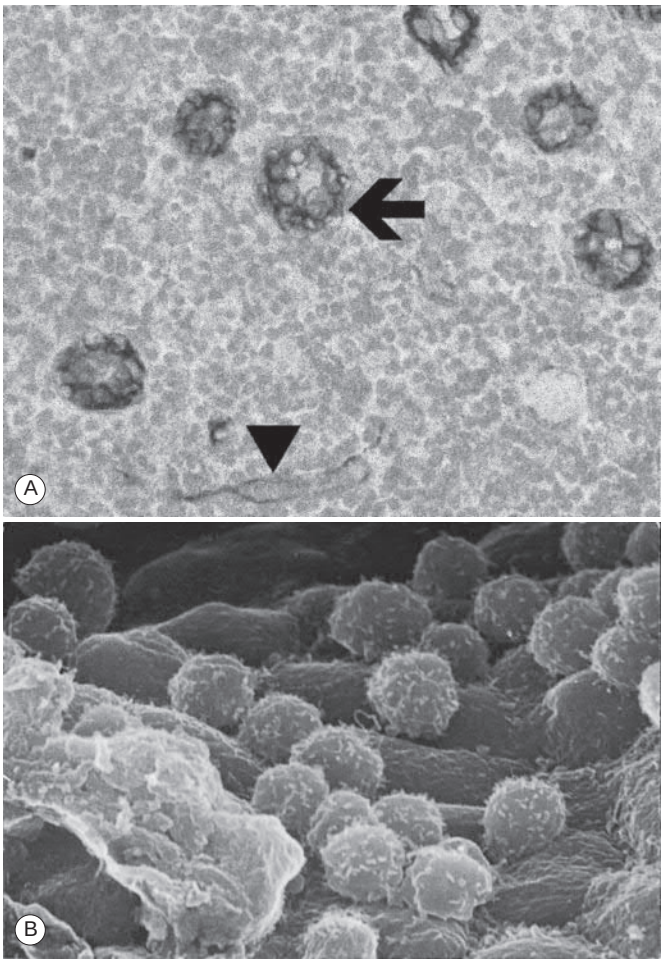


FIG. 16.2 Characteristic Features of High Endothelial Venules (HEVs). (A) In this immunoperoxidase staining with anti-CD31 antibody, six HEVs are seen with typical plump endothelial cells. One HEV is identified with an *arrow*; a vessel with flat endothelium is also seen in this figure *(arrowhead)*. (B) The scanning electron microscopy image shows lymphocytes adhering to the HEV.

change in the subsequent pattern of migration of the antigen-responsive cell. During the imprinting, local DCs give educational clues, such as vitamin A and D metabolites, to the lymphocytes. These lead to changes in the chemoattractant and adhesion receptor repertoire of the lymphocytes (Chapter 15).[10] Although these responding cells leave the node via the lymphatics and are carried back to the systemic circulation, unlike naïve cells, they no longer migrate randomly to any lymphoid tissue. Instead, imprinting primes cells to preferentially seek the peripheral tissues in which the inciting antigen was originally ingested by the DC. In this way, selective homing of lymphocytes according to their previous history allows the organism to focus the immune response to the tissues where the effector cells can do the most good.

Among activated T cells, distinct pools of short-lived T effector cells, long-lived central memory T cells, effector memory T cells, and tissue-resident memory T cells can be distinguished.[5,11] The different profile of adhesion molecule and chemokine receptor expression allows central memory T cells to continue migration through lymph nodes, whereas effector memory T cells are dispersed to patrol the peripheral tissues. In contrast, tissue-resident lymphocytes, including tissue-resident memory T cells, intraepithelial lymphocytes, γδ T cells, and innate lymphoid cells, mostly remain sessile in barrier tissues without the pattern of constant recirculation that is typical for the other lymphocyte subpopulations.

Under normal conditions, two distinct routes of lymphocyte recirculation can be discerned.[4,5] One targets lymphoid cells to the peripheral lymph nodes, and the second targets them to gut-associated lymphoreticular tissues (GALTs; Chapter 24). Although the common gut-associated lymphoreticular system has long been considered to include the intestinal, respiratory, and genitourinary tracts, differences in the fine specificity of lymphocyte homing among these targets do exist.

DISTINCT RECIRCULATION ROUTES IN THE SPLEEN

The spleen holds a unique place in the panoply of secondary lymphoid tissues.[12] It contains more lymphocytes than all of the peripheral lymph nodes combined, and the number of lymphocytes recirculating through it daily is the equivalent of the total pool of circulating lymphocytes. Based on the different anatomical structures of mouse and human spleen, the observations made in mouse are not directly applicable in man. For example, the human spleen does not have the marginal zone surrounding the white pulp, separating red pulp from the white pulp (Chapter 2). This is an important notion from the recirculation point of view, as the marginal zone has been considered to be a main entrance site

for lymphocytes in mouse. Recent studies, however, have demonstrated that lymphocytes can also enter the spleen via the red pulp vessels and thereafter, migrate within the *perivascular channels* to the white pulp in a chemokine receptor, CCR7-dependent manner, promoted by integrins leukocyte function–associated antigen-1 (LFA-1) and very late antigen 4 (VLA-4).

In the white pulp, T cells accumulate in the regions surrounding the central arteriole, a location known as the *periarteriolar sheath*. B cells are scattered in the corona that surrounds these T-cell areas. The spleen does not have afferent lymphatics, but efferent lymphatics may exist. The mechanisms that control the entry and exit of lymphocytes from the spleen remain incompletely understood but include chemokines, oxysterols, sphingosine-1-phosphate, and common lymphatic and vascular endothelial receptor-1 (Clever-1)/stabilin-1.

KEY CONCEPTS

Adhesion Molecules in Inflammation

- Adhesion molecules are important in directing leukocyte traffic to sites of inflammation.
- Numerous inflammatory mediators upregulate and/or induce expression of several endothelial cell adhesion molecules.
- Harmful inflammation can be prevented and cured by blocking the function of adhesion molecules.

INFLAMMATION-INDUCED CHANGES IN LEUKOCYTE TRAFFICKING

During an acute inflammatory response to an antigenic insult, leukocytes can migrate into all nonlymphoid sites. The inflammation-induced leukocyte migration takes place in characteristic waves.[13,14] First, the polymorphonuclear leukocytes rapidly (typically within 1 to 4 hours) infiltrate into the inflammatory focus. They are then followed by mononuclear cells (monocytes and lymphocytes). In a primary challenge, it may take 3 days or more before antigen-specific immunoblasts are seen at the peripheral site of inflammation. However, a secondary response by memory lymphocytes typically has a much shorter lag period. Different CD4 T-helper subpopulations, including Th1, Th2, Th17, and regulatory T cells (Tregs; Chapters 11 and 13), CD8 T-cytotoxic cells (Chapter 12), and B cells (Chapter 7) can all enter the inflamed tissue by using basically the same mechanisms. However, the ratio of these populations and the individual molecules employed can vary.[5,14] Successful resolution of an inflammatory reaction is also dependent on a coordinated program involving specialized proresolving lipid mediators (lipoxins, resolvins, protectins, and maresins), proteins (annexin A1), purines (adenosine), and gaseous mediators (e.g., hydrogen sulfide), all of which serve to halt the inflammatory cell recruitment and to initiate multiple antiinflammatory, tissue-repairing mechanisms.[15]

The normal vascular endothelium in nonlymphoid tissues has a flat, inactive morphology. With inflammation, a series of events renders postcapillary venules in these tissues capable of binding lymphocytes. The most important changes result from the proadhesive effects of a multitude of proinflammatory cytokines that are released by a variety of cell types after being subjected to inflammatory stimuli.[13] If inflammation becomes chronic, marked histological manifestations become apparent in the affected nonlymphoid tissues.[16] Most notably, the venules in these chronically inflamed tissues acquire many of the characteristics of HEVs. Immigrating lymphocytes can form lymphoid follicles that resemble those seen in lymph nodes. These alterations have consequences for lymphocyte recirculation pathways. For example, inflamed skin displays characteristics of lymphocyte homing that are clearly distinct from those of either mucosal or peripheral lymph node systems.[10,17]

MOLECULAR MECHANISMS INVOLVED IN LEUKOCYTE EXTRAVASATION FROM BLOOD INTO TISSUES

The Adhesion Cascade

Dynamic interactions between leukocytes and endothelial cells can be observed both in vitro and in vivo. For example, leukocyte adhesion can be followed in vivo in experimental animals and even in human tissues by intravital microscopy (Fig. 16.3). Leukocyte–endothelial cell interaction during the extravasation cascade can be divided into a series of phases, or steps, that all leukocyte subtypes, including lymphocytes, are thought to follow (Fig. 16.4).[4,5,13,18]

First, the leukocytes marginate out of the main bloodstream and begin to tether and roll on the endothelial cell surface. This step is mediated primarily by selectins and their mucin-like counterreceptors. This slow-velocity movement culminates in an activation phase, during which leukocyte chemokine receptors transmit activation signals by recognizing their chemokine ligands presented on the endothelial cell surface. This leads to avidity and/or affinity changes in leukocyte integrins, which bind leukocytes firmly to their immunoglobulin superfamily ligands on endothelial cells. Leukocytes then begin to crawl on the endothelium. After finding a proper place, they transmigrate through the endothelial cell layer. This transmigration process begins with interactions between leukocyte integrins, their counterreceptors, and other molecules. This step is followed by complex signaling events that lead to protein phosphorylation and dynamic clustering of the cytoskeleton.

Leukocytes typically transmigrate between endothelial cells at the endothelial cell junction (paracellular route), which opens transiently and subsequently closes by localized stimuli.[18] This process requires proteinases, such as matrix metalloproteinase-2 (MMP-2), which is induced in T cells upon adhesion to endothelial cells, as well as other repair mechanisms that are currently not well known. This subsequently closes the path of transmigration. Interestingly, leukocytes can also migrate through endothelial cells (transcellular route) in a subtype-specific fashion. For example, polymorphonuclear leukocytes prefer entering via the interendothelial junctions, whereas nonactivated lymphocytes may choose the transcellular route.[13,18]

KEY CONCEPTS

Leukocyte–Endothelial Cell Interactions

- Leukocytes interact with the vessel wall in a multistep fashion, using several leukocyte surface molecules that recognize their counterreceptors on endothelial cells.
- Selectins mediate the rolling and tethering of leukocytes on the vessel wall.
- Chemokines and their receptors activate leukocyte integrins.
- Only activated integrins are able to mediate firm adhesion between leukocytes and endothelium.
- The transmigration of leukocytes into the tissues requires proteinases and repair mechanisms.

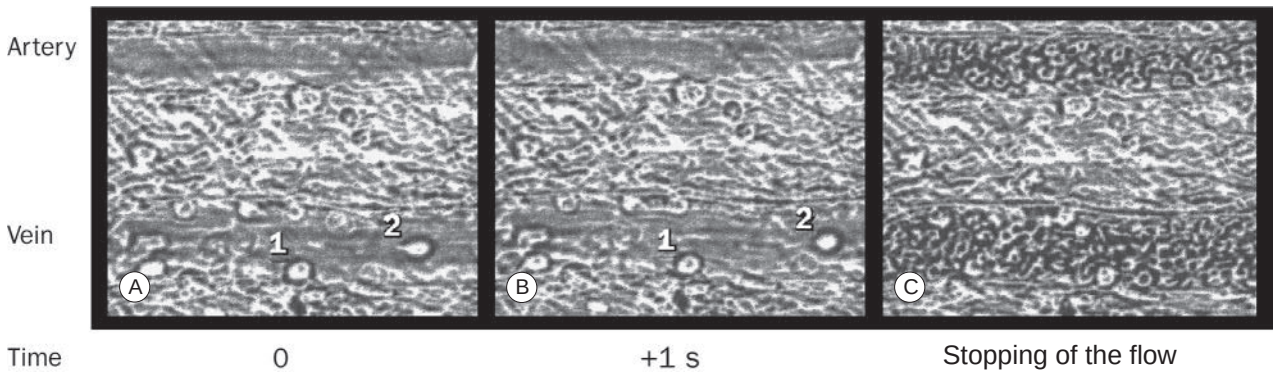


FIG. 16.3 Intravital Microscopy of Mesenteric Vessels. In these video frames taken at indicated intervals from the same field, a vein, an artery, leukocytes, and the transparent mesenteric membrane are seen. Within the vein, rolling and adherent leukocytes are visible, whereas no such cells are seen in the artery. Leukocyte *1* is attached to the vessel wall and leukocyte *2* is slowly rolling. Compare the locations of these cells in A and B to stationary leukocytes outside the vessels. Leukocyte 1 is at the same location in both panels, whereas leukocyte 2 has moved a distance corresponding to roughly the length of its diameter. Under normal conditions, freely flowing cells move so fast that they cannot be visualized. However, in panel C, the flow has been transiently stopped. Under this static condition, the large number of hematopoietic cells (mainly erythrocytes) traveling within the bloodstream can be seen. (Courtesy S. Tohka, University of Turku, Finland.)

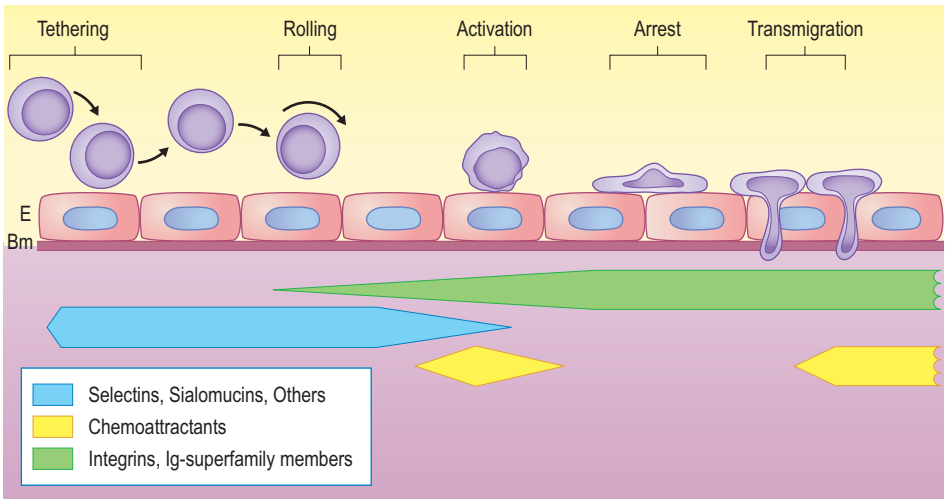


FIG. 16.4 The Multistep Cascade of Lymphocyte Extravasation. The blood-borne cell makes transient initial contacts with endothelial cells (tethering), which leads to the cell rolling along the vascular lining. If the cell becomes activated, it can subsequently adhere firmly to the endothelial cells (arrest). The adherent cell can then penetrate through the endothelial cell cytoplasm (transcellular migration) or seek for interendothelial junctions (crawling). It can then migrate between two endothelial cells (paracellular migration) and finally penetrate the basement membrane to enter the tissue. The contribution of major superfamilies of adhesion-associated molecules at each step is depicted below. *E*, Endothelial layer; *Bm*, basement membrane.

Certain endothelial molecules involved in the adhesion cascade show organ-specific expression patterns. Analogously, leukocyte-associated homing molecules display subtype-specific expression profiles. Only those leukocytes that have the proper set of molecules on their surface can enter a particular tissue because the entering leukocyte must find a correct endothelial partner molecule at each step of the adhesion cascade. Thus leukocyte–endothelial cell interactions take place in a well-coordinated multistep fashion, in which every step must be properly executed before the leukocyte can be guided into the tissue. The multistep nature of the leukocyte adhesion cascade is reminiscent of the cascades involved in blood clotting and complement-mediated killing.

RECEPTORS AND THEIR LIGANDS IN LEUKOCYTE–ENDOTHELIAL CELL INTERACTION

Various molecules belonging to several molecular families are expressed on both leukocyte and endothelial cell surfaces and participate in the complex extravasation process.[18] Most of these molecules exert their function in successive, but overlapping, phases of the adhesion cascade. In addition, proper functioning of multiple intracellular signaling molecules is needed for execution of successful extravasation.[4,18] In the following section, only the best-known surface molecules in this process are discussed (Fig. 16.5).

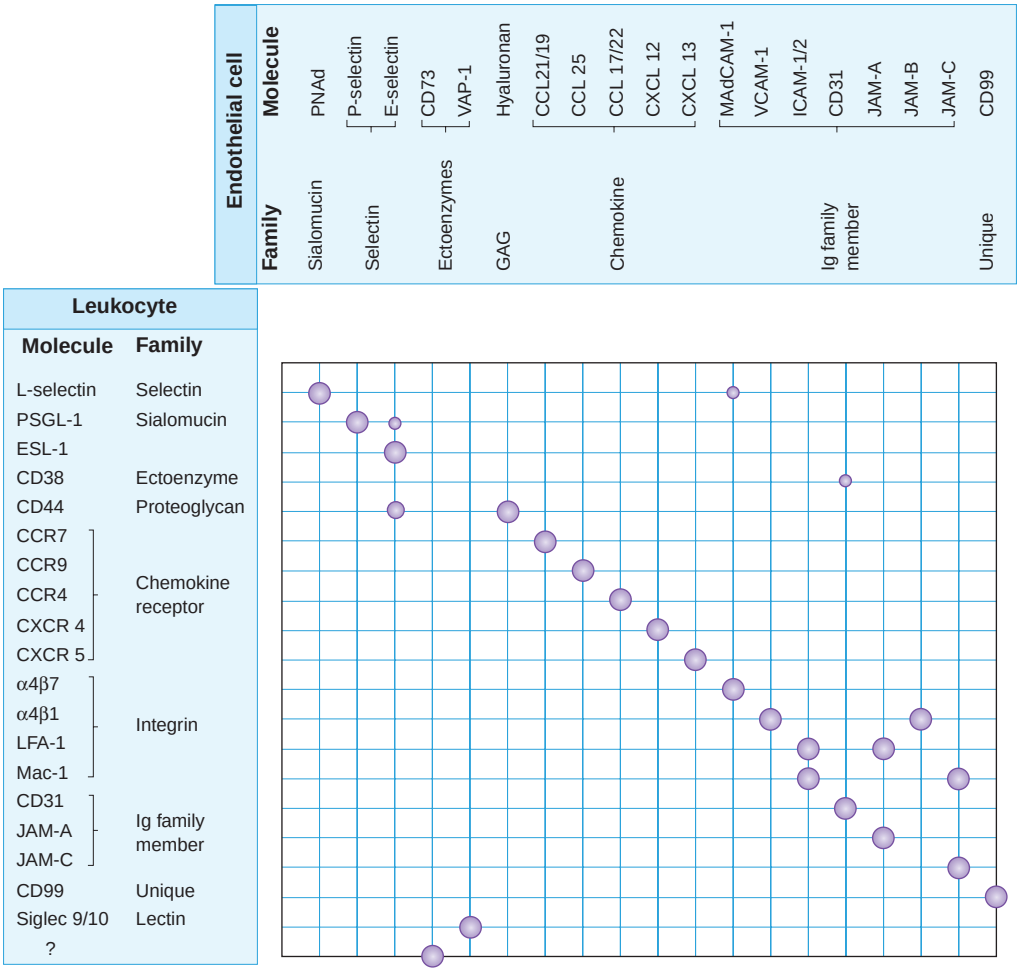


FIG. 16.5 Adhesion Molecules Mediating Leukocyte Traffic. The most relevant proteins involved in leukocyte–endothelial cell interactions are shown as receptor–ligand pairs. *GAG*, Glycosaminoglycan.

Selectins and Their Ligands

Three members of the selectin family mediate leukocyte trafficking. L-selectin is expressed on several leukocyte subpopulations. E-selectin is expressed on the endothelium, and P-selectin is expressed on both platelets and the endothelium. An important structural feature of selectins is the presence of a terminal lectin domain that is used to bind to their counterreceptors. The counterreceptor is typically decorated by a sialyl Lewis X (sLeX) carbohydrate, which is a prototype recognition domain for selectins in general.[19] The interaction between selectins and their counterreceptors is transient and weak, which allows leukocytes to form and break contacts with the endothelium effectively during tethering and rolling under shear stress.

L-selectin preferentially mediates lymphocyte migration to the peripheral lymph nodes.[13,19] However, it also participates in the homing of lymphocytes to organized mucosa-associated lymphoid tissue (MALT; e.g., Peyer patches) (Chapter 24). L-selectin is also an important contributor to the process of leukocyte trafficking to sites of inflammation. Peripheral lymph node addressins (PNAds) are the best-characterized counterreceptors for L-selectin. This group consists of at least six different molecules that are decorated with a sulfated and fucosylated sLeX, which serves as a recognition motif for L-selectin. PNAds include glycosylation-dependent cell adhesion molecule-1 (GlyCAM-1), CD34, podocalyxin, endomucin, nepmucin, and mucosal addressin cell adhesion molecule-1 (MAdCAM-1). MAdCAM-1 is decorated with recognition epitopes for L-selectin on HEVs in the organized lymphoid areas of the gut, but not on the flat-walled vessels in the lamina propria.

The importance of posttranslational carbohydrate modifications for the function of selectin ligands has been well demonstrated in knockout mice deficient in fucosyl, glucosaminyl, galactosyl, sialyl, or sulfotransferases.[19] For example, mice with a targeted disruption of fucosyltransferase VII (Fuc-TVII) are unable to glycosylate L-selectin ligands appropriately. These mice exhibit severe defects in lymphocyte homing and in the extravasation of leukocytes to sites of inflammation. In contrast, core 2 β 1,6-*N*-acetylglucosaminyltransferase (C2 β GlcNAcT) knockout mice that have deficient glycosylation of their P- and E-selectin ligands demonstrate normal lymphocyte homing to lymph nodes, but impaired leukocyte trafficking to sites of inflammation. Interestingly, L-selectin also facilitates entry of leukocytes into tissues by mediating leukocyte tethering and rolling on endothelium-bound leukocytes by binding to P-selectin glycoprotein ligand-1 (PSGL-1).

E- and P-selectin are inflammation-inducible molecules.[13,19] Within minutes, P-selectin can be translocated from intracellular storage granules onto the endothelial cell surface, where it

binds to its leukocyte receptor, PSGL-1 (see Fig. 16.5). P-selectin and PSGL-1 mediate rolling at early time points during inflammation. Platelet P-selectin can facilitate lymphocyte entry into tissues because it can simultaneously bind to PNAd on the endothelium and PSGL-1 on lymphocytes.

E-selectin upregulation requires new protein synthesis. It is maximally expressed about 4 hours after induction of inflammation. E-selectin is needed for slow rolling of leukocytes. It also has an affinity toward PSGL-1, as well as a specific glycoform of PSGL-1, cutaneous lymphocyte antigen (CLA). CLA specifically directs lymphocyte trafficking to inflamed skin.[17] In addition to PSGL-1, E-selectin has other leukocyte counterreceptors such as CD44 and E-selectin ligand-1 (ESL-1) (see Fig. 16.5).

Mice deficient in both E- and P-selectins have more drastic defects in their rolling and leukocyte migration to sites of inflammation than could have been anticipated from mice deficient in only one of the two selectins. This suggests that E- and P-selectins can compensate each other's functions.

Chemokines and Their Receptors

Chemotactic cytokines and their receptors (Chapter 15) are grouped into four different families based on their primary protein structure. The families are defined by a cysteine signature motif—CXC, CC, C, and CX3C—where C is a cysteine, and X any amino acid residue.[20] Most chemokines are small, soluble heparin-binding chemoattractants.

The relevant chemokines for leukocyte extravasation are presented to blood-borne lymphocytes by proteoglycan molecules present on endothelial cell surfaces. During the adhesion cascade, chemokines activate leukocyte integrins by signaling via serpentine receptors, which are pertussis-toxin sensitive and G-protein linked.

Different leukocyte subsets bear their own distinct sets of receptors, which enable them to respond to chemokines presented on the vascular endothelium as well as within tissues.[13,20] For example, CCL21 and CCL19 are preferentially expressed by HEVs that are found in the interfollicular areas within a lymph node. They preferentially attract T cells bearing the CCR7 receptor and thus draw T lymphocytes from blood and into these areas. Fractalkine, a CX3CL1 chemokine, can be produced in either a soluble or a membrane-anchored form. It can be found on the HEVs in peripheral lymph nodes and has potent chemoattractant activity for T cells. The major attractants for B cells are CXCL12 and CXCL13.

Although many chemokines are present in different organs in the body, some selectivity in their expression can guide tissue-selective leukocyte trafficking.[20] For example, CCL25 attracts CCR9-positive lymphocytes to the small intestine, and CCL17 and CCL22 attract CCR4-bearing lymphocytes to the skin. Chemokines can form heteromeric complexes to activate chemokine receptors. For example, CXCL13 can enhance triggering of CCR7 by CCL19 and CCL21. Conversely, formation of chemokine complexes may protect a chemokine from degradation.

Integrins and Their Immunoglobulin Superfamily Ligands

Integrins are a large family of heterodimeric molecules consisting of an α and a β chain.[4,21] Traditionally, they are thought to mediate firm adhesion between leukocytes and endothelial cells. However, in specific low-shear conditions, they can also participate in rolling. For leukocyte trafficking, the most important integrins are $\alpha_4\beta_7$, LFA-1, and $\alpha_4\beta_1$.

$\alpha_4\beta_7$ is a principal homing receptor for lymphocyte trafficking to mucosa-associated lymphatic tissues.[21] It binds to MAdCAM-1 on HEVs in organized MALTs, such as Peyer's patch and the appendix, and to flat-walled venules in the lamina propria. α_4-integrin can also pair with a β_1 chain to form $\alpha_4\beta_1$ dimers, which are utilized by lymphocytes primarily in inflammatory conditions. It binds to vascular cell adhesion molecule-1 (VCAM-1) on the endothelium and has been proven central in mediating lymphocyte trafficking into the brain in multiple sclerosis (MS).

LFA-1 (CD11a/CD18) is a member of the group of leukocyte integrins that contain a unique α chain (CD11 a, b, c, or d) but share a common β chain (β_2/CD18).[21] LFA-1 is present on practically all leukocyte subsets. It interacts with its counterreceptors, intercellular adhesion molecules (ICAM-1 and ICAM-2), or junctional adhesion molecule A (JAM-A) on the endothelial cell surface (see Fig. 16.5). ICAM-1 is upregulated at sites of inflammation, whereas ICAM-2 is constitutively present on the vascular endothelium.

To be functional, LFA-1 must be activated. Activation of LFA-1 is thought to be primarily a product of chemokine signaling. Alternative activation pathways include triggering through glycosyl-phosphatidylinositol (GPI)–linked molecules and CD44, as well as through other costimulatory lymphocyte surface molecules.[22] LFA-1–dependent pathways display no significant organ specificity in their function.

Mac-1 (CD11b/CD18) is also involved in leukocyte migration, although its contribution is overshadowed by LFA-1. Like LFA-1, Mac-1 uses ICAM-1 and ICAM-2 as its ligands (see Fig. 16.5). Both VCAM-1 and ICAM-1 also play an important role in initiating transmigration, as they mediate the signals to endothelial cells that are needed to change their shape and other properties in order to allow leukocyte entrance.[22]

Other Homing-Associated Molecules

Several other molecules belonging to various molecular families also participate in the adhesion cascade. CD44 is a multifunctional proteoglycan found on a large variety of different cell types.[23] Using endothelial hyaluronan as its ligand, it mediates lymphocyte rolling. It can form bimolecular complexes with $\alpha_4\beta_1$ that strengthen leukocyte–endothelial cell interaction. In vivo inhibition studies using function-blocking antibodies indicate that CD44 plays an important role in directing lymphocyte trafficking to sites of inflammation (e.g., skin and joints).

CD31, a member of the immunoglobulin superfamily, is found on many subsets of lymphocytes as well as in the continuous endothelium of all vessel types.[13] It is expressed primarily at the intercellular junctions and is involved in a stimulus-specific manner in transmigration, especially through the endothelial basement membrane. Other molecules involved in the transmigration process are CD99 and junctional adhesion molecules (JAMs) A and C, which are expressed on both leukocytes and endothelial cells.[13,18] These molecules interact sequentially in a homotypic fashion during diapedesis. Endothelial JAM-A can also use LFA-1, JAM-C can utilize Mac-1, and JAM-B can use $\alpha_4\beta_1$ as leukocyte ligands.

Also some ectoenzymes contribute to the adhesion cascade.[24] Vascular adhesion protein-1 (VAP-1), CD73, and CD38 have well-established roles in leukocyte trafficking. Because of their enzymatic properties, they can rapidly modify adhesive

interactions and modulate the microenvironment. VAP-1 is a homodimeric sialoglycoprotein that, under conditions of inflammation, is rapidly translocated onto the endothelial cell surface. It mediates early phases of leukocyte interaction with the endothelium, as well as transmigration. Besides its adhesive function, it also possesses an amine oxidase activity that can produce potent immunomodulators, such as H_2O_2 and aldehyde, as end products.

CD73 is present both on a subpopulation of lymphocytes and on the endothelium.[24] It is an ecto-nucleotidase. The main product of its enzymatic activity in dephosphorylation of adenosine monophosphate (AMP) is adenosine, which is highly antiinflammatory in nature and important in maintaining vascular integrity. Endothelial CD73 may also have a counterreceptor on the lymphocyte surface because lymphocyte binding to the endothelium inhibits the enzymatic activity of CD73. This facilitates the extravasation process of the lymphocyte. The importance of CD73 in dampening inflammation is clearly seen in mice deficient in CD73, as they are highly susceptible to inflammatory injuries as a result of leaky vasculature.

CD38, an adenosine diphosphate (ADP)–ribosyl cyclase, is expressed on most lymphoid cells and can use CD31 as its endothelial cell ligand.[24] It regulates calcium fluxes and the sensitivity of leukocytes to chemokine signals via its enzymatic activity.

INTRAORGAN LYMPHOCYTE LOCALIZATION

Following its extravasation from the blood vessels, a lymphocyte needs to interact with several matrix molecules, such as fibronectin, laminin, and collagens. Adhesive interactions between a lymphocyte and the extracellular matrix molecules are largely mediated by β_1 integrins, although lymphocytes can also use CD44 to interact with fibronectin and collagens. The directional movement and the final localization of a lymphocyte within the tissues are controlled by chemokines. Modern two-photon imaging has provided detailed information about the kinetics and directionality of lymphocyte movement in living tissues.[9]

Besides directing the entry of T cells into tissues, CCL21 and CCL19 also determine the final destination of the T lymphocyte within lymph nodes, the spleen, and Peyer patches.[20] CCL19 and CCL21 produced by stromal cells in the T-cell areas within lymphoid tissues guide the T lymphocyte into the interfollicular space. In an analogous manner, CXCL13, the B cell–attracting chemokine-1, is produced by a subset of follicular DCs found in the secondary lymphoid organs. It attracts B cells possessing CXCR5 to the light zone of the follicles. CXCL12, in contrast, guides CXCR4-positive B cells to the dark zone (Fig. 16.6).

The correct localization of a lymphocyte and an APC within the lymph node is a prerequisite for an optimal immune response. For example, B-cell collaboration with T cells is ensured by the upregulation of CXCR5 expression on a subpopulation of T cells and the upregulation of CCR7 on certain B cells. This promotes the movement of these cells to the B- and T-cell zone boundary.

The importance of CCL21 and CCL19 in lymphocyte trafficking is demonstrated by a spontaneous mutant mouse strain, plt, which has reduced expression of these chemokines.[20] In these mice, both lymphocyte entry via HEVs and the organization of T-cell areas within lymph nodes are defective. This phenotype is recapitulated in CCR7-knockout mice. CCL21/CCL19 and CXCL12/CXCL13 are currently the best-known determinants of lymphocyte localization within lymphoid tissues, although several other chemokines are needed for optimal encounters among various cell populations to create a crisp immune response. In addition to chemokines, other attractants (e.g., oxysterols) guide lymphocytes within the nodes. For example, Epstein-Barr virus–induced genes in B cells (EBI2) have a strong migratory response to 7α,25-dihydroxycholesterol, promoting B-cell positioning to the lymphoid follicles.[25]

CELL TRAFFICKING WITHIN LYMPHATICS

Although leukocyte trafficking within the lymphatics is an essential part of the recirculation process and the immune response in general, the molecular mechanisms that regulate this tightly controlled and cell type–selective migration remain poorly understood.[26] It is known that CCL21 secreted by the lymphatic endothelium generates haptotactic gradients in the skin and attracts CCR7-positive lymphocytes (and activated DCs) into the afferent lymphatics. Most leukocytes enter the afferent lymphatics at the blind-ended terminal lymphatic capillaries through button-like interendothelial junctions. They are then carried together with the lymph fluid via the collecting lymphatics into the draining lymph nodes. The afferent lymphatics open into the subcapsular sinus of the lymph nodes (see Fig. 16.6). At this location CCR7 on DCs and lymphocytes again play critical role in directing leukocyte migration through the floor lymphatic endothelial cells of the subcapsular sinus.[8] Leukocyte entry into the lymph node may also take place at the medullary sinusoids. The transmigration of leukocytes from the sinus into the parenchyma through lymphatic endothelial cells in lymph nodes is notoriously independent of classical adhesion molecules.

Lymphocytes, but not DCs, exit from the lymph node via efferent lymphatics at the cortical sinuses and medulla.[4,8] Lymphatic endothelial cells at the exit sites produce and secrete high levels of a multifunctional lipid messenger sphingosine-1-phosphate (S1P). S1P receptor 1 (S1PR1), which is expressed on the exiting T and B lymphocytes, guides their entry into the efferent lymphatics.[27] The role of adhesion molecules at the exit step remains largely unknown, although a few reports have implied a role for ICAM, macrophage mannose receptor, and Clever-1/stabilin-1 in this process.[8]

CLINICAL IMPLICATIONS

Leukocyte trafficking plays a pivotal role in the pathogenesis of all infectious and inflammatory diseases. Trafficking is essential for mounting a proper immune response against an invading microbe. However, in many other cases, inappropriate leukocyte migration also causes tissue destruction. Here, we have chosen a few representative examples to illustrate some of the general principles.

Immunodeficiencies

Various genetic defects in leukocyte trafficking have been identified during the past two decades among a multitude of other immunodeficiencies.[28] Although they often primarily affect neutrophil functions, the defective proteins are usually also present in lymphocytes, and the diseases very faithfully reproduce the leukocyte trafficking defects observed in experimental models.

Leukocytes from patients with Wiskott-Aldrich syndrome (WAS) (Chapter 34) show reduced migration to the chemokines CCL2, CCL3, and CXCL12 as a result of a mutation in the

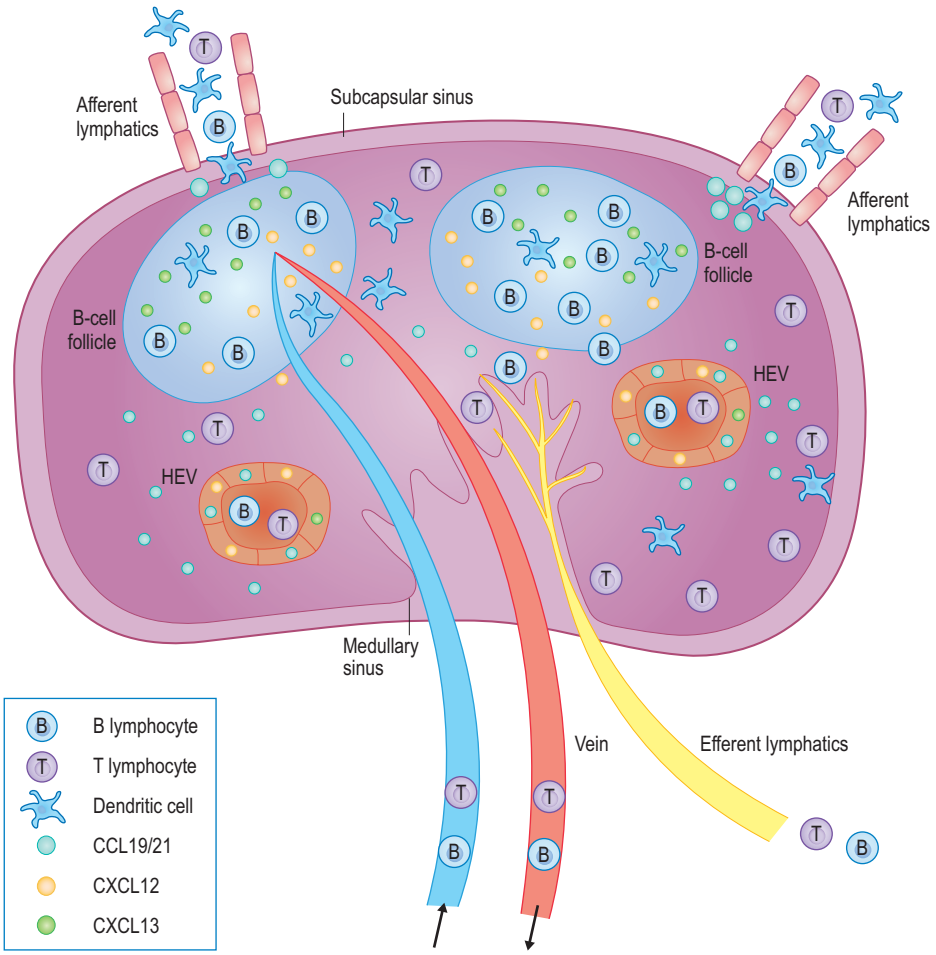


FIG. 16.6 The Role of Chemokines in Entrance and Intraorgan Localization of Lymphocytes. CCL19 and CCL21 are involved in lymphocyte entrance into the lymph node via the high endothelial venules *(HEVs)* and CCL21 also in entry of lymph-borne lymphocytes through the lymphatic sinuses. They also guide migration of T cells to the interfollicular areas within the node. In contrast, CXCL12 and CXCL13 attract B lymphocytes on the vessel wall and direct their movement to the follicles.

gene encoding an intracellular WAS protein that is responsible for proper cytoskeletal organization of the hematopoietic cells. Mutations of CXCR4 in patients with WHIM (warts, hypogammaglobulinemia, infections, and myelokathexis) lead to an enhanced chemotactic response to CXCL12 expressed on the bone marrow endothelium and leukocyte accumulation in bone marrow. A single case of inherited dysfunction of E-selectin has also been reported. The patient had recurrent infections but did not demonstrate neutrophilia. She could not synthesize E-selectin on the endothelium, although her serum level of soluble E-selectin had increased.[29]

Four different forms of leukocyte adhesion deficiency (LAD) syndromes have been described (Chapter 39).[30,31]

LADI: Patients with LADI have defects in the synthesis, pairing, or expression of β_2 integrins. The severe form leads to a drastic reduction or absence of CD11a/CD18 (LFA-1), CD11b/CD18 (Mac-1), CD11c/CD18, and CD11d/CD18. The patients manifest with defects in neutrophil arrest on endothelial cells. Patients suffering from a milder form of LADI (<10% of normal β_2 integrins) demonstrate impairment in leukocyte migration and suffer from frequent infections as well. In these patients, lymphocyte migration to inflammatory sites is close to normal, probably because lymphocytes can utilize the VLA-4–VCAM-1 pathway to compensate for the lack of β_2 integrins.

LADII: The cause of LADII is impaired transport of guanosine diphosphate (GDP)–fucose from the cytoplasm to the Golgi lumen as a result of mutations in a GDP–fucose transporter. The defect leads to impaired fucose modification of selectin ligands, most notably that of PSGL-1. Consequently, these patents demonstrate a marked decrease in leukocyte rolling under flow conditions, and, consequently, neutrophilia, recurrent infections, and impaired pus formation are typical in this disease.

LADIII: The third LAD syndrome involves defective inside-out signaling of β_1, β_2, and β_3 integrins due to mutations in FERMT3 (encoding Kindlin3), which is a major cytoplasmic activator of integrins. This results in impaired activation of multiple leukocyte integrins, including LFA-1 and VLA-4, by chemokine-triggered, G protein–coupled signals.

LADIV: The most recently identified LAD type is caused by defects in cystic fibrosis transmembrane conductance regulator (*CFTR*) gene. These relatively common mutations lead to defective activation of β_2 and $\alpha_4\beta_1$ integrins, and it appears to manifest especially with abnormal adhesion of monocytes to ICAM-1 and VCAM-1.

Autoimmune or Inflammatory Diseases

Multiple Sclerosis

Migration of T and B lymphocytes through the blood–brain barrier into the central nervous system (CNS) is a key pathogenic event that occurs during the development of MS and its widely used animal model, experimental allergic encephalomyelitis (EAE) (Chapter 66). Under normal conditions, lymphocytes only patrol outside of the CNS parenchyma in the leptomeningeal and perivascular compartments.[32] Naïve lymphocytes inefficiently bind to endothelial cells in meningeal microvessels (without a rolling step) using α_4 integrin and then extravasate through the fenestrated choroid vessels using mainly P-selectin and CCR6. Upon inflammation, meningeal and choroid vessels become more adhesive, and the endothelial cells within the CNS parenchyma start to express ICAM-1 and VCAM-1. These alterations allow effector lymphocytes, which have primarily been activated in the cervical lymph nodes that ultimately drain the CNS-derived antigens, to penetrate the blood–brain barrier and enter the CNS parenchyma.

In vivo animal studies have shown that the disease course can be dramatically altered by blocking the function of leukocyte α_4 integrin.[33] This treatment is able to prevent the disease, and even reverse the paralytic disease and lymphocytic infiltrations in the brain. Antibodies against $\alpha_4\beta_1$ integrin show this remarkable efficacy even if started a month after the onset of the clinical disease. Certain EAE models can also be blocked by targeting the α_4 integrin ligand VCAM-1 or by blocking the function of CD44. In contrast, inhibition of many other adhesion molecules, such as L-selectin, PSGL-1, and E- and P-selectins, fails to yield a consistent effect on EAE. It should also be noted that although $\alpha_4\beta_1$ integrin is very important for lymphocyte homing to brain, it is not a brain-specific homing molecule. It is also involved in leukocyte trafficking to other organs, such as the pancreas and the gut.

ON THE HORIZON

Adhesion Molecules as Targets for Novel Diagnostic and Therapeutic Agents

- Use of adhesion molecules as targets for the diagnosis of immune deficiencies, inflammatory disorders, and cancers.
- Use of adhesion molecules as targets for imaging inflammatory responses.
- Use of adhesion modulating therapies and other manipulations of lymphocyte trafficking, including small molecules, to treat infections, immune deficiencies, autoimmune disorders, transplant rejection, ischemia-reperfusion injuries, and cancers.

Inflammatory Bowel Diseases

As described above, effector lymphocytes activated in the inductive sites (Peyer patches and mesenteric lymph node) of the intestine selectively migrate to the effector sites of the gut (mainly the lamina propria) (Chapter 24).[34,35] The key tissue-specific adhesion code for the intestinal entry is $\alpha_4\beta_7$ integrin on lymphocytes and its endothelial counterreceptor MAdCAM-1. Although the canonical CCR9-CCL25 interaction is important for attracting T and B lymphocytes to the small intestine, the critical chemotactic receptors for the entry into the large intestine are GPR15 (on T cells except for Treg cells) and CCR10 (on plasmablasts).

In humans, the homing of regulatory T cells to the colon is GPR15 independent. This creates pathways by which the trafficking of proinflammatory versus antiinflammatory lymphocyte subtypes in the large bowel can be altered. In inflammatory bowel disease (IBD) (Chapter 75), the expression of MAdCAM-1 is upregulated in the vessels of the lamina propria, together with other vascular adhesins, such as VCAM-1 and ICAM-1. Increased expression of CCL20 in the IBD mucosa enhances the recruitment of both T and B cells to the inflamed gut through ligation of CCR6. Successful pharmacological and genetic targeting of these key gut-selective lymphocyte trafficking pathways alleviates inflammation in multiple animal models of IBD.

Cancer

Immunoevasion is one of the main hallmarks of cancer, and insufficient lymphocyte trafficking to solid tumors contributes to tumor progression. In the neoangiogenic vessels of tumors, immature and disorganized endothelial cells often express insufficient amounts of adhesion and attraction molecules.[36] Moreover, tumor-derived factors often render the local endothelial cells unresponsive to inflammatory cytokines. Nevertheless, chronic inflammation in tumors can induce the formation of HEV-like vessels, which show augmented expression of trafficking molecules. The homing patterns of leukocytes and those of cytotoxic T cells in particular, into the tumors (inflamed, excluded, or deserted), often associates with the outcome of the disease.[37]

Lymphocyte migration can be harnessed to improve the outcome of cellular immunotherapy for cancer.[38] Infusion of activated or genetically modified effector T cells, and cytotoxic CD8 T cells in particular, would benefit from a more effective targeting of the cells into the tumor. Homing of CAR-T cells to the tumors could be enhanced by transduction with adhesion molecules and chemokine receptors, possibly with simultaneous induction of tumor vessel normalization. Similarly, antiadhesive therapies targeting inappropriate accumulation of Treg in the tumor would also enhance anticancer immune responses.

THERAPEUTIC PRINCIPLES

- Proadhesive strategies have been mainly developed using gene therapy.
- Inappropriate inflammation associated with many diseases can be dampened by antiadhesive therapeutics.
- Function-blocking monoclonal antibodies (mAbs) are effective antiadhesive molecules.
- Cell migration can also be blocked by ligand and receptor analogues, small-molecule inhibitors, and genetic means (e.g., RNA interference).

ADHESION MOLECULES AS DIAGNOSTIC TARGETS

Immunodeficiency Disorders

Analyses of CD18 integrin expression, presence of fucosylated adhesion molecules (e.g., sLeX epitope), and the activation state of integrins are useful in diagnosis of different forms of LAD (Chapter 39).

Soluble Adhesion Molecules

Most adhesion molecules are found in soluble forms in body fluids. They can be generated by alternative splicing of messenger RNA (mRNA) with deletion of transmembrane anchors, by

proteolytic cleavage of cell-bound proteins by various sheddases (proteases), or by cleavage of the GPI linkage. Either of the soluble forms can function as molecular sinks, competing for their specific ligand(s) with the membrane-bound forms. Alternatively, they can trigger biological responses by interacting with their ligand-bearing cells.

The availability of commercial kits for measuring the levels of soluble adhesion molecules has led to numerous reports describing an increase or a decrease of certain adhesion molecules in different diseases. They may provide additional diagnostic or predictive value derived from the use of these tests compared with more traditional parameters of inflammatory activity. However, at present, the indications for routine measurements of soluble adhesion molecules, or chemokines, in inflammatory diseases and cancers remain to be defined.

Imaging

The use of neutrophil scans or radioactively labeled nonspecific molecules to localize inflammatory foci is not satisfactory in terms of timing, expense, biohazards, specificity, or sensitivity. Hence, there have been trials to radioactively or nonradioactively label monoclonal antibodies (mAbs) or peptides that recognize endothelial adhesion molecules, infuse them intravenously, and monitor their accumulation by appropriate imaging devices. Inflammation-inducible molecules, such as E-selectin, MAdCAM-1, and VAP-1, have been used as target antigens. In the case of radioactively labeled E-selectin antibodies and a VAP-1 binding peptide, the utility of this approach has been verified in patients.

CLINICAL PEARLS

- Blocking of α_4 integrin with natalizumab ameliorates disease activity in multiple sclerosis (MS) and in Crohn disease.
- Blocking of α_4/β_7 with vedolizumab is an effective treatment for inflammatory bowel disease (IBD).

THERAPEUTIC APPLICATIONS OF ADHESION-MODULATING THERAPIES

Pro- and antiadhesive therapies have long been an obvious pharmaceutical goal in the field of leukocyte trafficking. Inflammation-promoting strategies would be beneficial for treating many immune deficiencies, persistent infections, or cancers. Proadhesive control of lymphocyte traffic would benefit specific areas, such as vaccine development (Chapter 87) and bone marrow cell transplantation (Chapters 90 and 92). In practice, lymphocyte recirculation routes are already being empirically exploited, for example, by varying the anatomical site (skin vs. intestine) of vaccination to optimize the immunization response. Antiadhesive therapy, on the other hand, can be seen as a form of treatment applicable to all disease categories that involve an inflammatory component. In addition, it could provide novel precision drugs individually tailored for treating organ-specific inflammatory disorders, which would be predicted to diminish the problems of generalized immunosuppression.

Antibodies and Small-Molecular Drugs

A number of specific antagonists of adhesion molecules have been developed.[21,34,35] Function-blocking mAbs against adhesion molecules and chemokines have often been the drug candidates used for proof-of-principle experiments. In parallel, recombinant ligand or receptor analogs have been developed. Ultimately, knowledge of the structure of the adhesion molecules has allowed the design of rational, small-molecular drugs, such as those affecting the conformational state of leukocyte integrins. Possibilities to modulate mRNA expression of adhesion molecules through antisense oligonucleotides and RNA interference have added another potential tool to the pharmaceutical armamentarium.

Adhesion-Modulating Drugs in Clinical Use

Although several forms of antiadhesive therapy have been enormously successful in a panoply of animal models, transfer to the clinic has been slow. However, a small number of very potent drugs targeting adhesion molecules have already been approved.

The first selective adhesion molecule (SAM) inhibitor was natalizumab. The way for natalizumab, a humanized anti-α_4 integrin antibody, was paved by the excellent results of α_4 blocking in EAE (see above). In patients with relapsing MS, a monthly intravenous injection of natalizumab leads to a significant reduction in the numbers of new lesions, numbers of relapses, and the risk of sustained disability.[33] The beneficial effects of natalizumab treatment are evident also in patients who do not respond to interferon-β (IFN-β) therapy. Because natalizumab also inhibits $\alpha_4\beta_7$, the gut homing receptor, natalizumab also alleviates inflammation in the gut, and it has been shown to be effective in patients with Crohn disease.[34]

Another clinically approved selective modulator of an adhesion molecule is vedolizumab.[34,35] It is a humanized mAb that binds to a combinatorial epitope in α_4/β_7 integrin but does not interact with α_4/β_1 and α_E/β_7 integrins. This makes vedolizumab a remarkably selective inhibitor of lymphocyte homing to the intestine. It has been approved for the treatment of both ulcerative colitis and Crohn disease (Chapter 75). An mAb, etrolizumab, which targets the β_7 integrin (thus blocking both α_4/β_7 and α_E/β_7, but not α_4/β_1), which can induce remission in ulcerative colitis, and an anti-MAdCAM-1 antibody (ontamalimab) that appears to be effective in ulcerative colitis, but not in Crohn disease, are in phase III clinical trials.

These therapies also have adverse effects. The most feared adverse effect of natalizumab treatment is the possible reactivation of polyoma John Cunningham (JC) virus, which leads to the development of potentially fatal progressive multifocal leukoencephalopathy (PML).[39] Up to 1% of natalizumab-treated patients develop PML in the high-risk groups. However, the benefits of this therapy in MS still outweigh the risks (Chapter 66). With good risk stratification and vigilance, the drug has maintained its status as a very effective biological treatment option for MS. Theoretically, α_4/β_7 blocking should not interfere with lymphocyte surveillance functions in the CNS, and in line with this, no PML cases have been reported among vedolizumab-treated patients. Mild adverse effects with SAM therapy include an expected increase in susceptibility to infections and injection site reactions. However, apart from the risk of PML with natalizumab, both natalizumab and vedolizumab seem to have good overall safety profiles.

Lymphocyte migration can also be targeted therapeutically by modulating lymphocyte exit from the lymph nodes.[27] Fingolimod/FTY720 causes S1P receptor internalization and degradation. Consequently, it is a functional antagonist of SIP

receptors 1, 3, 4, and 5. It inhibits lymphocyte egress from the lymphoid organs. It may selectively retain CCR7-positive T cells, including central memory T cells, which may be particularly important for the pathogenesis of brain inflammation, while sparing CCR7-negative effector memory cells. It is used as the first-line treatment for relapsing MS, and it is the first orally active disease-modifying drug for this immunological disease. However, a more specific S1PR modulator, siponimod, targeting S1PR1 and S1PR5, was approved in 2019 by the FDA for the treatment of relapsing MS, including active secondary progressive MS.[40] Another selective S1PR1 antagonist, ozanimod, has shown preliminary efficacy, also in IBD. Although used in certain clinical indications, targeting of chemokine/chemokine receptors for manipulating lymphocyte migration has not yet proceeded to clinical use.

REFERENCES

1. Ratajczak MZ. A novel view of the adult bone marrow stem cell hierarchy and stem cell trafficking. *Leukemia*. 2015;29(4):776–782.
2. Zhang SL, Bhandoola A. Trafficking to the thymus. *Curr Top Microbiol Immunol*. 2014;373:87–111.
3. Gowans JL, Knight EJ. The route of re-circulation of lymphocytes in rat. *Proc Roy Soc Lond Ser B*. 1964;159:257–282.
4. Girard JP, Moussion C. Forster R. HEVs, lymphatics and homeostatic immune cell trafficking in lymph nodes. *Nat Rev Immunol*. 2012; 12(11):762–773.
5. Fu H, Ward EJ, Marelli-Berg FM. Mechanisms of T cell organotropism. *Cell Mol Life Sci*. 2016;73(16):3009–3033.
6. Druzd D, Matveeva O, Ince L, et al. Lymphocyte circadian clocks control lymph node trafficking and adaptive immune responses. *Immunity*. 2017;46(1):120–132.
7. Ager A. High endothelial venules and other blood vessels: critical regulators of lymphoid organ development and function. *Front Immunol*. 2017;8:45.
8. Jalkanen S, Salmi M. Lymphatic endothelial cells of the lymph node. *Nat Rev Immunol*. 2020;20:566–578.
9. Germain RN, Robey EA, Cahalan MD. A decade of imaging cellular motility and interaction dynamics in the immune system. *Science*. 2012;336(6089):1676–1681.
10. Sigmundsdottir H, Butcher EC. Environmental cues, dendritic cells and the programming of tissue-selective lymphocyte trafficking. *Nat Immunol*. 2008;9(9):981–987.
11. Fan X, Rudensky AY. Hallmarks of tissue-resident lymphocytes. *Cell*. 2016;164(6):1198–1211.
12. Lewis SM, Williams A, Eisenbarth SC. Structure and function of the immune system in the spleen. *Sci Immunol*. 2019;4(33):eaau6085.
13. Nourshargh S, Alon R. Leukocyte migration into inflamed tissues. *Immunity*. 2014;41(5):694–707.
14. Qi H, Kastenmüller W, Germain RN. Spatiotemporal basis of innate and adaptive immunity in secondary lymphoid tissue. *Annu Rev Cell Dev Biol*. 2014;30:141–167.
15. Headland SE, Norling LV. The resolution of inflammation: principles and challenges. *Semin Immunol*. 2015;27(3):149–160.
16. Buckley CD, Barone F, Nayar S, et al. Stromal cells in chronic inflammation and tertiary lymphoid organ formation. *Annu Rev Immunol*. 2015;33:715–745.
17. Ho AW, Kupper TS. T cells and the skin: from protective immunity to inflammatory skin disorders. *Nat Rev Immunol*. 2019;19(8):490–502.
18. Vestweber D. How leukocytes cross the vascular endothelium. *Nat Rev Immunol*. 2015;15(11):692–704.
19. McEver RP. Selectins: initiators of leucocyte adhesion and signalling at the vascular wall. *Cardiovasc Res*. 2015;107(3):331–339.
20. Eckert N, Permanyer M, Yu K, et al. Chemokines and other mediators in the development and functional organization of lymph nodes. *Immunol Rev*. 2019;289(1):62–83.
21. Ley K, Rivera-Nieves J, Sandborn WJ, et al. Integrin-based therapeutics: biological basis, clinical use and new drugs. *Nat Rev Drug Discov*. 2016;15(3):173–183.
22. Kinashi T. Overview of integrin signaling in the immune system. *Methods Mol Biol*. 2012;757:261–278.
23. Jordan AR, Racine RR, Hennig MJ, et al. The role of CD44 in disease pathophysiology and targeted treatment. *Front Immunol*. 2015;6:182.
24. Salmi M, Jalkanen S. Ectoenzymes in leukocyte migration and their therapeutic potential. *Semin Immunopathol*. 2014;36(2):163–176.
25. Cyster JG, Dang EV, Reboldi A, Yi T. 25-Hydroxycholesterols in innate and adaptive immunity. *Nat Rev Immunol*. 2014;14(11):731–743.
26. Randolph GJ, Ivanov S, Zinselmeyer BH, et al. The lymphatic system: integral roles in immunity. *Annu Rev Immunol*. 2017;35:31–52.
27. Cyster JG, Schwab SR. Sphingosine-1-phosphate and lymphocyte egress from lymphoid organs. *Annu Rev Immunol*. 2012;30:69–94.
28. Badolato R. Defects of leukocyte migration in primary immunodeficiencies. *Eur J Immunol*. 2013;43(6):1436–1440.
29. DeLisser HM, Christofidou-Solomidou M, Sun J, et al. Loss of endothelial surface expression of E-selectin in a patient with recurrent infections. *Blood*. 1999;94(3):884–894.
30. Das J, Sharma A, Jindal A, et al. Leukocyte adhesion defect: Where do we stand circa 2019? *Genes Dis*. 2020;7(1):107–114.
31. Hanna S, Etzioni A. Leukocyte adhesion deficiencies. *Ann N Y Acad Sci*. 2012;1250:50–55.
32. Dendrou CA, Fugger L, Friese MA. Immunopathology of multiple sclerosis. *Nat Rev Immunol*. 2015;15(9):545–558.
33. Steinman L. Immunology of relapse and remission in multiple sclerosis. *Annu Rev Immunol*. 2014;32:257–281.
34. Neurath MF. Targeting immune cell circuits and trafficking in inflammatory bowel disease. *Nat Immunol*. 2019;20(8):970–979.
35. Habtezion A, Nguyen LP, Hadeiba H, et al. Leukocyte trafficking to the small intestine and colon. *Gastroenterology*. 2016;150(2):340–354.
36. Peske JD, Woods AB, Engelhard VH. Control of CD8 T-cell infiltration into tumors by vasculature and microenvironment. *Adv Cancer Res*. 2015;128:263–307.
37. Chen DS, Mellman I. Elements of cancer immunity and the cancer-immune set point. *Nature*. 2017;541(7637):321–330.
38. Fucà G, Reppel L, Landoni E, et al. Enhancing chimeric antigen receptor T-cell efficacy in solid tumors. *Clin Cancer Res*. 2020;26(11):2444–2451.
39. Calabrese LH, Molloy E, Berger J. Sorting out the risks in progressive multifocal leukoencephalopathy. *Nat Rev Rheumatol*. 2015;11(2):119–123.
40. Derfuss T, Mehling M, Papadopoulou A, et al. Advances in oral immunomodulating therapies in relapsing multiple sclerosis. *Lancet Neurol*. 2020;19(4):336–347.

17

Death Pathways and Immunogenicity

Wulf Tonnus, Alexia Belavgeni, and Andreas Linkermann

For more than a century and a half, we have known that the process of regulated cell death (RCD) is an integral part of life. In 1972, the first insights into the mechanisms underlying RCD came from the identification of apoptotic pathways.[1] For almost 40 years, the apoptotic process of RCD was viewed as the opposite of necrosis, which was considered accidental (and therefore unregulated). In 2008, however, it was realized that necrosis could be antagonized under certain conditions. This led to the appreciation that necrosis could also be a genetically determined, regulated process.

The current challenge is to elucidate the various genes and pathways that take part in regulated physiological and pathophysiological cell death and to better understand why there are multiple pathways. It has become clear from these studies that the various pathways create differences in the immunogenicity of the byproducts of necrosis. This is especially the case when cellular necrosis is not limited to a single cell but affects a functional unit (*e.g.*, ferroptosis). As tissue injury and inflammation are tightly linked, these discoveries fuel the need to better understand how necrosis influences pathologies, including autoimmune diseases, transplant rejection, ischemic injury, and cancer. This chapter focuses on the most clearly identified regulated necrosis (RN) pathways. Our "selection" of the RN pathways discussed is based on the current appreciation of the importance of these pathways, either physiologically or pathophysiologically, to differences in the immunogenicity of these RN pathways and their role in diseases (Fig. 17.1).

CELL DEATH AND DAMAGE-ASSOCIATED MOLECULAR PATTERNS—THE CONCEPT OF NECROINFLAMMATION

Pattern recognition receptors (PRRs) bind a wide range of damage-associated molecular patterns (DAMPs) (Chapter 3). These include exogenous stimuli such as lipopolysaccharides (LPS) as well as an array of intracellular content.[2]

It is now understood that necrosis triggers inflammation and that inflammation can, in turn, lead to further RN. This observation gave rise to the concept of necroinflammation.[3] Necroinflammation can create an autoamplificatory feedback loop that results in DAMP release between organs. For example, ischemic or traumatic tissue in the lung can initiate a positive feedback loop that leads to acute respiratory distress syndrome (ARDS).[4,5] Recent observations indicate that the DAMPs released by this type of feedback after renal transplantation can lead to RN in the lung as well.[6]

KEY CONCEPTS

Necroinflammation—An Autoamplificatory Feed-Forward Loop

- Induced by cells dying by necrosis (*e.g.*, in hypoxia)
- Associated proinflammatory damage-associated molecular pattern (DAMP) release
- Immune cell infiltration
- Regulated necrosis (RN) in parenchymal and endothelial cells induced by immune cells
- RN in inflammatory cells (*e.g.*, macrophages undergoing pyroptosis)
- Surrounding cells die of necrosis
- DAMP transfer to remote organs
- Remote organ injury

For years, immunology focused on the issue of self versus non-self recognition. However, within the past two decades, the danger/injury model has become more and more prominent. Briefly, this model proposes that cellular stress activates the innate immune system, thus leading to an inflammatory response that can, under certain circumstances, then induce a specific adaptive response.

Key to this is the concept are DAMPs (Chapter 3). DAMPs include a wide array of different stimuli, including exogenous stimuli (*e.g.*, LPS) and intracellular contents (*e.g.*, uremic acid and high-mobility group box-1 [HMGB1]).[7,8] These are sensed by either "classic" PRRs or "nonclassic" receptors found on the surface or within cells of innate immunity. When cells die by RN, these DAMPs become accessible to the immune system in the extracellular space. They are sensed especially by cells of the innate immune system. Dendritic cells (DCs) (Chapter 6) are activated by directly sensing DAMPs via surface receptors (*e.g.*, calreticulin [CALR]–CD91, adenosine triphosphate [ATP]–P2X7R, or HMGB1–Toll-like receptor 4 [TLR4]). Sensing of DAMPs by monocytes leads to activation of inflammasomes, which then promote expression of mature inflammatory cytokines (*e.g.*, interleukin-1β [IL-1β] and IL-18) (Chapter 14). Natural killer (NK) cells (Chapter 12) also directly interact with injured cells. However, both monocytes and NK cells give rise to inflammation and thus support the maturation of immature DCs.

The infiltration of innate immune cells into the inflamed tissues and their subsequent inflammatory response is thought to account for a significant part of the overall damage to the organ, above and beyond initial necrosis itself. Mechanisms that mediate this deterioration may include the edema that arises following capillary leakage, induction of cell death by innate immune cells with the effect of decreased organ

function, and dysregulated partial oxygen pressure within the area of inflammation/necrosis. Also, mature DCs interact with CD8 T cells (Chapter 12) via major histocompatibility complex (MHC) class I interaction and CD4 T cells via MHC class II interaction (Chapter 5). This shifts the immune response to an antigen-specific response as cytotoxic T cells and active B cells now specifically attack stressed cells (Fig. 17.2).

In addition to these DAMPs, other intracellular content also has the potential to be sensed as a target, which could lead to the development of autoimmunity. A prominent example is the breaking of intracellular organelle membranes into pieces with exposed neoepitopes. These larger particles are removed by a process referred to as LC3-associated phagocytosis (LAP). Failure to remove these neoepitopes may result in the generation of antinuclear antibodies (ANAs) or antibodies against double-stranded DNA (dsDNA).

In mouse models, genetically induced breakdown of LAP drives a lupus-like phenotype.[9] ATG16L is required for both autophagy and LAP. In humans, ATG16L mutations are markedly associated with the development of autoimmunity.[10]

Usually, IL-10 is secreted by monocytes engulfing dying cells to dampen the immune response. However, LAP-deficient monocytes actively produce proinflammatory IL-1β and IL-6 instead. Orchestrated with other elevated proinflammatory cytokines (IL-17, IL-18, IL-23), a low immunogenic stimulus can thus induce a strong immune response. In the presence of this stimulus, common self-epitopes are classified as DAMPs by the immune system, which explains why mice deficient in LAP develop a lupus-like autoinflammatory disease. As a result, functional LAP may also be required for prevention of memory B-cell priming during solid-organ transplantation.

When challenged with injurious conditions, such as ischemia–reperfusion in solid-organ transplantations, options for cellular fate include directly succumbing via accidental cell death (ACD), restoring cellular homeostasis, or dying by RCD. ACD happens passively under certain conditions, such as extreme heat or mechanical trauma, and occurs nearly immediately in an uncontrolled fashion. Because of the direct loss of membrane integrity, huge amounts of DAMPs are released. This leads to massive recruitment of innate immunity and local

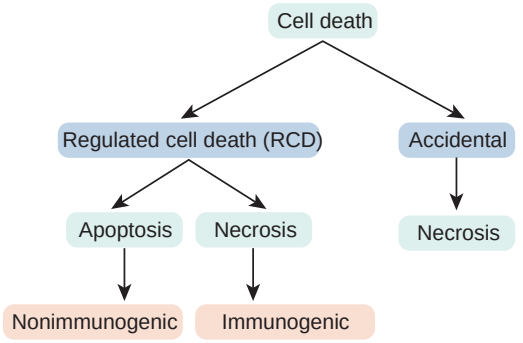


FIG. 17.1 Regulated Cell Death. When apoptosis was first identified, two models of regulated cell death were recognized: apoptosis and necrosis. More recently, however, cell death has been divided into regulated cell death (including apoptosis and regulated necrosis) and accidental cell death (instant necrosis). A second parameter classifies cell death as either immunogenic or nonimmunogenic.

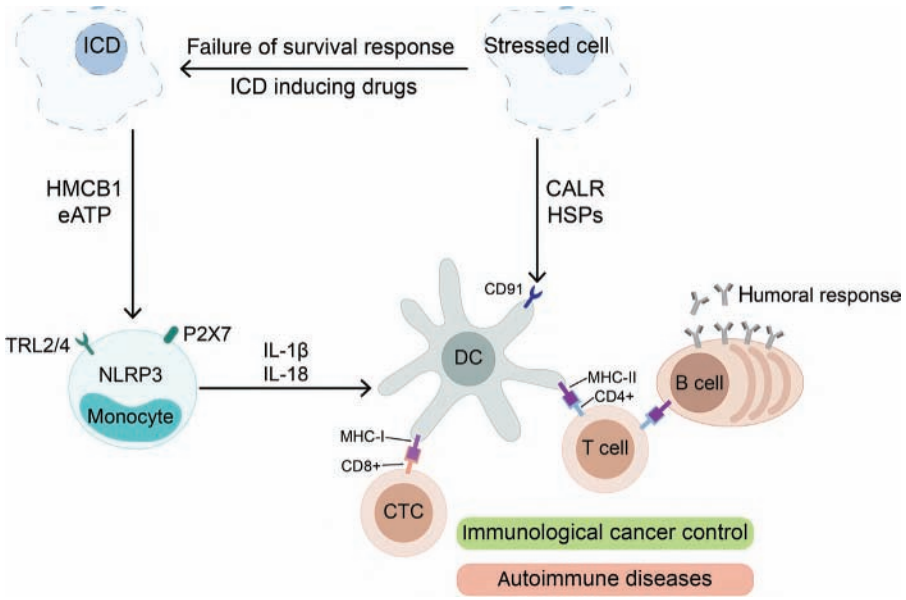


FIG. 17.2 Role of Dendritic Cells *(DCs)* in Necroinflammation. Innate immunity plays a crucial role in necroinflammation. When cells are stressed, they release damage-associated molecular patterns (DAMPs), such as calreticulin *(CALR)* and heat shock proteins *(HSPs)*. These DAMPs are sensed by scavenger receptors, such as CD91. This leads to partial activation of inactive DCs. If those cells fail to restore metabolic balance, they succumb to immunogenic cell death *(ICD)* and release additional DAMPs, which activate inflammasomes, such as LRR and PYD domains–containing protein 3 *(NLRP3)* in monocytes. Activation of the inflammasomes leads to the release of interleukins, which help DCs to gain full activation. These fully activated DCs use major histocompatibility complex *(MHC)* class I and class II antigen presentation to stimulate cytotoxic CD8 T cells *(CTC)* and CD4 T cells, which interact with B cells to trigger a humoral response. Activation of the immune response by necroinflammation can thus promote the control of cancer by immunogenic means. However, the same mechanisms can promote the development of autoimmune diseases.

inflammation. However, if these cells do not explode directly, they balance on the edge. On the one side is restoration to a normal state (*e.g.*, by means of autophagy or the unfolded protein response following endoplasmic reticulum [ER] stress) or, if the damage is too severe, loss of balance and the induction of RCD.

Within the cells that succumb to cell death, the necroinflammatory loop starts with perturbation of intracellular homeostasis (class V DAMPs). This triggers the heat shock response, a system of critical importance for correct protein folding. Secreted or surface-exposed heat shock proteins (HSPs) can be sensed by either classic (*e.g.*, TLR2/4) or nonclassic (*e.g.*, CD91) receptors on DCs. In parallel, ER stress, which is often the result of reactive oxygen species [ROS] generation, leads to secretion of calreticulin (an ER chaperone referred to as *CALR*). CALR then acts extracellularly as a class I DAMP by binding to CD91. This causes inactive dendritic cells (iDCs) to be activated. For full activation, however, an inflammatory milieu is required.

An inflammatory milieu is primarily generated by monocytes. These become active as the survival response fails and the formerly stressed cell succumbs to RN. Necroptosis and ferroptosis are typical modes of RN within this process. The ferroptotic cell can "leech" redox equivalents, such as nicotinamide adenine dinucleotide phosphate (NADP), from neighboring cells (see below), which subsequently undergo necrosis in a wave-of-death-like manner. This noncell autonomous induction of necrosis may involve necroptosis, although in this case the mechanism of necroptosis induction has not yet been elucidated. RN releases high amounts of DAMPs; thus in this context it can also be named *immunogenic cell death* (ICD).

As a DAMP, ATP can reach the extracellular space (eATP). When the plasma membrane is intact, the autophagy machinery seems to be necessary for ATP export. However, upon RN the membrane ruptures, eliminating the need to export ATP. eATP represents a class II DAMP as it activates the purinergic-receptor, ligand-gated ion channel P2X7R. Activation of P2X7R leads to potassium influx, which is sensed by the NACHT, LRR, and PYD domains-containing protein 3 (NLRP3) inflammasome and leads to IL-1β and IL-18 maturation. eATP is thus extremely immunogenic.

Upon membrane rupture, HMGB1 also reaches the extracellular space and is therefore the prototype of a RN-DAMP. iHMGB1 acts via binding to TLR4 of monocytes and also activates the NLRP3 inflammasome.

The DAMPs from the stressed cells and the inflammatory cytokines fully activate iDC, which then create a per-se cytotoxic inflammatory environment. They also prime naïve CD4 and CD8 T cells, thereby inducing an antigen-specific response. This response is the basis for certain conditions, such as chemotherapy in cancer. To stably control (or wipe out) a cancer, a specific response against tumor epitopes is required. Thus, induction of RN can be beneficial. In contrast, in some settings, such as solid-organ transplantation (Chapter 89), an antigen-specific response induced by DAMPs can give rise to antibody-mediated rejection (ABMR), which typically starts after cortisone tapering. In this setting, RN is detrimental.

The activation of an antigen-specific response results in additional cells being attacked. As they succumb, these dying cells replenish the DAMP pool, further attracting cells of innate immunity and further promoting DC maturation and T-cell priming. Thus, tissue injury amplifies while the loop closes (Fig. 17.3).

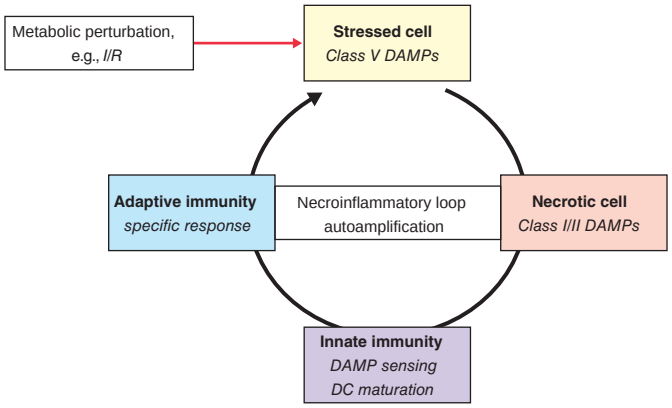


FIG. 17.3 The Autoinflammatory Loop. The necroinflammatory loop is triggered by metabolic perturbations (*e.g.*, ischemia–reperfusion *[I/R]*). The so-stressed cells release class V danger-associated molecular patterns *(DAMPs)*. In the next step, now-necrotic cells release class I/II DAMPs, which can be sensed by cells of innate immunity, such as dendritic cells *(DCs)*, which are then stimulated to mature. Via mechanisms described in *Fig. 17.2*, adaptive immunity is empowered to unleash a specific response to (neo)antigens. This leads to further stressed and dying cells, which, in turn, can themselves then release DAMPs and thus create an autoinflammatory loop.

REGULATED CELL DEATH REGULATES ITS IMMUNOGENICITY IN AN ACTIVE MANNER

Apoptosis is the prototype of non-ICD. Avoidance of immunogenicity is achieved by active processing and covering of DAMPs during the apoptosis program. For example, DNA is fragmented, cell organelles are consumed, and proteins degraded. Moreover, everything is packed in blebbing membranes. It is a matter of debate, however, whether apoptosis should be also considered somehow immunogenic, as some DAMPs (*e.g.*, HMGB1) are still released and phosphatidylserines, which are usually located in the inner leaflet of the plasma membrane, are flipped to the outer leaflet.

Phosphatidylserine surface expression serves as an "eat me" signal for macrophages and other phagocytes, which can then act to remove these cells in an immunologically silent manner. Thus, apoptotic cells recruit cells of innate immunity, but in a manner that is programmed to not induce further inflammation. Therefore, the debate about the inflammatory potential of apoptosis might be more or less of an academic nature. The sophisticated program of apoptosis appears to have evolved to prevent necroinflammation and is therefore favored in physiological settings that result in regular cell turnover.

Under metabolic pressure, such as ischemia–reperfusion injury, tissue damage occurs primarily by RN, and thus, the necroinflammatory loop *per se* tends to refuel itself. Therefore, mechanisms are required to keep this in check. One such mechanism is the active production of IL-33 by cells undergoing necroptosis (see below), which acts to limit the immunogenicity to a certain microenvironment,[11] as IL-33 stabilizes regulatory T cells (Treg) through ST2 receptors.[12] Cells undergoing pyroptosis (see below) actively produce and secrete IL-1β and IL-18 upon their demise. Both of these ILs

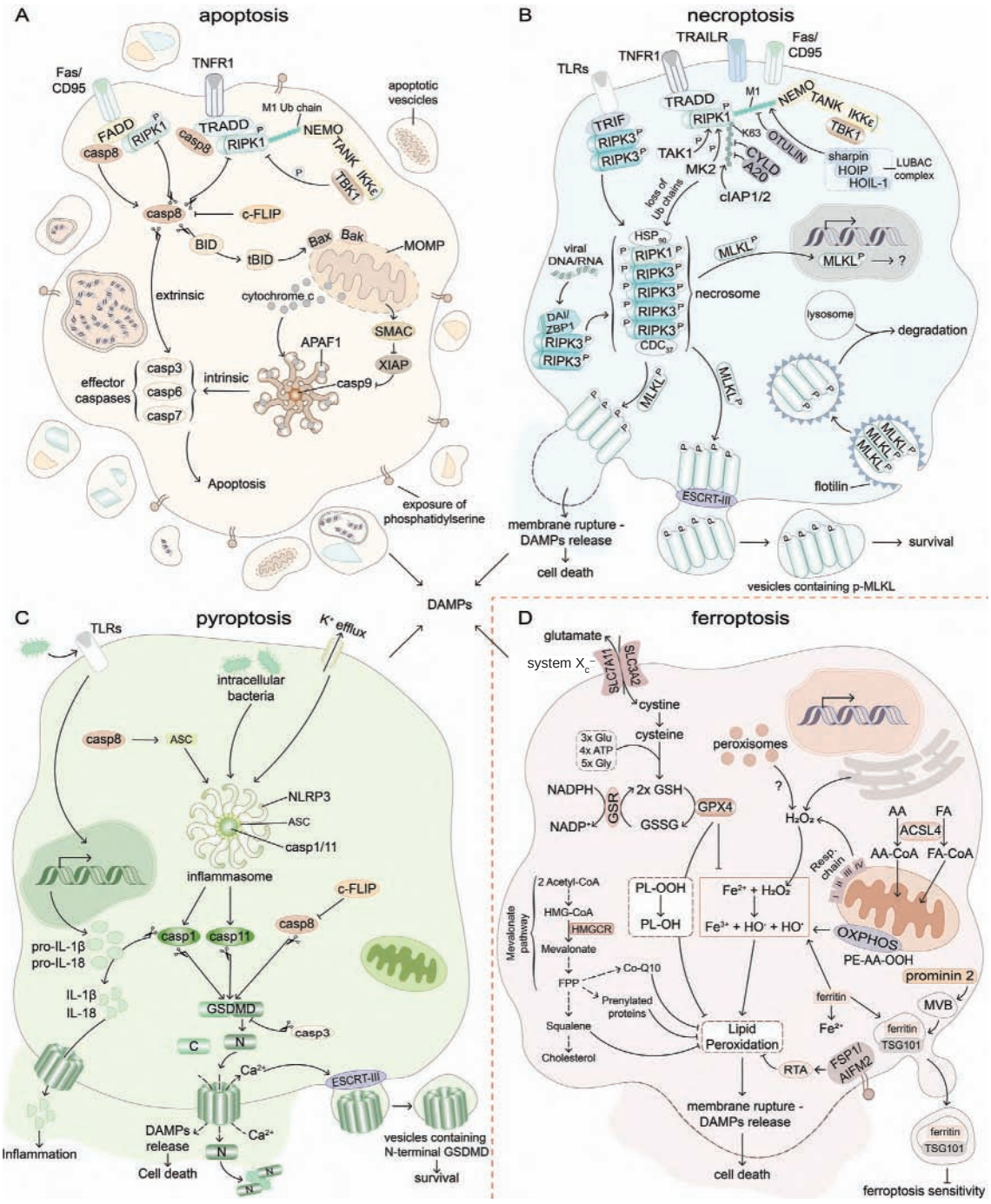


FIG. 17.4 Pathways of Regulated Cell Death *(RCD).* Apoptosis (A) represents a noninflammatory pathway that is mediated by caspases. Two distinct signaling pathways of apoptosis, extrinsic and intrinsic apoptosis, have been characterized. In extrinsic apoptosis, death receptors such as *TNF-R1*, *CD95* (Fas) and *TRAIL-R*, through the engagement of various intracellular adaptor proteins (*FADD, TRADD*) lead to the activation of the master regulator caspase-8. Upon homodimerization, caspase-8 cleaves the effector caspases-3/-6/-7 to propagate the apoptosis program. Upon loss of mitochondrial outer membrane potential *(MOMP)* and the *BAX-BAK*-mediated release of cytochrome c from the mitochondria into the cytosol, the intrinsic apoptotic pathway is triggered. Within the cytosol, cytochrome c, *APAF1*, and caspase-9 form the apoptosome, which activates the *effector caspases-3/-6/-7*. The typical morphology includes nuclear condensation, early loss of cellular volume (shrinking), subsequent membrane blebbing, and exposure of phosphatidylserine *(PS)*. Importantly, PS exposure represents an eat mesignal to macrophages that eliminate apoptotic cells. Importantly, the plasma membrane does not lose its integrity during this process. While apoptosis depends on the activation of caspases, necroptosis (B) is mediated by kinases. Dependent on its RHIM domain, *RIPK3* forms an amyloid-like structure referred to as the *necrosome*, the central relay of necroptosis. Therein, RIPK3 phosphorylates the pseudokinase mixed lineage kinase domain-like protein *(MLKL)*. By unknown mechanisms, the phosphorylated form of *MLKL* (*p-MLKL*) triggers plasma membrane rupture, a process that was demonstrated to be counteracted by the membrane repair *ESCRT-III* complex. The necrosome can be engaged by death receptor signaling in condition, in which caspases are absent or inhibited (*e.g.*, by viral proteins), and *RIPK1* no longer intercalates with RIPK3. Other ways

are highly proinflammatory. As pyroptosis is typical for cells of innate immunity, this might be an alert function of this first-line defense.

In conclusion, even in RN, cells regulate their inflammatory potential in both proinflammatory and antiinflammatory ways. When balanced, this generates a beneficial environment for both defense and regeneration. The exception to this rule is ferroptosis, where lipid peroxides recruit neutrophils (see below), and it is still unclear whether this response can be considered beneficial.[13]

SIGNALING PATHWAYS OF REGULATED CELL DEATH

RCD is an umbrella term for any genetically determined signaling pathway that results in cellular demise.[14] RCD thus includes both nonimmunogenic apoptosis and immunogenic RN. Fig. 17.4 provides an overview of the four most important pathways of RCD, comprising apoptosis, pyroptosis, necroptosis, and ferroptosis. Whereas the first three of these pathways are controlled by an intertwined caspase/kinase system, ferroptosis appears to be somewhat different in nature.

Apoptosis

Apoptosis is a complex program for nonimmunogenic cellular demise that is regulated by caspases. During the first hours of apoptosis, cells maintain their plasma membrane integrity and therefore, by definition, are not necrotic. The apoptotic program may be activated via the intrinsic (mitochondrial) or the extrinsic (death receptor) pathways.

The intrinsic pathway is controlled by different members of the Bcl-2 protein family, such as Bcl-2 or BAX, and therefore reacts to intracellular changes (*e.g.*, DNA damage) by forming transient pores, a process referred to as mitochondrial outer membrane permeabilization (MOMP). The intrinsic pathway reacts to internal stimuli, whereas the default activation of the extrinsic pathway occurs following an outside-in signaling. The latter is typically mediated by death receptors, such as Fas (also known as CD95 or Apo1) or tumor necrosis factor receptor 1 (TNFR1). Upon trimerization/hexamerization, these death receptors recruit downstream molecules via death domains (DDs), such as TRADD or FADD, to form a receptor-associated platform, the death-inducing signaling complex (DISC).

Nuclear factor-κB (NF-κB) signaling is the canonical response to TNFR1 activation. If NF-κB is inhibited and/or receptor-interacting protein kinase 1 (RIPK1) polyubiquitination is lost, the DISC complex is capable of activating initiator caspases (*e.g.*, caspase-8 [CASP-8] or caspase-10 [CASP-10]). Together with RIPK1, RIPK3, and FADD, a CASP-8–cFLIP heterodimer forms a cellular signaling platform, termed a ripoptosome, which usually prevents necroptotic signaling (see below) by cleaving RIPK1, RIPK3, and cylindromatosis (CYLD) while not proteolytically activating downstream effector caspases such as CASP-3, CASP-6, and CASP-7. However, should CASP-8 be activated, it will form homodimers to activate downstream caspases, which induce apoptosis.

RIPK1 polyubiquitination represents a major checkpoint for a cellular decision to either survival or to undergo RCD. For this reason, polyubiquitination is tightly regulated. Inhibitors of apoptosis 1 and 2 (cIAP1/2) and the linear ubiquitination complex (LUBAC) attach ubiquitin chains, whereas OTULIN, CYLD, and A20 remove them. Polyubiquitination is required for NEMO-dependent canonical NF-κB signaling and mitogen-activated protein kinase (MAPK) activation. Both provide a survival signal. In contrast, on loss of polyubiquitination RCD is licensed. Remarkably, both pro-survival NF-κB signaling and necroptosis lead to local inflammation, whereas apoptosis is noninflammatory.

Following the observation that CASP-8-deficient mice (which die in utero) can be rescued on a RIPK3-deficient background to become viable and fertile, the physiological role of apoptosis has been questioned. However, CASP-8 is of critical importance only for the extrinsic pathway of apoptosis. Mice deficient for both BAX and BAK, which cannot undergo intrinsic apoptosis, are not viable. Thus, apoptosis appears to be important for normal development and, particularly in the case of CASP-8, for the inhibition of necroptosis.

Other than hereditary autoimmune syndromes (*e.g.*, autoimmune lymphoproliferative syndrome [ALPS]), no clear role for apoptosis in diseases has been unequivocally reported. To date, inhibition of caspases in certain diseases has either worsened or failed to provide benefits to disease outcomes.

Pyroptosis

Pyroptosis (from "pyro" = fever/inflammation and "ptosis" = to fall) is a form of RCD induced by inflammatory stimuli, transduced via inflammatory caspases and (probably) executed by gasdermins. Inflammatory caspases include CASP-1, CASP-4/-5 (human), and CASP-11 (murine). Unlike apoptosis, caspase activation in this context leads to a necrotic phenotype that is mediated by members of the gasdermin family. In the prototype pyroptosis pathway, CASP-1/-11 cleave gasdermin D (GSDMD). Proteins of the gasdermin family are composed of a self-inhibitory C-terminal and a death-inducing

of engaging the necrosome are through Toll-like receptors (*TLRs*) via the RHIM-containing adaptor molecule *TRIF*, or by activation of the protein *ZBP1/DAI* in response to sensing intracellular oligonucleotides. Finally, and without any clear connection to apoptosis and necroptosis, ferroptosis (C) is a failsafe rather than a typical cell death pathway. In cellular homeostasis, H_2O_2 concentrations iron catalyzed and fenton reactions are limited by diverse cellular antiredox systems. The best-studied system relies on glutathione *(GSH)*, which is generated intracellularly and depends on supply via system X_{c-}, a cys/glu antiporter in the plasma membrane, or products of the mevalonate pathway. Upon sufficient GSH concentrations, glutathione peroxidase 4 (GPX4) prevents lipid peroxidation that otherwise leads to plasma membrane rupture by unknown mechanisms. In contrast, the oxidoreductase *FSP1* (also known as *AIF-M2*) prevents lipid peroxidation upon myristoylation-dependent recruitment to the plasma membrane in a GSH-independent manner. Finally, the targeted release of ferritin from multivesicular bodies into the extracellular space might represent another anti-ferroptosis mechanism that depends on the protein prominin2. Ferroptosis occurs as a non cell-autonomous pathway in a process referred to as synchronized regulated necrosis (SRN).

N-terminal fragment.[15] Upon caspase-mediated cleavage of GSDMD, the N-terminal fragment loses its self-inhibiting C-terminal fragment and therefore becomes active. This, through an as-yet to be defined mechanism, results in the for mation of a plasma membrane pore that allows secretion of active cytokines and necrosis. An exceptionally potent inflam matory stimulus for pyroptosis is LPS, a typical component of bacterial membranes. In pyroptosis, the axis LPS–CASP-11–GSDMD –pyroptosis is also referred to as "noncanonical."

The canonical pathway is initiated by the inflammasomes (Chapter 3). The NLRP3 inflammasome forms upon stimuli that include pathogen-associated molecular patterns (PAMPs), DAMPs, and absent in melanoma-2 (AIM2) as a response to cytosolic DNA. These inflammasomes recruit CASP-1/-11 via adapter molecule ASC (apoptosis-associated speck-like protein containing a C-terminal caspase recruitment domain) and lead to GSDMD cleavage and IL-1β/IL-18 maturation. Indeed, active production of systemically effective inflammatory cytokines during RN has been described for pyroptosis, thus rendering this cell death modality highly inflammatory. Pyroptosis is typically observed in macrophages during gram-negative infection and upon culture with bacterial or viral intracellular pathogens. Thus, pyroptosis is thought to mediate the immunogenic destruction of colonized niches. The other members of the gasdermin family (GSDMA, GSDMB, GSDMC, and GSDME) share structural analogies and potentially represent additional pore-forming molecules. In line with this, it was recently demonstrated that GSDME can be cleaved by Caspase 3 and granzyme B; thus, leading to pyroptosis as well.

Necroptosis

Necroptosis is the best-characterized mode of RN. It was discovered as a type of necrotic cell death in apoptosis-resistant cell lines. Classically, it is induced upon tumor necrosis factor-α (TNF-α) stimulation with concomitant inhibition of apoptosis (*e.g.*, by a virally expressed caspase inhibitor). In this context, TNFR1, Fas, and other death receptors (described at pathways of extrinsic apoptosis, see above) transduce this signal into the cell (see Fig. 17.4).

RIPK1, a key checkpoint of cellular fate, contains a motif next to its DD termed *RIP homotypic interacting motif* (RHIM). This motif is preserved in only four proteins within the mammalian proteome, and all these proteins are associated with the regulation of necroptosis. The necrosome, a higher-order structure with a poly-RIPK3 backbone, is central to necroptosis execution. Importantly, the necrosome can also be engaged downstream of TLR3 and TLR4 signaling through the protein TIR-domain-containing adapter-inducing interferon-β (TRIF) .

Viral recognition is integrated by the protein DNA-dependent activator of interferon regulatory factors (DAIs), also referred to as ZBP1.[16] Recently, DAI has been demonstrated to mediate in utero lethality of RIPK1-deficient mice by forcing RIPK3 oligomerization, which places inactive RIPK1 as an inhibitor of necroptosis. The RIPK1 kinase inhibitor necrostatin-1s (Nec-1s) seems to stabilize this conformation and thereby inhibit RIPK3 oligomerization.

Downstream of the active necrosome, mixed lineage kinase domain-like protein (MLKL) becomes phosphorylated by RIPK3. After phosphorylation of MLKL, this pseudokinase forms oligomers targeting the plasma membrane via its four-helical bundle (4HB) motif to mediate plasma membrane rupture. This process is controlled by the ESCRT-III complex, but the precise mechanisms of necrotic cell death and potential pore formation currently remain undefined. Clearly, phosphorylated MLKL (pMLKL) is required, but it is not sufficient to execute necroptosis. This is of significant interest as pMLKL targets multiple intracellular membranes, translocates to the nucleus to induce CXCL1/IL-33, and is stably expressed in terminally differentiated cells (*e.g.*, podocytes and endothelia) without killing them. Taken together, this might point to a still-unknown physiological role of pMLKL. Mice deficient in RIPK3 or FADD and MLKL die following challenge with influenza A virus. Therefore, RNA viruses, such as influenza A (and potentially also SARS-CoV-2, Chapter 25), drive necroptosis through activation of ZBP1/DAI.[17]

Bacterial infection is sensed via TLR3 and TLR4. This also recruits a RHIM domain–containing adapter protein named TRIF that engages RIPK3 to drive necrosome formation and necroptosis. It is likely that necroptosis exemplifies an evolutionary conserved program to defend the host against viruses and certain bacteria. In keeping with this hypothesis, some virus express caspase inhibitors, such as crmA (*e.g.*, cowpox virus), whereas viral protein M45 (*e.g.*, cytomegalovirus [CMV]) specifically inhibits necroptosis. CMV is a member of the herpesvirus family, which is characterized by its persistence within the host. M45 contains a viral RHIM domain and thereby suppress ZBP1/DAI-induced RIPK3 oligomerization within the necrosome.

Following activation of the necroptosis pathway, chemokines (Chapter 15) and cytokines (Chapter 14) are actively produced to be released in addition to DAMPs. These include CXCL1 and IL-33, a stimulator of ST2 signaling on regulatory T cells. These observations suggest that necroptosis limits the inflammatory response to a certain microenvironment and thereby prevents a systemic inflammatory response syndrome (SIRS) and death. As immunogenic as necroptosis may be on a local scale, on its own it may not induce the release of acute-phase proteins from the liver or with a fever.

CLINICAL RELEVANCE

Selection of Clinically Relevant Conditions Associated With Necroptosis

- Acute kidney injury
- Acute liver failure
- Acute respiratory distress syndrome
- Autoimmune disorders
- Cancer (necrosis in the center of solid tumors)
- Myocardial infarction
- Solid-organ transplantation
- Stroke
- Transplant rejection

Necroptosis critically contributes to diverse pathophysiological settings, including ischemia–reperfusion injury in solid organ transplantations, myocardial infarction, stroke, and SIRS. RIPK3- and MLKL-deficient mice have been demonstrated to be protected from preclinical models of such diseases by several independent groups. As a result, inhibitors of necroptosis (RIPK1 kinase inhibitors, RIPK3 kinase inhibitors, and MLKL inhibitors) have entered phase I and phase II clinical trials. As of the writing of this chapter, no cell death–preventing therapy had been approved by the US Food and Drug Administration (FDA). However, preclinical and first clinical data are very promising. Necroptosis inhibitors may soon become the first-in-class compounds to prevent RN.[18]

Ferroptosis

Ferroptosis is an important RN pathway with respect to ischemic injury of the brain, the heart, and the kidneys. It has been implicated as a mechanism to target cancer. Unlike extrinsic apoptosis and necroptosis, ferroptosis is not initiated by specific receptors. In renal tubules, it mediates an event referred to as *synchronized regulated necrosis* (SRN) of an entire functional unit. Thus, it provides a biochemical basis for the clinical observation of necrotic casts in the urine sediment of patients with acute kidney injury (Chapter 69). Key ferroptosis molecules have been associated with renal clear cell carcinomas.

Ferroptosis is critically mediated by the loss of NADPH (the major cellular redox equivalent) abundance as result of lipid peroxidation. This peroxidation was initially attributed to lipoxygenase ALOX5. However, it has become clear that ferroptosis is predominantly driven by the loss of function of glutathione (GSH) peroxidase 4 (GPX4) and/or ferroptosis-suppressor protein 1 (FSP1, also known as AIFM2), depending on the cell type. Upon GPX4-dysfunction, ferroptosis occurs rapidly in all metabolically active cells by means of ROS that are generated in the mitochondria.[4,19]

GPX4 requires GSH as a redox equivalent to function. Inhibition of the glutamate/cystine-antiporter system X_c^- in the plasma membrane depletes intracellular cysteine required for GSH synthase. Inhibitors of the antiporter system X_c^- are referred to as type 1 ferroptosis inducers (FINs). The default type 1 FIN is the compound "erastin," which was found in a screen for lethal compounds against Ras-transformed tumor cells. Type 2 FIN directly inactivate the active center of the selenoprotein GPX4. Other inhibitors of ferroptosis have also been identified through screens (*e.g.*, the first-in-class compound ferrostatin-1[20] and in cancer cell lines (*e.g.*, diffuse large B-cell lymphomas and clear cell renal carcinomas.[21]

GPX4 catalyzes the reduction of oxidized phospholipids and sphingolipids to the respective alcohols. This is associated with protection from cell death by ferroptosis, indicating that lipid peroxidation drives the loss of plasma membrane integrity. However, precise mechanisms remain to be determined. Defects in enzymes required for generation of polyunsaturated fatty acids (PUFAs), such as acyl-CoA synthetase long-chain family member 4 (ACSL4), prevent cells from undergoing ferroptosis. This further emphasizes the role of lipid peroxidation.

NADPH depletion has been identified as a downstream event of lipid peroxidation. This might explain the nature of cell death propagation during ferroptosis, which is not restricted to a single cell but causes necrosis beyond the plasma membrane limits. This phenomenon has been referred to as a *wave of death*, and was confirmed to occur in cell culture, in fish, and in renal tubules. NADPH may freely diffuse between neighboring cells through intercellular pores, such as gap junctions, leading to RCD in adjacent cells.[22,23]

CLINICAL RELEVANCE

Selection of Clinically Relevant Conditions Associated With Ferroptosis

- Cancer (necrosis in the center of solid tumors)
- Myocardial infarction
- Hemorrhagic stroke
- Acute kidney injury and acute tubular necrosis
- Rhabdomyolysis
- Solid-organ transplantation

At the time of writing this chapter, no immunomodulatory role for cells that die by ferroptosis had been described, and the immunogenicity of ferroptosis is certainly high. It is clear, however, that ferroptosis in ischemic myocardial tissue recruits neutrophils, and that addition of ferrostatins reduces the infarcted area and the scarring.[13] Post-myocardial infarction inflammatory syndromes (Chapter 37), such as Dressler syndrome, may be partially explained by this mechanism. Finally, it remains unclear why some heart attacks are not accompanied by neutrophil infiltration.

The role of mitochondria in ferroptosis has been recently elucidated.[24] Mitochondria are one source of the ROS that drive lipid peroxidation during ferroptosis. These findings, and the identification of FSP1 (see above) also suggested that ferroptosis might substantially overlap with what was previously referred to as "mitochondrial necrosis" or mitochondrial permeability transition-induced regulated necrosis (MPT-RN). The distinction of the RN pathways is of major relevance because of the development of specific small molecule inhibitors. Clear data exist to demonstrate that widespread depletion of mitochondria does not affect necroptosis.[25] Thus, while necroptosis does not require mitochondria and mitochondrial cell death does not require RIPK3, ferroptosis does. Evidence from isolated mitochondria, as well as from immunofluorescence and electron microscopy, suggests a central role of swelling of this organelle and MOMP during MPT-RN, as in apoptosis. To the best of our knowledge, a role of mitochondrial swelling in ferroptosis has not been determined.

Distinction of Other Regulated Necrosis Pathways—Mitochondrial Permeability Transition-Induced Regulated Necrosis and Parthanatos

MPT-RN is the consequence of mitochondrial permeability transition, a highly effective shortcut between the mitochondrial matrix and the cytosol. MPT is mediated through a pore (the MPT pore [MPTP]), the composition of which has been a matter of debate for at least two decades. Currently, a widely accepted model assumes a multiprotein complex physically or functionally involving proteins from the mitochondrial matrix, inner and outer mitochondrial membranes, the transmembrane space, and the cytosol. The pore is controlled by a cyclophilin named *cyclophilin D* (CYPD), a key modulator of MPTP opening and, thus, MPT-RN. Genetic absence of *ppif*, the gene that encodes for CYPD, protects mice from ischemic challenges, including stroke, myocardial infarction, and renal ischemia–reperfusion injury. As with other cyclophilins, the immunosuppressant cyclosporine (CsA) (Chapter 85) inhibits the opening of the MPTP and therefore prevents MPT-RN. This effect may well account for some of the immunosuppressive function of CsA. Other models favor CYPD to interact with the c subunit of F_1F_0–ATPase complex.[26] However, the role CsA in ferroptosis remains unclear, as does the extent to which ferroptosis and MPT-RN express two variations of the same RN pathway.

Parthanatos is defined as cell death that occurs following so-called overactivation of the DNA repair enzyme poly (ADP-ribose) polymerase 1 (PARP1). PAR polymers are formed and translocate to the outer mitochondrial membrane by unknown targeting mechanisms. The default induction of PARP1 overactivation includes a wide array of stimuli, ranging from DNA damage (*e.g.*, through irradiation), over ROS, stress, and induction by toxins, such as methylnitronitrosoguanidine (MNNG). Comparable to MPT-RN, parthanatos results in apoptosis-inducing factor (AIF)

release. It has been suggested that AIF requires active PARP1 to transfer ATP-ribose groups from NAD^+ onto its targets.

Most of our knowledge of PARP1 derives from cancer research. Many tumors have been demonstrated to overactivate PARP1. PARP inhibitors are in the FDA approval process for diverse cancers as a result of promising outcomes of phase III clinical trials. For *BRCA1/2* mutated ovarian and breast cancer, PARP-inhibitor olaparib has already been introduced in clinical practice. As with MPT-RN and ferroptosis, it is possible that parthanatos and ferroptosis are variations of the same RN pathway. Future experiments will have to clarify this issue.

KEY CONCEPTS

Prototype Inhibitors of Regulated Necrosis

- Inhibitors of necroptosis
  - Necrostatin-1 (Nec-1)
  - Nec-1s (Nec-1 stable)
  - Ponatinib
- Inhibitors of ferroptosis
  - Necrostatin-1 (Nec-1)
  - Ferrostatin-1 (Fer-1)
  - 16–86
  - Liproxstatin-1 (Lip-1)
  - Deferoxamin (DFO)
- Inhibitors of mitochondrial permeability transition (MPT)–induced regulated necrosis (MPT-RN)
  - Cyclosporine (CsA)
  - Sanglifehrin A (SfA)
- Inhibitors of parthanatos
  - Olaparib and many others
- Inhibitors of pyroptosis
  - zVAD-fmk (nonspecific caspase inhibitor)
  - emricasan

CONCLUDING REMARKS AND IMPLICATIONS FOR SOLID-ORGAN TRANSPLANTATIONS

Understanding the pathways of RN will allow screening for and/or design of specific inhibitors of the enzymes involved. Two major effects are anticipated from inhibiting RN. First, in clinically relevant conditions in which necrosis is the main determinant, such as stroke, myocardial infarction, sepsis, transplantation, acute liver failure, pancreatitis, and the center of solid tumors, it remains to be investigated how beneficial an antinecrosis therapy might be. Second, and probably of at least equal importance in solid-organ transplantation, necroinflammation deteriorates and amplifies the primary organ damage.

In transplantation, standard immunosuppression is carefully individualized to prevent proliferation of immune cells, but it does not avoid the priming of memory B cells. To transplant an organ that contains DAMPs and necrotic debris—a frequent scenario following a period of organ transfer—creates a strong stimulus for memory B cells. Expansion of these cells following the engagement of the BCR is a target for immunosuppression. The memory phenotype may be favored upon these special conditions. This scenario is comparable to a vaccination against intracellular components of the graft. When immunosuppression is tapered, ABMR becomes a major factor, lasting for years.[27] Under these circumstances, ABMR may be triggered by necroptosis of transplanted cells that have been virally infected. Finally, transplantation of necrotic debris hardly happens in living donor transplantation despite HLA mismatch and blood group mismatch and blood group incompatibilities. The upcoming years of RN and transplantation research should clarify how an anti-RN therapy may improve the outcome of transplants, with a particular focus on ABMR (Table 17.1).

TABLE 17.1 Classification of Damage-Associated Molecular Patterns

Classes of DAMPs[a]		Categories of Cognate Recognition Receptors/Sensors (Cell Bound, Humoral)
Class Ia DAMPs	DAMPs such as HMGB1, HSPs, nucleic acids including mitochondrial and cytosolic DNA	Sensed via binding to "classical" recognition receptors (*e.g.*, PRRs such as TLRs, RLRs, ALRs) on/in innate immune cells (*e.g.*, phagocytes and DCs), thereby triggering signaling pathways
Class Ib DAMPs	DAMPs such as CALR and eATP	Recognized by "nonclassical" recognition receptors (*e.g.*, the purinergic receptors P2X7.) thereby contributing to phagocytes including DCs activation
Class II DAMPs	DAMPs (*e.g.*, eATP, uric acid) operating as second signals to activate the NLRP3 inflammasome	Sensed by NLRP3 receptor to form assembly of the NLRP3 inflammasome contributing to phagocytes including DC activation
Class III DAMPs	DAMPs exposed on stressed cells such as MICs and ULBPs	Recognized by the activating NKG2D receptor (*e.g.*, on NK cells, thereby contributing to NK cell activation)
Class IV DAMPs	DAMPs in terms of neoantigens/neoepitopes (*e.g.*, NMHC-II, oxidized phospholipids, and the actin cytoskeleton.)	Recognized by binding to preexisting natural IgM antibodies to activate the complement cascade, thereby contributing to inflammation
Class V DAMPs	Dyshomeostasis-associated molecular patterns (*e.g.*, accumulation of unfolded proteins in the ER; intracellular ion perturbations, hypoxia, and redox imbalance)	Sensed by sensors of the UPR (*e.g.*, PERK) or sensed by NLRP3 receptor, thereby contributing to inflammation and DC activation
Class VI DAMPs	Metabolic DAMPs (*e.g.*, succinate)	Recognized by the "nonclassical" recognition receptor GPR91. Thereby promoting inflammation
Class VII[b] DAMPs	Nociceptor-activating DAMPs (*e.g.*, osmotic challenges, low and high temperature, capsaicin)	Sensed by nociceptors (*e.g.*, TRPA1 channels and TRPV1)

[a]The attempt to classify DAMPs as depicted in this table is restricted for this chapter only and with focus on their crucial role in allograft rejection.
[b]Class VII DAMPs sensed by nociceptors have been tentatively included in this table to show that DAMP-induced responses of the innate immune defense system may exceed the traditional phenomena of inflammation and adaptive immunity. Of course, this approach is debatable and we freely admit that there are still some deficits in our classification waiting for a final resolution.

CD, Cluster of differentiation; *DAMPs*, damage-associated molecular patterns; *DC*, dendritic cells; *eATP*, extracellular ATP; ER, *endoplasmic reticulum*; *GPR91*, G protein-coupled receptor 91; *HMGB1*, high mobility group box 1; *HSPs*, heat shock proteins; *IgM*, immunoglobulin M; *MICs*, MHC class I chain-related proteins; *NK*, natural killer; *NKG2D*, natural killer group 2 member D; *NLRP3*, NLR family, pyrin domain-containing protein 3; *NMHC-II*, nonmuscle myosin II-A heavy chain; *PERK*, the protein kinase R (PKR)-like endoplasmic reticulum kinase; *PRRs*, pattern recognition receptors; *P2X7*, purinergic receptor P2X7; *RLRs*, retinoic acid-inducible gene (RIG)-I-like receptors; *TLR*, Toll-like receptor; TRPA1, transient receptor potential cation channel subfamily A member 1; *TRPV1*, transient receptor potential vanilloid subtype 1; *ULBPs*, UL16 binding proteins.

ON THE HORIZON

In Vivo Interference With Regulated Necrosis

- RIPK1 inhibitors have shown promising results in animal studies/human phase I and II studies. They are the most likely first-in-class cell death inhibitors.
- Ferroptosis inhibitors may be useful for protection against the wave-of-death phenomenon seen in myocardial infarction and acute kidney injury. Perfusion of transplantable organs with ferrostatins may be a clinical application that could soon be tested.
- Immunogenicity of necrosis may boost cancer immunotherapy. Examples for this are approaches of gasdermin-mediated necrosis of tumor cells that are combined with checkpoint inhibitors.

ACKNOWLEDGMENTS AND CONFLICT OF INTEREST STATEMENT

This manuscript was supported by funding of the German Research Foundation, and the Heisenberg-Professorship to W.T., A.B., and A.L. and also supported by the TRR205 Adrenal of the DFG. A.L. was issued a patent for a combined inhibitor of necroptosis and ferroptosis, Nec-1F (patent # 20160943.5). Apart from this, all authors declare no competing interests.

REFERENCES

1. Kerr JF, Wyllie AH, Currie AR. Apoptosis: a basic biological phenomenon with wide-ranging implications in tissue kinetics. *Br J Cancer*. 1972;26(4):239–257.
2. Sarhan M, Land WG, Tonnus W, et al. Origin and consequences of necroinflammation. *Physiol Rev*. 2018;98(2):727–780.
3. Linkermann A, Stockwell BR, Krautwald S, Anders HJ. Regulated cell death and inflammation: an auto-amplification loop causes organ failure. *Nat Rev Immunol*. 2014;14(11):759–767.
4. Linkermann A, Green DR. Necroptosis. *N Engl J Med*. 2014;370(5):455–465.
5. Mulay SR, Linkermann A, Anders HJ. Necroinflammation in kidney disease. *J Am Soc Nephrol*. 2016;27(1):27–39.
6. Zhao H, Ning J, Lemaire A, et al. Necroptosis and parthanatos are involved in remote lung injury after receiving ischemic renal allografts in rats. *Kidney Int*. 2015;87(4):738–748.
7. Land WG, Agostinis P, Gasser S, et al. Transplantation and damage-associated molecular patterns (DAMPs). *Am J Trans*. 2016;16(12):3338–3361.
8. Land WG, Agostinis P, Gasser S, et al. DAMP—Induced allograft and tumor rejection: the circle is closing. *Am J Trans*. 2016;16(12):3322–3337.
9. Martinez J, Cunha LD, Park S, et al. Noncanonical autophagy inhibits the autoinflammatory, lupus-like response to dying cells. *Nature*. 2016;533(7601):115–119.
10. Hampe J, Franke A, Rosenstiel P, et al. A genome-wide association scan of nonsynonymous SNPs identifies a susceptibility variant for Crohn disease in ATG16L1. *Nat Genet*. 2007;39(2):207–211.
11. Rickard JA, O'Donnell JA, Evans JM, et al. RIPK1 regulates RIPK3-MLKL-driven systemic inflammation and emergency hematopoiesis. *Cell*. 2014;157(5):1175–1188.
12. Schiering C, Krausgruber T, Chomka A, et al. The alarmin IL-33 promotes regulatory T-cell function in the intestine. *Nature*. 2014;513(7519):564–548.
13. Li W, Feng G, Gauthier JM, et al. Ferroptotic cell death and TLR4/Trif signaling initiate neutrophil recruitment after heart transplantation. *J Clin Investig*. 2019;129(6):2293–2304.
14. Vanden Berghe T, Linkermann A, Jouan-Lanhouet S, et al. Regulated necrosis: the expanding network of non-apoptotic cell death pathways. *Nat Rev Mol Cell Biol*. 2014;15(2):135–147.
15. Broz P, Pelegrín P, Shao F. The gasdermins, a protein family executing cell death and inflammation. *Nat Rev Immunol*. 2020;20(3):143–157.
16. Linkermann A, Skouta R, Himmerkus N, et al. Synchronized renal tubular cell death involves ferroptosis. *Proc Natl Acad Sci U S A*. 2014;111(47):16836–16841.
17. Zhang T, Yin C, Boyd DF, et al. Influenza virus Z-RNAs induce ZBP1-mediated necroptosis. *Cell*. 2020;180(6):1115–1129.e13.
18. Degterev A, Linkermann A. Generation of small molecules to interfere with regulated necrosis. *Cel Mol Life Sci: CMLS*. 2016;73(11–12):2251–2267.
19. Friedmann Angeli JP, Schneider M, Proneth B, et al. Inactivation of the ferroptosis regulator Gpx4 triggers acute renal failure in mice. *Nat Cell Biol*. 2014;16(12):1180–1191.
20. Dixon SJ, Lemberg KM, Lamprecht MR, et al. Ferroptosis: an iron-dependent form of nonapoptotic cell death. *Cell*. 2012;149(5):1060–1072.
21. Yang WS, Stockwell BR. Ferroptosis: death by lipid peroxidation. *Trends Cell Biol*. 2016;26(3):165–176.
22. Shimada K, Hayano M, Pagano NC, et al. Cell-line selectivity improves the predictive power of pharmacogenomic analyses and helps identify NADPH as biomarker for ferroptosis sensitivity. *Cell Chem Biol*. 2016;23(2):225–235.
23. Tonnus W, Linkermann A. "Death is my Heir"—Ferroptosis connects cancer pharmacogenomics and ischemia-reperfusion injury. *Cell Chem Biol*. 2016;23(2):202–203.
24. Gaschler MM, Hu F, Feng H, et al. Determination of the subcellular localization and mechanism of action of ferrostatins in suppressing ferroptosis. *ACS Chem Biol*. 2018;13(4):1013–1020.
25. Tait SW, Oberst A, Quarato G, et al. Widespread mitochondrial depletion via mitophagy does not compromise necroptosis. *Cell Rep*. 2013;25(4):878–885.
26. Bonora M, Bononi A, De ME, et al. Role of the c subunit of the FO ATP synthase in mitochondrial permeability transition. *Cell Cycle (Georgetown, Tex)*. 2013;12(4):674–683.
27. Sellares J, de Freitas DG, Mengel M, et al. Understanding the causes of kidney transplant failure: the dominant role of antibody-mediated rejection and nonadherence. *Am J Transplant*. 2012;12(2):388–399.

Human Genomics in Immunology

Jennifer M. Puck and Robert L. Nussbaum

The completion of the Human Genome Project in 2003, 50 years after the landmark 1953 publication of the double-helical structure of DNA by James Watson and Francis Crick, was a major milestone in modern biology. The sequence provided a comprehensive and accurate view of the genetic makeup of humans, with an error rate of <0.001% and only a few hundred gaps.[1,2] This detailed picture of a composite human genome is for human genetics what Vesalius's publication of the structure of the human body, *De humani corporis fabrica*, was for anatomy. Like Vesalius's work, it continues to serve as a foundation for further discovery in such areas as genetic variation, gene function, human physiology, and the genetic basis for disease.

The human genome has about 21,000 protein-coding genes distributed on 23 pairs of chromosomes.[3] The total length of the haploid genome is ~3 billion nucleotide base pairs (bp). The protein-coding segments, called exons, are interspersed with noncoding DNA sequences, called introns. The aggregate protein-coding sequences, referred to as the "exome," account for ~1% to 1.5% of the genome, with exons plus introns ~20% and the remainder is DNA located between genes. Some of the noncoding DNA contains regulatory elements that direct gene expression, act as origins for DNA replication, signal where exons start and end so that introns can be spliced out of messenger RNA, control chromatin conformation, encode small regulatory RNA molecules, and participate in three-dimensional looping to produce the large-scale structure of chromosomes. About 40% of the total DNA consists of families of repeated sequences. These repeat elements are generally silent, but they may be involved in some types of gene regulation and can participate in mutations by facilitating deletion, duplication, and insertion. Each cell expresses only a subset of the entire gene repertoire at a given point in time. "Housekeeping" genes are expressed in almost all tissues and cell types, where they perform basic metabolic and structural functions. Other genes are under very specific control, with their expression restricted to one or a few cell types. Differential gene expression specifies the unique composition and functions of cells (e.g., immunoglobulin [Ig] in B cells, and T-cell receptors [TCRs] in T cells). Some genes encode transcription factors, which translocate to the cell nucleus to coordinate the expression of groups of tissue-specific target genes. A few of these apparently act as "master" genes during particular developmental processes or in specific cell lineages. Over 450 genes are known that when mutated result in immune system defects, affecting either innate or adaptive immune cells by altering processes governing cell growth, differentiation, effector functions, or cell death such as by apoptosis.[4,5]

KEY CONCEPTS

- The sequence of the human genome and catalogue of genome variation among humans has revolutionized our approach to heritable immune disorders.
- Genomic DNA sequence is the finest scale physical map of the genome, describing the exact locations of each gene on each of the 22 paired autosomes and the X and Y sex chromosomes.
- The human genome encompasses ~3 billion nucleotide base pairs of DNA with approximately 21,000 protein-coding genes, with each cell expressing only a subset of the genes.
- Coding sequences of genes, referred to as the exome, comprise only 1%–1.5% of the genome.
- Areas of the genome that are critical for normal function of gene products have been conserved through evolution; thus, species conservation suggests functional significance.
- A variety of types of DNA variation are recognized, including single nucleotide changes, insertions or deletions of a few to many nucleotides, copy-number variants consisting of deletions or duplications or triplication of many hundreds to thousands of nucleotides, and structural variants such as inversions or translocations.
- Interpretation of the significance of an observed variant in DNA sequence may require consideration of its location, frequency in the population, inheritance in a family, and specific effect on the resulting gene product.

GENOME ANNOTATION

Each individual inherits genomic material from both parents. The maternal and paternal copies of the genome are nonidentical, and the difference in sequence (a *variant*) at any one location (*locus*) is called an *allele*. The two alleles at a single locus on the two copies of a chromosome (except for the X chromosome in males, when there is only one allele) constitute the *genotype*. Genotype should not be confused with *haplotype*, which is the set of alleles present at a series of loci on a single chromosome. The individual's genotype interacts with the environment throughout life to create the *phenotype*. Some phenotypes, such as body weight, are simple to measure, whereas others are based on complex laboratory evaluation (e.g., T-cell proliferation). Phenotypes may be discrete traits (normal vs. abnormal) or quantitative traits that have a continuous range of values. Relatively common variants, or polymorphisms, account for some of the phenotypic variation between healthy individuals or between populations, and the cumulative percentage of phenotypic variation explained by genetic variation is called *heritability*. In monogenic diseases (also called *mendelian* or single-gene disorders), the presence of variant(s) is usually considered necessary and sufficient to cause disease; when an abnormal genotype does not cause disease in everyone who has that

genotype, the phenomenon is referred to as *reduced penetrance*. The landscape of genetic variation in each individual includes a continuum of gene effects ranging from weak (common single-nucleotide polymorphisms [SNPs] detected in disease-association studies) to strong (rare, damaging variants detected in single-gene disorders).

The consensus sequence of the human genome has been only the first step in exploring normal biological functions and how variants cause disease. The Human Genome Project has matured into a number of basic and applied research areas: (i) acquiring a comprehensive catalogue of human variation and the impact of individual variants on phenotype, including disorders of human development; (ii) comparing human genomes with those of other organisms and human ancestors; and (iii) learning how to interpret all the sequence elements within the genome, not just the codons that determine the amino acid makeup of proteins. Almost two decades after initial "completion" of the human genome sequence a fully complete and accurate single, contiguous reference haploid genome is still being constructed, and updated versions continue to be released. The greatest challenges to completing the human genome sequence are posed by regions that contain segmental duplications of nearly the same sequence.

HUMAN VARIATION

The first publicly available human genome sequence was constructed from a small number of individuals, a composite haploid sequence rather than an actual sequence of a single individual. It was neither a "normal" nor a "control" genome; instead, it was a reference, providing a universally available sequence against which the genomes of individual humans, as well as other species, could be compared and any differences, or variants, determined. Even before the human genome sequence was completed, the need to discover as broad a range of human variations as possible in populations from around the world was recognized to be essential for understanding how genetic variations lead to differences in phenotypic traits and disease susceptibilities. The first effort to catalogue human genetic diversity was the dbSNP (database of single nucleotide polymorphisms) project, followed by the 1000 Genomes project. These catalogues have been supplemented enormously by more comprehensive efforts, including the NHLBI GO Exome Sequencing Project (ESP), and the Exome and Genome Aggregation Consortium databases (EXaC and gNOMAD), which have made publicly available a vast number of variants and their frequencies from hundreds of thousands of individuals.

Variants are classified as rare or common. Most variants (85%) have allele frequencies substantially below the 1% cutoff for being a called a polymorphism and are, instead, considered rare, sometimes restricted to a single ethnic group or even a single kindred.[6]

DNA variation is also classified according to the *type* of DNA change (Table 18.1). Single nucleotide variants (SNVs), insertion/deletion variants (indels), copy number variants (CNVs), and structural variants (SVs) can have different consequences, depending on their location and the number and identity of nucleotides affected. The simplest and most common of all variants are SNVs, in which one nucleotide in the reference sequence is substituted by another. A locus characterized by an SNV usually has only two alleles, corresponding to the more common (major allele) and less common (minor allele) base found at that particular location, although, theoretically, four alleles at any base position are possible (either an adenine [A], cytosine [C], guanine [G], or thymidine [T] nucleotide). SNVs are observed on average once every 1000 bp in the genome but are not distributed evenly and have different frequencies in different populations. Most SNVs are not located within exons or other known functional elements; moreover, over half of coding SNVs do not alter the predicted amino acid sequence of the encoded protein and are thus termed *synonymous*, whereas the remainder that do alter the amino acid sequence are *nonsynonymous*. Other SNVs introduce or change a stop codon, and still others alter a known splice site; such SNVs are likely to have significant impact on the expression of the gene containing the SNV as well as possible phenotypic consequences.

A second general class of variation is the result of insertion and/or deletion compared to the reference sequence, ranging from 1 up to an arbitrary cutoff of ~300 to 1000 bp. When reference nucleotides are simply deleted or duplicated, the variant is referred to as a "del/dup." When the reference sequence has some nucleotides deleted and replaced by another inserted

TABLE 18.1 Types of DNA Variation

Type	Description	Methods for Detection
Single nucleotide variant (SNV)	Sequence change in which, compared with a reference sequence, one nucleotide is substituted for another	Readily found by Sanger or massively parallel sequencing with short or long reads
Deletion or duplication variant (del/dup)	Sequence change involving 2 to ~1000 base pairs (bp) in which reference nucleotides are either missing (deleted) or duplicated and inserted directly 3′ to the reference nucleotides	Small del/dup detected by short-read sequencing, but long reads may be required
Insertion/deletion variant (indel)	Sequence change involving between 2 and ~1000 bp in which one or more reference nucleotides are replaced by one or more other nucleotides and is not an SNV or SV	Most often found by short-read sequencing, but may require long reads
Copy number variant (CNV)	The del/dup or indel is arbitrarily set as larger than ~1000 bp	Difficult to find by short-read sequencing without specialized CNV detection software; long reads or other specialized tools may be required
Structural variant (SV)	Inversion, translocation	May be difficult to find by any sequencing method other than whole-genome sequencing, depending on where the inversion or translocation occurs; karyotype analysis required for large-scale cytogenetic changes

From Human Variome Society, Sequence Variant Nomenclature, version 20.05. Available at: http://varnomen.hgvs.org (accessed January 25, 2021).

sequence, the variant is referred to as an "indel." Each individual carries many hundreds of thousands of indels.

Some del/dup variants are multiallelic because of variable numbers of the identical segment of DNA inserted in tandem at a particular location, thereby constituting what is referred to as a *microsatellite*, or *short tandem repeat* (STR). Microsatellites are segments of DNA composed of units of 2, 3, or more nucleotides, such as $(TG)_n$, or $(CAA)_n$, with n between 2 and several dozen. Many tens of thousands of polymorphic microsatellites are known to exist throughout the human genome, making them useful for tracking inheritance of DNA blocks within kindreds (linkage analysis) as well as for forensic identification of an individual's DNA identity or fingerprint. STR DNA segments within or adjacent to coding exons rarely, but famously, can expand to become hundreds or thousands of nucleotides long, thereby causing such human disorders as fragile X syndrome or Huntington disease. Even without expansion, STRs within exons, some of which may be as small as 9 to 25 bp, have an outsized impact on the frequency of human disease, since they confer a five- to sixfold increase in the frequency of rare disease-causing indel mutations compared with neighboring exon sequences that do not contain an STR.[7]

One subclass of indel variants arises from mobile elements. Nearly half of the human genome consists of widely dispersed families of repetitive elements, of which the two most common are *Alu* (a ~300 bp short interspersed nuclear element) and LINE (long interspersed nuclear element). Although most of the copies of these repeats are stationary, some of them contribute to human genetic diversity through *retrotransposition*, or insertion of a DNA segment generated through transcription of an *Alu* or LINE element into an RNA that is then reverse-transcribed back into DNA. Each mobile element indel consists of two alleles, one with and one without the inserted element. Mobile element polymorphisms are found on all human chromosomes; some have been implicated in gene disruptions underlying human disease.

Another important type of human variation includes CNVs, variations in the number of copies of DNA segments >1000 bp to many hundreds of kilobase pairs (kb). CNVs >500 kb occur in 5% to 10% of the general population, whereas those >1 million bp (1 Mb) occur in 1% to 2%. Not only are CNVs a significant contributor to human variation and disease but also the areas of the genome where they are found are often sites of segmental duplications, some of the most difficult regions in which to develop an accurate reference sequence because the most widely used sequencing technologies generate only short reads of a few hundred bp.[8,9] Many CNVs include genes; thus, CNVs are frequently implicated in diseases that result from altered gene dosage. One well-known human immunological and multisystem disorder, DiGeorge syndrome, is caused by heterozygous deletion CNVs between four sets of repeated DNA elements on chromosome 22q11.2.[10] The deletion occurs de novo in about 1 in 5000 individuals and encompasses dozens of genes, the most important of which for the DiGeorge phenotype is the transcription factor gene *TBX1*.

The most common SVs in the human genome are inversions, which range from a few bp to up to several Mb. Inversions are present in either of two orientations in the genomes of different individuals.[11] Most do not involve gain or loss of DNA, allowing each opposite-oriented allele to achieve a substantial frequency in the general population. Inversions can, however, cause significant gains or losses of DNA in the offspring of inversion carriers because of aberrant recombination during meiosis, leading to serious syndromes brought about by chromosomal imbalance. Furthermore, if inversions interfere with normal gene expression by disrupting a gene or altering the physical relationship between a gene and its regulatory elements, disease may ensue.

CLINICAL IMPACT OF HUMAN VARIATION

One of the greatest challenges facing human geneticists is linking variation to phenotype.[12] The significance for the health of the vast majority of variants of any type is unknown, and yet this knowledge is essential if we are to apply genomics to clinical care. The impact of variants ranges from completely benign to highly pathogenic, the latter causing devastating disorders of the immune system that may occur as new mutation dominants or as autosomal recessively inherited syndromes. Even common polymorphic variants may affect health or longevity, although their being common means that they are likely to produce a relatively subtle alteration of disease susceptibility rather than directly cause a serious illness.[13] Working out the functional impact of human variation will occupy genomics researchers for many years to come. An essential component of this work is to make databases of genetic variants and their impact on human health available to the research and clinical communities, as is being done with the ClinVar database hosted by the National Library of Medicine in the United States and the Leiden Open-source Variation Database (LOVD).[14–16]

COMPARATIVE GENOMICS

Evolution at work is nowhere better illustrated than in the field of comparative genomics, which deals with similarities in the sequence, structure, and chromosomal location of genes between different species whose evolutionary paths diverged up to hundreds of millions of years ago. Direct sequence comparison has revealed that an enormous number of human proteins have orthologues (genes derived from a common ancestor) in other organisms, ranging from 87% in chimps to 79% in mice, 63% in zebrafish, 39% in the fruit fly (*Drosophila melanogaster*), and 31% in the nematode (*Caenorhabditis elegans*). The study of the human genome and the genomic basis for human disease has benefited from studies in other organisms, particularly multiple strains of mice, in which many decades of inbreeding and gene manipulation have permitted characterization of the roles of many gene products and phenotypic consequences of variation. Mouse models of immunity have been highly relevant and informative for humans.

However, orthologous genes may serve different functions in different species; therefore, one cannot assume that a disease-causing variant in humans will cause a similar defect when its orthologue is similarly mutated in mice, and vice versa. For example, the V(D)J recombination activating genes *RAG1* and *RAG2* in humans and *Rag1* and *Rag2* in mice appear to have identical functions, with knockout mice showing the same inability to recombine T- and B-lymphocyte antigen-receptor genes as humans with RAG1 or RAG2 deficient severe combined immunodeficiency (SCID; Chapter 34).[17] As a result, the phenotype of RAG1/2 deficiency in both species is absent T and B cells with normal natural killer (NK) cells, referred to as $T^-B^-NK^+$ SCID. A contrasting situation occurs in other SCID genotypes. For example, humans lacking the common γ chain

(γc) of receptors for interleukin-2 (IL-2) and other cytokines, caused by mutations in the X-linked gene *IL2RG* have SCID in which T and NK cells are absent but nonfunctional B cells are present in normal to high numbers, $T^-B^+NK^-$ SCID. Mice with the orthologous gene *Il2rg* mutated or removed can make T cells but have no B cells, which gives them a $T^+B^-NK^-$ phenotype.[18]

Genes other than *IL2RG* alone must be responsible for this difference between species. Such genes, known as *modifiers*, have not yet been identified. Notably, different strains of mice can also have important phenotypic differences in the presence of a single gene mutation under study; for example, some strains such as nonobese diabetic (NOD) and Murphy Roths Large (MRL) mice are highly prone to developing autoimmunity, while others such as C57BL/6 (B6) are resistant.[19]

FUNCTIONAL GENOMICS

Immediately after the Human Genome Project was declared to have been complete in 2003, an important follow-on project was launched to identify the functional segments of DNA, particularly portions in the 98% to 99% lying outside of the coding exons of genes, for which there was no simple sequence code that could be understood the way the triplet codon code can be read. This project, termed ENCODE (for "encyclopedia of DNA elements") set out to address the problem by studying functional DNA elements in genomes of humans and model organisms.[20] Before ENCODE, estimates were that 3% to 8% of the human genome had some role in function, given that this fraction of the genome appeared to be highly conserved among species with only very limited variation. This estimate was far too low, as it did not take into account rapidly evolving functional elements or those restricted to particular evolutionary lineages, nor did it include segments of DNA too small to show conservation with statistical significance, or functional elements in repetitive DNA not reliably scored as being evolutionarily conserved.

Since the same genome is present but functions differently in different cells of an individual, ENCODE used a number of different tissues for its studies. A comprehensive catalogue of every segment of DNA that is transcribed into RNA in any tissue was required. A second project, GTEx (Genotype-Tissue Expression) was launched to catalogue tissue-specific gene expression and regulation from 54 non-diseased tissue sites across nearly 1000 individuals.[21] ENCODE has analyzed not only total whole-cell RNAs but also those located in the nucleus or cytosol, because subcellular localization of RNAs is important in understanding their functions. Assays for functionality of segments of DNA are to a large extent circumstantial and include biochemical evidence, such as identifying (i) segments of DNA located in chromatin loops that allow chromatin-chromatin interaction; (ii) regions of open chromatin, which are accessible to transcription; (iii) motifs that transcription factors recognize and bind to; (iv) regions associated with histones that have been modified to either promote or suppress transcription; and (v) regions with differential methylation of cytosine residues in different tissues, with methylation being associated with inactivity.

The circumstantial nature of the evidence developed by ENCODE assays associating DNA elements and chromatin changes with gene expression means that not all elements so implicated will turn out to have functional significance. A stringent threshold for ascribing a functional role to DNA segments (i.e., direct effect on gene expression and phenotype of at least one human cell type) suggests that 10% to 20% of human noncoding DNA functions in gene regulation. As of 2020, the ENCODE database contained 1 million functional human elements comprising 7.9% of the human genome and about half that many in mouse.[22] More research will be needed before the ENCODE project delivers its final assessment of the fraction of the human genome that plays a role in gene regulation—that is, in how different cells use their genomes.

APPLYING HUMAN GENOMICS TO DISCOVERY OF DISORDERS OF THE HUMAN IMMUNE SYSTEM

Scientists and clinicians dealing with rare, single-gene disorders must cope with both locus and allele heterogeneity. *Locus heterogeneity* means that the same or a similar phenotype results from mutation in one of several different genes. For example, SCID can result from mutation in the adenosine deaminase, *IL2RG*, *JAK3*, and so on. *Allelic heterogeneity* means that the disease is caused by different mutations in the same gene. In X-linked disorders, allelic heterogeneity is typically high because affected individuals have reduced reproduction (negative evolutionary selection), and most mutations are lost in the population after a few generations. At autosomal loci some mutations have reached appreciable frequency as a result of demographic processes (e.g., *founder effect*).[23] In the heterozygous state, recessive mutations can be weakly deleterious, neutral, or actually confer a small advantage.

If the mutant alleles are common in a population, one can test affected individuals and potential carriers directly for those mutations. However, none of the immune deficiencies result from common mutations that would lend themselves to screening of the general population, except for special instances when one particular variant is responsible for the vast majority of cases of a disease in a particular population (e.g., the particular *DCLRE1C* variant in the Artemis gene causing SCID among Navajo Native Americans). One DNA-based population screening test that has proven highly effective does not identify a specific gene variant, but rather detects T-cell receptor excision circles (TRECs), DNA byproducts of successful T-cell receptor gene recombination in developing thymocytes (Chapter 9).[24] Because TRECs are a biomarker for normal T-cell production, their absence in infant dried blood spots constitutes a newborn screen to identify essentially all cases of SCID as well as other conditions with very low T-cell numbers, regardless of genotype.

Genome research has provided geneticists with a catalogue of essentially all known human genes, knowledge of their location and structure, and an ever-growing list of variants in DNA sequences found among individuals in different populations. In the past, geneticists followed two approaches to identifying the genetic basis for human disorders. The first, linkage analysis, is *family based* (Fig. 18.1). Linkage analysis takes advantage of family pedigrees to follow the inheritance of a disease among family members and test a few hundred DNA variants distributed throughout the genome for consistent, repeated coinheritance, or *segregation*, with the disease. A demonstration of significant coinheritance with a *variant* or *variants* located in a particular region of the genome indicates that the disease-causing mutation is also located in a gene nearby. The marker

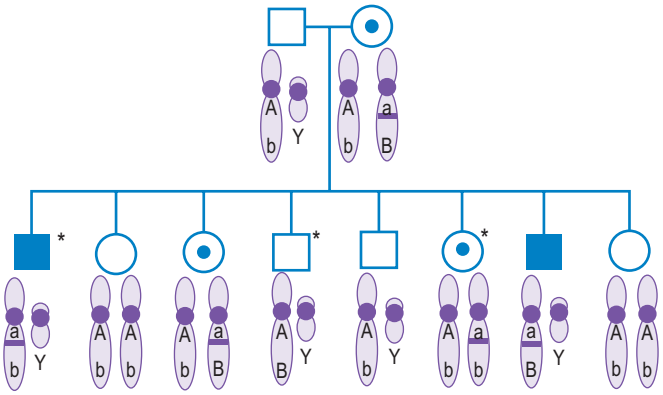


FIG 18.1 An Example of Family-Based Linkage Analysis as a Means to Identify a Disease Gene. The pedigree of a family with X-linked severe combined immunodeficiency (SCID) caused by an *IL2RG* mutation in the proximal long arm of the X chromosome *(marked with dark bar)* is shown. Affected males have *filled in square*; female carriers are designated by *dot in the center* of their symbols. Two loci, one with alleles A and a, and the other with alleles B and b, are shown in the proximal and distal long arm of the X (Xq). There are no recombinations during female meiosis between allele a and the *IL2RG* mutation in any of the eight children, while three children *(marked with asterisk)* show a recombination with the locus on the distal Xq. The father's X chromosome is passed on without any recombinations to each of his daughters and is shown on the left of each pair of X chromosomes in the daughters.

variants segregating with the disease are usually not the variants responsible for the disorder, but are close enough that recombination during meiosis between the marker and the gene mutation responsible for the disease is uncommon.

The second approach, genome-wide association analysis (GWAS), is *population based*. A sample of affected individuals, or "cases," taken from the population, is chosen along with a matched set of unaffected "control" individuals from the same population. Then, a large number of variants throughout the genome are analyzed for an increased or decreased frequency in cases versus controls. The alleles used to test for association need not be the actual variants functionally responsible for the disease association—in fact, this would be highly unlikely. Instead, GWAS, like linkage, depends on one or a few of the vast number of marker alleles tested for association to be located close enough to the functionally responsible alleles for association to have been carried along, on the same chromosome in a haplotype block, through many generations (Fig. 18.2). Association analysis does not require pedigrees and is useful for complex diseases that do not show strict mendelian inheritance.

Linkage analysis and association studies have limitations in investigating the genetic basis for human immune disorders. Linkage analysis is problematic if the disorder is a rare autosomal recessive condition, with a consequence that there are not enough families with two carrier parents to enable such a study; nonetheless, the increased frequency of consanguineous matings in some populations has been utilized to overcome this limitation.[23] Another challenge with linkage analysis is if the disorder is genetically lethal so that it is never inherited and always occurs sporadically as a result of a new mutation. Detecting an association in a case-control study is also a problem when the frequency of any particular allele associated with the disease is too low among the cases to give a detectable association. For example, if the disorder arises from different, independent mutations and if these mutations are found on many different haplotypes in affected individuals, it may be very difficult to establish a significant association with any one variant marker.

Genomic sequencing has become an important alternative to linkage analysis and GWAS. Vastly improved methods of DNA sequencing have opened up new possibilities to discover the genes and mutations responsible for rare mendelian disorders. One can generate a whole-genome sequence (WGS) or, in what has often proven a cost-effective compromise, a sequence of only about 2% of the genome, the part containing the exons of genes, referred to as a whole-exome sequence (WES).

FINDING DISEASE-CAUSING VARIANTS BY ANALYZING DEEP SEQUENCE DATA

Given the high background diversity of genome sequences between individuals, how can pathogenic variants be distinguished from benign polymorphisms? Variant assessments require extensive use of public genomic databases and software tools, including the human genome reference sequence, databases of variants with their allelic frequencies, software that assesses how an amino acid substitution might alter gene function, collections of known disease-causing mutations, and databases of functional networks and biological pathways.[5]

In a recent report of the clinical use of WES in rare disorders, 2000 individuals with a variety of disorders that had escaped diagnosis despite thorough conventional clinical evaluations underwent WES.[25] A likely causative mutation or pair of mutations was found in 504 persons (25.2%). Of these mutations (of which many were severe developmental defects), half were de novo, not present in either parent. Also of interest was that 23 (4.6%) of these 504 patients in whom a diagnosis was made through WES had two different genetic disorders, resulting in a combined phenotype, most likely obscuring the diagnosis of either of the individual disorders had they been present alone.

As an example of what is now possible, suppose that there is a family "trio" consisting of a child affected with a rare immunodeficiency and the child's parents. WES is carried out, yielding typically over 4 million SNVs, indel, and CNV differences in the child compared with the human genome reference sequence. Which of these variants would be responsible for the disease? The extraction of useful information from such a massive amount of data relies on creating a variant filtering scheme based on a variety of reasonable assumptions about likely responsible explanations for the disease; application of filtering in one instance is shown in Fig. 18.3.[26]

1. *Location with respect to protein-coding genes*. Keep variants that are within or near exons of protein-coding genes and discard variants deep within introns or intergenic regions. It is possible, of course, that the responsible mutation might lie in a noncoding RNA gene (e.g., *RMRP*, the gene mutated in the immunodeficiency and dwarfing syndrome cartilage hair hypoplasia). However, these are currently more difficult to assess, and thus, as a simplifying assumption, it is reasonable to focus initially on protein-coding genes.

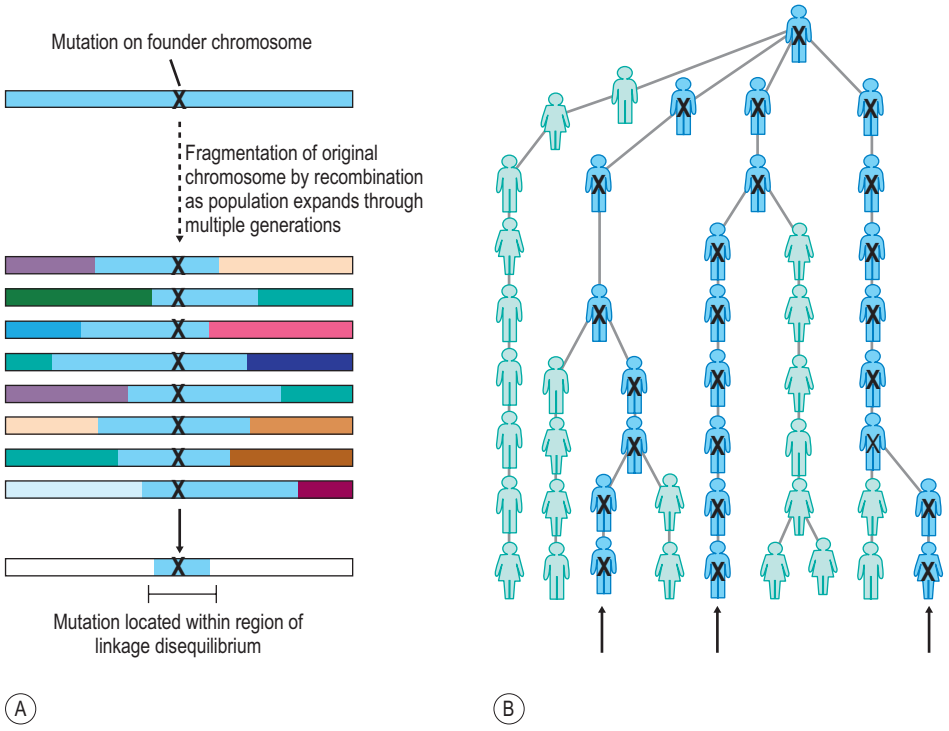


FIG 18.2 An Example of a Population-Based Genome-Wide Association Analysis As a Means to Identify a Disease Gene. (A) A mutation (X) that predisposes to a disease first occurs on a chromosome with a certain set of alleles at polymorphic loci along that chromosome *(symbolized by the blue color)*. With each generation, meiotic recombination exchanges the alleles that were initially present at polymorphic loci on the "blue chromosome" for other alleles present on homologous chromosomes *(symbolized by other colors)*. Over many generations, the only alleles that remain associated with the mutation are those at loci so close to the mutant locus that recombination between them has not occurred. These alleles constitute a disease-associated haplotype. (B) Affected individuals in most recent generation (*arrows*) carry the mutation (**X**) and are enriched for the disease-associated haplotype (*individuals in blue*) compared with unaffected individuals. Depending on the age of the mutation and other population genetic factors, a disease-associated haplotype ordinarily spans a region of DNA of a few kilobases to a few hundred kilobases. (From Nussbaum R, McInnes RR, Willard HF. *Thompson and Thompson Genetics in Medicine*. 8th ed. Toronto, Canada: Elsevier Canada; 2016: 177, Fig. 10.8, with permission.)

2. *Population frequency*. Keep rare variants from Step 1 and discard common variants with allele frequencies <0.05 (or some other arbitrary number between 0.01 and 0.1) because common variants are highly unlikely to be responsible for a disease whose population prevalence is much less than the q^2 predicted by the Hardy-Weinberg equilibrium.
3. *Deleterious nature of the mutation*. Keep variants from Step 2 that cause nonsense or nonsynonymous changes in codons within exons, cause frame-shift mutations, or alter highly conserved splice sites. Discard synonymous changes that have no predicted effect on protein function (unless there is reason to suspect that they influence splicing or expression, such as the last nucleotide of an exon, which is typically "G").
4. *Consistency with likely inheritance pattern*. If the disorder is considered most likely to be autosomal recessive, keep any variants from Step 3 for which an affected child has two variants in the same gene and each parent has one of the variants. The child need not be homozygous for the same deleterious variant but could be a compound heterozygote for two different deleterious mutations in the same gene.

If there were consanguinity in the parents, the candidate genes and variants might be further filtered by requiring that the child be a true homozygote for the same mutation derived from a single common ancestor. If the autosomal recessive mode of inheritance is correct, then the parents should both be heterozygous for the variants. The disease-causing variant in a male could also be X-linked, in which case any variant found in an X-linked gene for which the mother is a heterozygote would be a candidate. For either the autosomal or X-linked case, the disorder could be the result of a new dominant mutation; in this case, keep variants from Step 3 that are de novo changes in the child and are not present in either parent.

In the end, millions of variants can often be filtered down to a handful of SNVs, indels, or CNVs affecting a reasonable number of genes. Once the filtering reduces the number of genes and alleles to a manageable number, they can be assessed individually for other characteristics. Do any of the genes have a known function or tissue expression pattern that would be expected if it were the potential disease gene? Is the gene already implicated in other disease phenotypes, or does it have a role in pathways in which mutations are known to cause similar or different phenotypes? Finally, is the same gene mutated in other patients or in an animal model with a similar phenotype? Finding mutations in one of the candidate genes in additional patients would support that gene's role in the disease of the patient under study. In some cases, one gene remaining on the list in Step 4 may become a leading candidate because its involvement makes biological or genetic sense or it is known to be mutated in other affected individuals. In other cases, however, the gene

responsible may turn out to be entirely unanticipated on biological grounds or may not be mutated in other affected individuals because of locus heterogeneity.

In some cases, the difficulty in making a diagnosis was the result of there being a variant in a gene, previously unknown to cause human disease, but implicated by finding variants that fit the inheritance pattern (de novo dominant or autosomal recessive), and then later confirmed by finding other variants in phenotypically similar patients. In others, the gene involved had been previously described, but not in association with the phenotype for which the patient was undergoing diagnostic exome analysis. For example, in the kindred depicted in Fig. 18.3, two children had an undiagnosed autoimmune syndrome consisting of early-onset bullous pemphigoid, nephritis, autoimmune anti-factor-VIII hemophilia, and inflammatory bowel disease. WES was performed for both children, their unaffected sister, and both unaffected parents.[26] Eighty-eight genes were found in which the two affected children were compound heterozygotes for two different rare mutations, whereas the unaffected sister and both parents were carriers for only one of the two mutations found in the children. Of these, only one gene was known to have any immunological function: *ZAP-70*, previously known to be mutated in combined immunodeficiency but not with strictly autoimmune disease of this type (see Fig. 18.3). Both affected children were compound heterozygotes for two missense mutations, p.R192W and p.R360P, in *ZAP-70*. The mother carried the R192W allele, whereas the father and unaffected sister carried the R360P allele. Neither mutation had been reported previously, nor were they near mutations in the same gene that had been previously associated with human disease. Functional studies revealed the p.R192W had reduced activity, whereas the R360P encoded a modestly hyperactive protein because of disruption of an autoinhibitory mechanism. The combination of these alleles, as proven in laboratory studies, fully explained the autoimmune phenotype.

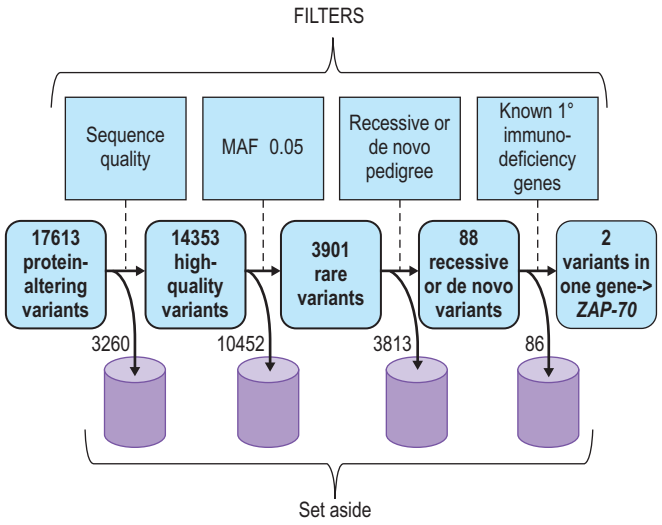


FIG 18.3 Variant Filtering Used to Sort Through Whole-Exome Sequence to Analyze a Familial Disease. Protein-altering variants detected in the affected children, their unaffected sibling, and parents were first filtered by sequencing quality and allele frequency to yield a set of 88 rare, damaging or non-synonymous variants that fit an autosomal recessive inheritance pattern and were then candidates for causing the autoimmune disorder. This group was then further analyzed for genes involved in immune function, which brought two variants in *ZAP-70*, for which the affected children were both compound heterozygotes, to the top of the list as possible causative mutations. (Adapted from Chan A, Punwani D, Kadlecek T, et al. A novel human autoimmune syndrome caused by combined hypomorphic and activating mutations in ZAP-70. *J Exp Med*. 2016;213:155–165, with permission.)

INTERPRETING VARIANTS DISCOVERED THROUGH GENE SEQUENCING

Once a gene and certain variants within it are implicated in causing a particular disease, how is the information applied in clinical diagnostics when the next patient presents with a similar clinical picture and a different, novel variant in that gene? Variation is common and not all variants cause disease. In 2015, much-needed rules were promulgated to measure and standardize the evidence needed to assert that a previously unknown variant is responsible for a disease.[27] The rules stipulated that, by combining five kinds of evidence (Table 18.2), all

TABLE 18.2 Interpreting Variants Found During Clinical Diagnostic Testing

Data Category	In Favor of Benign Score	In Favor of Pathogenic Score
Frequency in the population	• Too frequent, given disease incidence	• Rare and enriched in affected vs. unaffected individuals
Type of alteration in the DNA sequence	• Synonymous single nucleotide variant (SNV) • Del/Dup or indel without frame shift	• Stop gain (nonsense) • Canonical splice site disruption • Gene deletion or disruption
Functional assays in vitro and in model organisms	• No effect on protein expression, enzymatic activity, cellular function, or animal phenotype	• Deleterious effect on protein expression, enzymatic activity, cellular function, or animal phenotype
Segregation pattern in one or more families	• Not present in everyone affected • Absent from some affected individuals	• Co-inherited with disease in all affected individuals and usually not present in unaffected (unless there is incomplete penetrance)
De novo vs. inherited		• Present in affected child but not in either parent (and paternity/maternity confirmed)
Other kinds of data	• In silico predictors • Labelled as B (benign) in other databases	• In silico predictors • Labelled as P (pathogenic) in other databases • Highly suggestive patient phenotype and present with a known pathogenic variant on the other chromosome (for autosomal recessive inheritance)

Adapted from Richards S, Aziz N, Bale S, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet in Med*. 2015;17:405–424.

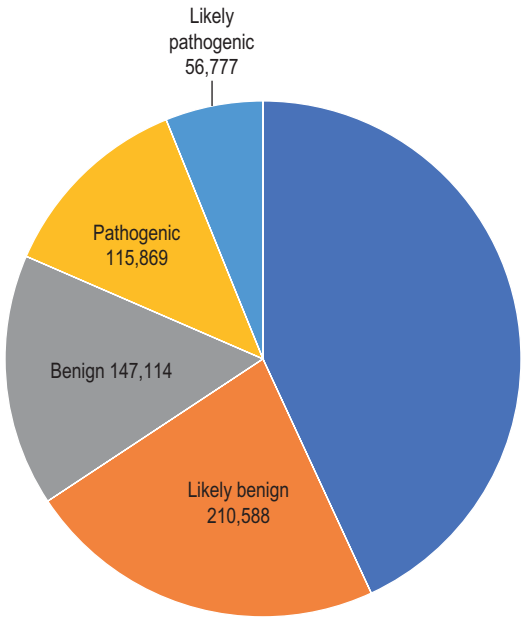


FIG 18.4 Distribution of >930,000 Variant Records in ClinVar by Interpretation. Fully 43% of submitted variants are currently interpreted as variant of uncertain significance (VUS). (According to Richards S, Aziz N, Bale S, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet in Med*. 2015;17:405–424. Data derived from ClinVar Data Miner at https://clinvarminer.genetics.utah.edu/variants-by-significance, accessed January 25, 2021.)

variants found during clinical testing were to be assessed for pathogenicity and reported as belonging to one of five categories: pathogenic (P), likely pathogenic (LP), likely benign (LB), benign (B), or of uncertain significance. The term "likely" is used to indicate a lack of certainty but with odds of the variant being pathogenic or benign, for LP or LB, respectively, being at least 10:1. Clinical laboratories assess each available piece of evidence and weigh its strength; they then combine all of the levels of evidence to come to an interpretation. Unfortunately, many rare variants are too infrequent to reach B or LB status, while for other variants information is lacking or is contradictory. Any such variant is reported as a variant of uncertain significance (VUS), which under current guidelines is considered non-diagnostic and not to be used in most cases to make a diagnosis or to direct medical management. In ClinVar, for example, there are 1.3 million variants submitted by hundreds of laboratories from around the world, discovered primarily in the course of clinical diagnostic testing.[28] Fully 43% of unique variants currently in ClinVar were called a VUS by at least one submitter. However, many variants initially interpreted as VUSs have over time accumulated enough evidence to allow them to be reinterpreted to be concordant in ClinVar. When reinterpretation becomes possible, ~75% of VUSs are reinterpreted as B or LB, while ~25% are upgraded to P or LP, depending on clinical area and the nature of the VUS (i.e., missense, splice site)[29] (Fig. 18.4).

ON THE HORIZON

- With falling costs and rapid turnaround times, whole-genome sequencing will allow detection of mutations in noncoding regulatory regions as well as exons.
- Data storage and computer speed will need to be improved to handle 50-fold more data from each individual.
- New analytical paradigms for the analysis of variation in noncoding DNA and the interpretation of DNA variants that are not clearly associated with deleterious effects will be required.
- Better catalogues of sequences from diverse racial/ethnic backgrounds will be necessary to distinguish pathogenic changes from natural, non-disease-causing variants in individuals other than Northern Europeans.
- Proving causality of new gene variants causing immune disorders will depend on future cataloguing of deleterious mutations in each gene and on molecular studies of the direct effects of the variants.

FUTURE DIRECTIONS

The application of WES and WGS to rare mendelian disorders, begun in 2009, has greatly accelerated disease gene discovery. As of 2019, over 3500 parings of genotypes with disease phenotypes had been found, including several hundred single-gene immune disorders.[5,30] New discoveries will continue, and we can anticipate that the number of known mendelian disorders, including those characterized by primary immunodeficiency and/or autoimmunity, will continue to grow. While WES has been a useful strategy, future diagnostic sequencing for immune disorders may start with rapid, low-cost, optimized panels of genes, followed by WGS if a pathogenic genotype is not readily identified.

Genetic and genomic analysis will not only continue to potentiate discovery of inborn errors of immunity and, in turn, deepen our knowledge of immune system networks but also the knowledge gained will directly lead to new treatments, using gene addition, gene editing, and epigenetic modifiers therapeutically.

However, the genetic basis for complex disorders that show increased familial occurrence, and yet are not single-gene disorders and do not follow mendelian inheritance, will require further technological and analytical tools to be fully unraveled. Increased genomic data from humans and model organisms will help elucidate these more challenging human immune disorders.

REFERENCES

1. International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature*. 2004;431:931–945.
2. Steinberg K, Schneider V, Graves-Lindsay T, et al. Single haplotype assembly of the human genome from a hydatidiform mole. *Genome Res*. 2014;24:2066–2076.
3. https://www.genome.gov/About-Genomics/Introduction-to-Genomics. Accessed 25.01.21.
4. Tangye S, Al-Herz W, Bousfiha A, et al. Human Inborn Errors of Immunity: 2019 update on the classification from the International Union of Immunological Societies Expert Committee. *J Clin Immunol*. 2020;40:24–64.
5. Chinn I, Chan A, Chen K, et al. Diagnostic interpretation of genetic studies in patients with primary immunodeficiency diseases: A working group report of the Primary Immunodeficiency Diseases Committee of the American Academy of Allergy, Asthma & Immunology. *J Allergy Clin Immunol*. 2020;145:46–69.

6. Gonzaga-Jauregui C, Lupski J, Gibbs R. Human genome sequencing in health and disease. *Annu Rev Med*. 2012;63:35–61.
7. Madsen B, Villesen P, Wiuf C. Short tandem repeats in human exons: a target for disease mutations. *BMC Genomics*. 2008;9:410.
8. Carvalho C, Lupski J. Mechanisms underlying structural variant formation in genomic disorders. *Nat Rev Genet*. 2016;17:224–238.
9. Goodwin S, McPherson J, McCombie W. Coming of age: ten years of next-generation sequencing technologies. *Nat Rev Genet*. 2016;17: 333–351.
10. Morrow B, McDonald-McGinn D, Emanuel B, et al. Molecular genetics of 22q11.2 deletion syndrome. *Am J Med Genet A*. 2018;176:2070–2081.
11. Puig M, Casillas S, Villatoro S, et al. Human inversions and their functional consequences. *Brief Funct Genomics*. 2015;14:369–379.
12. Amendola L, Jarvik G, Leo M, et al. Performance of ACMG-AMP variant-interpretation guidelines among nine laboratories in the Clinical Sequencing Exploratory Research Consortium. *Am J Hum Genet*. 2016;98:1067–1076.
13. Tryka K, Hao L, Sturcke A, et al. NCBI's Database of Genotypes and Phenotypes: dbGaP. *Nucleic Acids Res*. 2014;42:D975–D979.
14. Landrum M, Lee J, Benson M, et al. ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res*. 2016;44: D862–D868.
15. Harrison S, Riggs E, Maglott D, et al. Using ClinVar as a resource to support variant interpretation. *Curr Protoc Hum Genet*. 2016;89: 8.16.1–8.16.23.
16. Fokkema IF, Taschner P, Schaafsma G, Celli J, Larow J, den Dunnen J. LOVD v.2.0: the next generation in gene variant databases. *Hum Mutat*. 2011;32:557–563.
17. Revy P, Buck D, le Deist F, de Villartay J. The repair of DNA damages/modifications during the maturation of the immune system: lessons from human primary immunodeficiency disorders and animal models. *Adv Immunol*. 2005;87:237–295.
18. Kovanen P, Leonard W. Cytokines and immunodeficiency diseases: critical roles of the gamma(c)-dependent cytokines interleukins 2, 4, 7, 9, 15, and 21, and their signaling pathways. *Immunol Rev*. 2004;202:67–83.
19. Choi Y, Simon-Stoos K, Puck J. Hypo-active variant of IL-2 and associated decreased T cell activation contribute to impaired apoptosis in autoimmune prone MRL mice. *Eur J Immunol*. 2002;32:677–685.
20. Diehl AG, Boyle AP. Deciphering ENCODE. *Trends Genet*. 2016;32:238–249.
21. https://gtexportal.org/home/. Accessed 25.01.21.
22. The ENCODE Project Consortium, Moore J, Purcaro M, et al. Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature*. 2020;583(7818):699–710.
23. Al-Herz W, Naguib K, Notarangelo L, et al. Parental consanguinity and the risk of primary immunodeficiency disorders: report from the Kuwait National Primary Immunodeficiency Disorders Registry. *Int Arch Allergy Immunol*. 2011;154:76–80.
24. Currier R, Puck J. SCID newborn screening: What we've learned. *J Allergy Clin Immunol*. 2021;147(2):417–426.
25. Yang Y, Muany D, Xia F, et al. Molecular findings among patients referred for clinical whole-exome sequencing. *JAMA*. 2014;312:1870–1879.
26. Chan A, Punwani D, Kadlecek T, et al. A novel human autoimmune syndrome caused by combined hypomorphic and activating mutations in ZAP-70. *J Exp Med*. 2016;213:155–165.
27. Richards S, Aziz N, Bale S, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med*. 2015;17:405–424.
28. https://www.ncbi.nlm.nih.gov/clinvar/submitters/. Accessed 25.01.21.
29. Harrison S, Rehm H. Is 'likely pathogenic' really 90% likely? Reclassification data in ClinVar. *Genome Med*. 2019;11:72 PMID: 31752965.
30. Posey J, O'Donnell-Luria A, Chong J, et al. Insights into genetics, human biology and disease gleaned from family based genomic studies. *Genet Med*. 2019;21:798–812.

19

Regulatory RNA in Immunologic Diseases

Thomas M. Aune

Most would agree that the biologic complexity of humans far exceeds the biologic complexity of worms, as exemplified by intellectual, physical, and behavioral capacities: the numbers of discrete organ systems and cell lineages, each with unique biologic functions (Fig. 19.1). A corollary is that humans also develop many more diseases of greater diversity and complexity than worms. However, both humans and worms have about the same number of protein-coding genes (~20,000) in their genomes and the proteins they encode largely carry out similar biologic functions. This has led to the question: where does the information that programs the greater biologic complexity of humans reside? Protein-coding genes in humans take up approximately 2% of the human genome. Thus, one possibility is that this information is encoded in the non-protein-encoding portion of the genome, which is 30-fold larger in humans than in the worm. More than 75% of the human genome is transcribed into RNA, in some human cell lineage at some stage of human development. This led to the hypothesis that these additional transcribed noncoding RNAs (ncRNAs) contribute to the differences in biologic complexity between humans and other species.[1–4] A corollary is that this vast array of transcribed ncRNAs might also contribute to the greater diversity and complexity of human disease.

KEY CONCEPTS

- Only ~2% of the human genome is used to transcribe protein-coding genes,
- >70% of the human genome is transcribed into noncoding RNA (ncRNA), with regulatory functions,
- Newly discovered classes of ncRNA include long noncoding RNAs (lncRNAs), microRNAs (miRNAs), and enhancer RNAs (eRNAs).
- ncRNAs impact critical aspects of mRNA and protein biology via multiple mechanisms.
- ncRNAs are new "master regulators of gene expression and cell fate," and help confer biologic complexity into the human genome.

CLASSES OF REGULATORY RNAS

Classes of regulatory RNAs include long noncoding RNAs (lncRNAs), enhancer RNAs (eRNAs), and microRNAs (mi RNAs).[5–11] lncRNA genes look like protein-coding genes and many, but not all, have exon-intron structures. By definition, lncRNAs are more than 200 nucleotides in length. This definition by length distinguishes lncRNAs from miRNAs and other small RNAs. Most lncRNAs are 5′ capped and polyadenylated, similar to most mRNAs. In contrast to the similarity in numbers of protein-coding genes, human genomes contain 20 times more lncRNA genes than do worm genomes. However, lncRNAs are not transcribed into proteins. They carry out their biologic functions as RNAs (although there are exceptions to this rule). Current classification or naming of lncRNAs largely draws from locations and orientations of lncRNA genes in the genome relative to neighboring protein-coding genes. These include intergenic lncRNAs if located between two protein-coding genes, intronic if located largely in a single intron of one protein-coding gene, and antisense if transcribed from the opposite DNA strand as the neighboring protein-coding gene. This latter category also includes divergent lncRNA:mRNA pairs that are transcribed from opposite DNA strands but have transcriptional start sites that are typically within 1000 bp. Proximity suggests shared promoters and enhancers. It is worth noting that while these classifications provide information regarding their position in the genome relative to protein-coding genes, they do not necessarily provide information regarding lncRNA function or mechanism of action.

Several general classifications can also be made regarding functions of lncRNAs.[5,8,12] First, some lncRNAs can rewrite the histone code at target protein-coding gene loci by recruiting epigenetic machinery to a locus. This often occurs *in cis* with the lncRNA gene and an adjacent protein-coding gene. The lncRNA rewrites the histone code at promoters and enhancers of the adjacent protein-coding gene by producing either activating or inhibitory histone marks to either stimulate or reduce mRNA expression, respectively (Fig. 19.2A). Many lncRNAs target only one protein-coding gene, but others can target multiple protein-coding gene loci. Multitargeting oftentimes involves gene families, such as the globin gene locus, the Hox gene locus, or the interleukin (IL)4-IL5-IL13 Th2 cytokine gene locus. LncRNAs can also regulate expression of genes on different chromosomes. Although these genes may not share a chromosome, they may encode proteins that are involved in common biologic processes.[5,12,13]

Second, certain lncRNAs bind and thus inhibit miRNA or mRNA function. Binding of miRNAs can lead to increased expression of protein-coding genes targeted by a given miRNA. LncRNAs can act on mRNAs to modulate pre-mRNA splicing, interfere with translation or alter mRNA degradation rates (see Fig. 19.2B).

Third, lncRNAs can bind transcription factors to either aid or prevent the recruitment of transcription factors to promoters and enhancers, thereby either increasing or reducing mRNA transcription (see Fig. 19.2C).

Last, lncRNAs also tether individual proteins that work together in common biologic pathways to affect their function or prevent protein-protein interactions (see Fig. 19.2D). Undoubtedly, additional mechanisms by which lncRNAs function will be added to this list as we continue study of this diverse class of RNAs.

Differences between human and worm genomes

	humans	worms	fold difference
genome size (bp)	3,000,000,000	100,000,000	30X
# protein-coding genes	20,000	20,000	1X
# long noncoding RNA genes	60,000	3,000	20X
# enhancer RNAs	400,000	5,000	80X
# microRNAs	2,000	200	10X

FIG. 19.1 Biologic Complexity and Genome Size. Illustration of humans *(upper left)* and roundworms (*Caenorhabditis elegans*, *upper right*). Differences in genome size, number of protein-coding genes, long noncoding RNAs (lncRNAs), enhancer RNAs (eRNAs), and microRNAs (miRNAs) between humans and worms *(lower)*.

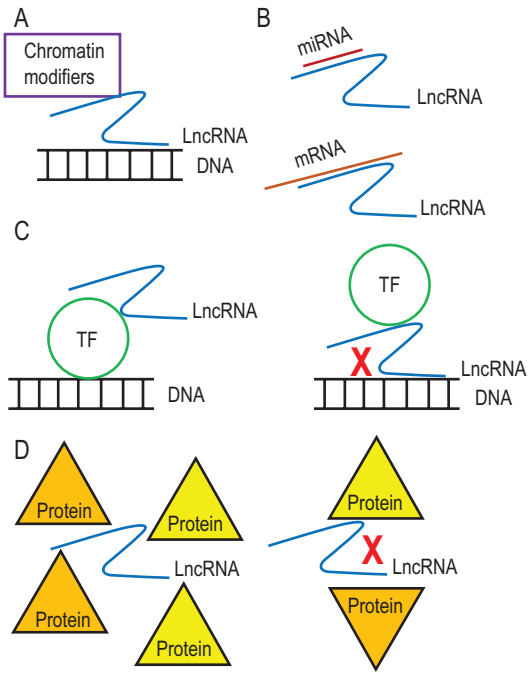


FIG. 19.2 Common Functions of Long Noncoding RNAs *(lncRNAs)*. LncRNAs (A) recruit chromatin modifiers to target DNA, (B) act as sponges and bind *microRNAs (miRNAs)* or mRNAs, (C) bind to transcription factors *(TFs)* to tether them to DNA or block transcription factor binding to DNA, and (D) facilitate or prevent protein-protein interactions.

LNCRNAS AND THE DEVELOPMENT AND FUNCTION OF THE IMMUNE SYSTEM

Individual lncRNAs regulate many key immune processes. This includes the development and function of both innate and adaptive immunity.[12,14] Some examples include lncHSC-1, lncHSC-2, H19, and others that regulate the differentiation of hematopoietic stem cells and progenitor cells into mature hematopoietic cell lineages, as well as the self-renewal of both short- and long-term hematopoietic stem cells. These processes are necessary to maintain numbers of all the mature hematopoietic cell lineages in the periphery. They are also required to sustain both the innate and adaptive arms of the immune system. These functions are achieved by regulation of both protein-coding genes critical for establishing and maintaining these developmental programs and the activity of key transcription factors that guide hematopoietic differentiation programs.

LncRNAs also regulate critical aspects of the innate immune system (Chapter 3). This includes the development of macrophages and dendritic cells from common myeloid progenitor cells (lnc-DC); the lifespan of short-lived myeloid cells, eosinophils, and neutrophils (Morrbid); and the expression of proinflammatory mediators, such as cytokines. Deletion of these regulatory lncRNAs seriously disrupts the function of the innate immune response in a variety of animal models.

LncRNAs also regulate key functions of the adaptive immune response. Many of these studies have focused on the process of differentiation of naïve CD4 T cells to the different T helper lineages, Th1, Th2, and Th17, which are defined by the cytokines they express (Chapter 11). LncRNAs have been identified that regulate expression of GATA3, the key transcriptional regulator required for Th2 lineage commitment. The interferon-gamma-antisense 1 (IFNG-AS1) lncRNA regulates expression of interferon-γ (IFN-γ) in Th1 cells and TH2LCRR regulates IL-4, IL-5, and IL-13 in Th2 cells.[13,15,16] Conversely, other lncRNAs (e.g., linc-MAF-4 and lncRNA-CD244) can negatively regulate T helper lineage commitment by inhibiting expression of transcription factors or stimulating development of opposing lineages.[5,17] By these opposing functions, lncRNAs may guide creation of the proper balance of these critical T helper cell lineages whose primary functions are to orchestrate the adaptive immune response, which is critical for establishing and maintaining lifelong immunity

to pathogens. Studies such as these in the field of immunology as well as other fields have led to the labeling lncRNAs as the "emerging master regulators of gene expression and cell fate."[18]

LNCRNAS AND IMMUNE-MEDIATED DISEASE

Studies have also emerged that associate lncRNA expression and, in some cases, downstream function, to human immune-mediated diseases.[19–21] Following completion of the Human Genome Project and identification and annotation of lncRNA genes, it became possible to use high-throughput methods, such as microarrays or whole-genome RNA sequencing (RNA-seq), to ascertain the expression levels of many, if not all, annotated lncRNAs using fairly standard computational approaches. These same RNA-seq methods could also be used to identify novel or not previously identified lncRNAs. These methods enabled examination of the contribution of lncRNAs to the increased genomic and biologic complexity of humans and the immune-mediated diseases they manifest. In order to illustrate how lncRNA dysregulation may contribute to immune-mediated disease, specific examples are provided below (Table 19.1).

CLINICAL RELEVANCE

- The magnitude of dysregulation of noncoding RNAs (ncRNAs) in immune-mediated disease is vast.
- NcRNA dysregulation disrupts cell phenotypes through the loss of negative regulators and the gain of positive regulators.
- Genetic variants linked to immune-mediated disease preferentially target ncRNA coding regions of the genome.
- It is likely that many immune-mediated diseases reflect as yet undefined dysregulation of ncRNA.

LINKS BETWEEN IFNG-AS1 AND BRUCELLOSIS, INFLAMMATORY BOWEL DISEASE, AND SYSTEMIC LUPUS ERYTHEMATOSUS

IFN-γ is produced by effector and memory CD4 and CD8 T cells and by natural killer (NK) and natural killer T (NKT) cells (Chapter 12), but not by naïve CD4 and CD8 T cells (Chapter 9). This cytokine plays a critical role in regulating the immune response to intracellular infections. It is a major activator of macrophage functions (e.g., phagocytosis and antigen presentation) (Chapter 2), it plays a critical role in stimulating antiviral responses by CD8 cytotoxic T cells (Chapter 12), and it possesses antiviral properties (Chapter 25). Given its proinflammatory nature and results from various immune models, IFN-γ can play critical roles in human immune-mediated diseases (Chapter 14). Induction of the lncRNA IFNG-AS1 (Tmevpg1, NeST) plays a key role during the development of these immune response programs by guiding the writing of the epigenetic code at the IFNG locus. It acts to create a positive transcriptional environment, which is required for efficient production of IFN-γ by these diverse cell types during their response to external stimuli.[5,18]

Levels of IFNG-AS1 have been shown to be elevated in a number of human immune-mediated diseases. For example, in brucellosis, which is the most common zoonotic infection in humans, elevated IFNG-AS1 levels during infection correlate with increased IFN-γ expression.

TABLE 19.1 Examples of lncRNAs Dysregulated in Immune-Mediated Disease

lncRNA	Disease	Level in Disease	Normal Function
lincRNA-p21	RA	Decreased, restored by MTX	Inhibits NF-κB activity
H19	RA	Increased	Increases NF-κB, IL-6, TNF
HOTAIR	RA	Decreased in synoviocytes	Increases MMP-2, MMP-13
HOTAIR	RA	Increased in monocytes	Macrophage activation/migration
MALAT1	RA	Increased in synoviocyte	Apoptosis
C5T1lncRNA	RA	Increased	Increases C5
GATA3-AS1	Allergy	Increased	Induces GATA3/Th2 responses
IFNG-AS1	Brucellosis	Increased	Induces IFN-γ
	IBD	Increased	Induces
	UC	Increased	Induces
	SLE	Increased	Induces
Lnc13	CeD	Decreased	Represses CeD-related genes
LINC00305	CAD	Increased	Increases NF-κB activity
ANRIL	CAD/T2D	Increased	Induction of IL-6 and IL-8 by TNF
Linc00513	SLE	Increased	Increases IFN pathway
RP11-2B6.2	SLE	Decreased	Inhibits SOCS1, neg. regulator of IFN pathway
NEAT1	SLE	Increased	Positive regulator TLR-mediated pathways

CAD, Coronary artery disease; *CeD*, Celiac disease; *IBD*, inflammatory bowel disease; *IFN*, interferon; *IL*, interleukin; *MMP*, matrix metalloproteinase; *MTX*, methotrexate; *NF-κB*, nuclear factor kappa B; *RA*, rheumatoid arthritis; *SLE*, systemic lupus erythematosus; *T2D*, type 2 diabetes; *TNF*, tumor necrosis factor. *TLR*, Toll-like receptor; *UC*, ulcerative colitis.

In Crohn's disease and ulcerative colitis, which are the most common forms of inflammatory bowel disease (IBD) (Chapter 75), colon biopsy samples exhibit increased expression of IFNG-AS1. This suggests that the IFNG-AS1/IFNG axis may also play an important role in this organ-specific immune-mediated disease. In support of this hypothesis, genome-wide association studies (GWAS) have identified genetic variants in the IFNG-AS1/IFNG locus that are linked to increased risk for developing IBS. These genetic risk factors are linked to altered expression of lncRNA and of cytokine genes within this region.[20]

A similar case is seen in systemic lupus erythematosus (SLE) (Chapter 52). GWAS studies have identified genetic risk factors localized to the IFNG-AS1/IFNG genomic region. Studies have shown that the expression of genes within this locus, as well as the lncRNA IFNG-AS1, are dysregulated in SLE.

GATA3-AS1 AND ALLERGY

Widespread dysregulation of lncRNAs has also been observed in allergy and asthma (Chapter 43). One example is the lncRNA GATA3-AS1. GATA3-AS1 is critical for effector T helper cell differentiation programs (Chapter 11).[16] The GATA3 gene and GATA3-AS1 transcriptional start sites are positioned about

1000 bp apart in the human genome. GATA3-AS1 is transcribed from the 3′ DNA strand and the GATA3 mRNA is transcribed from the 5′ strand, thus meeting the definition of a divergent lncRNA/mRNA gene pair. GATA3-AS1 acts as a guide to rewrite the epigenetic code at the GATA3 gene promoter and enhancer loci, thus creating a positive transcriptional environment to induce expression of GATA3. As a transcription factor, GATA3 is considered to be the master regulator of Th2 differentiation and the expression of Th2 cytokines. Not only is GATA3-AS1 required for expression of GATA3 but also it is required for downstream expression of Th2 cytokines.

In general terms, allergy and asthma are considered to be Th2-mediated diseases. In allergic individuals, GATA3-AS1 levels are elevated and GATA3-AS1 is also specifically induced by allergen exposure. Taken together, this suggests a model where GATA3-AS1 contributes to early stages of allergy by promoting Th2 differentiation and function. Responses by differentiated Th2 effector cells are thought to play important roles in the induction of allergic and asthmatic responses and associated pathologies.

LNCRNAS, NUCLEAR FACTOR KAPPA B (NF-KB), AND THE PATHOGENESIS OF RHEUMATOID ARTHRITIS

Three unique biologic functions have been ascribed to LincRNA-p21.[5,21] LincRNA-p21 was originally identified because it plays a critical role in executing the cellular response to DNA damage. This lncRNA also regulates the Warburg effect, which refers to how, in order to generate energy, cancer cells seem to prefer glycolysis rather than the more efficient oxidative phosphorylation pathway, even under aerobic conditions. By regulating levels of a transcription factor that is key to governing cellular responses to hypoxia, hypoxia inducing factor 1-α (HIF-1α) the balance is shifted away from oxidative phosphorylation to glycolysis. A third function of lincRNA-p21 is to bind to certain mRNAs via a base-pairing mechanism and thus interfere with translation of these mRNAs into proteins. Specific examples include blocking translation of RELA and JUN mRNA. RelA protein is a necessary component of the key proinflammatory transcription factor NF-κB, whereas Jun protein is a key component of the AP-1 transcription factor. AP1 also provides critical functions in a variety of cellular responses, including responses to stress.

Levels of lincRNA-p21 are severely depressed in rheumatoid arthritis (RA) (Chapter 53). Basal and induced activities of NF-κB are elevated in RA, suggesting an inverse relationship between lincRNA-p21 levels and NF-κB activity, which has been demonstrated experimentally. In support of this idea, lincRNA-p21 and NF-κB activity normalize in RA patients on methotrexate, which is one of the most common therapies for RA. Further studies support a model where methotrexate therapy directly induces increased expression of lincRNA-p21. LincRNA-p21 is able to bind to RELA mRNA and inhibit its translation, thus reducing basal and stimulus-induced levels of NF-κB activity. In this regard, lincRNA-p21 plays a critical antiinflammatory function. Loss of lncRNA-p21 appears to be an important contributor to proinflammatory responses mediated by NF-κB. In addition to RA pathogenesis, this suggests that lncRNA-p21 may also contribute to the pathogenesis of other autoimmune disorders, especially those in which elevated NF-κB activity plays a central role.

In addition to lncRNA-p21, levels of the lncRNA H19, originally described for its function in hematopoiesis, are also markedly elevated in RA synovium. Elevation of H19 results in increased NF-κB activity and induction of the proinflammatory cytokines IL-6 and tumor necrosis factor (TNF), both of which are NF-κB inducible. H19 controls expression of an imprinted network of genes via epigenetic modifications. H19 lncRNA also contains miRNA members of the miR-675 family. Thus, H19 may directly or indirectly influence NF-κB proinflammatory activity via either alterations in imprinting networks or expression of the miRNAs embedded within it.[21]

Other lncRNAs impact overall NF-κB activity by different mechanisms. NKILAe inhibits IKB phosphorylation, and thus dampens both basal levels and stimulus-induced NF-κB activity. Lethe inhibits binding of the NFκB transcription factor complex to target DNA transcriptional elements, thus inhibiting induction of NF-κB transcriptional response genes. Although these lncRNAs have not been shown to exhibit different expression levels in RA, age-related loss of Lethe expression has been described in mice. Thus, Lethe may be a natural negative regulator of proinflammatory NF-κB function. It is possible that its loss could contribute to age-related increases in NF-κB proinflammatory activity in the immune system.[21]

The lncRNA HOTAIR was originally described to play a critical role in the epigenetic silencing of the HOXD gene locus during early development. HOXD plays an important role in the epigenetic regulation of morphogenesis—for example, how the skin develops over the surface of the body. HOTAIR is also expressed in fully differentiated cells, although its function in these cells is less well understood. Importantly, HOTAIR is expressed at elevated levels in certain cancers and appears to play an important role in cancer and cancer metastasis. This is an active area of investigation. In RA synovium, HOTAIR is expressed at reduced levels and is associated with increased expression of the matrix metalloproteinases (MMP)-2 and MMP-13, which may contribute to tissue destruction. Conversely, HOTAIR is expressed at elevated levels in monocytes/macrophages isolated from RA patients. Elevated HOTAIR levels appear to increase macrophage activation and migration, also potentially contributing to RA inflammation.[21]

In summary, many of the lncRNAs that exhibit dysregulation in RA have been shown to impact the activity of the critical proinflammatory transcription factor complex NF-κB and to influence the expression levels of known NF-κB target genes that are recognized to play important roles in RA pathogenesis.

LNCRNAS AND SYSTEMIC LUPUS ERYTHEMATOSUS

Many autoimmune diseases exhibit increased expression of genes known to be induced by the type 1 interferons IFN-α and IFN-β, which is referred to as the *IFN signature*. This is most notable in SLE but is also seen in other autoimmune diseases, especially at times of increased disease activity. The major stimuli leading to induction of interferons and downstream interferon-stimulated genes are viral infections. Origins of this IFN signature in autoimmune disease in the apparent absence of viral infection are incompletely understood. Positive regulators of the type 1 interferon signaling pathway are among the lncRNAs whose levels are elevated in SLE. One lncRNA, named linc00513, has been shown to increase the phosphorylation of Stat1 and Stat3, two critical positive transcriptional regulators of the interferon signaling pathway. Expression levels of this lncRNA are linked to genetic variants that are themselves linked to SLE disease risk.

A second lncRNA, named RP11-2B6.2, inhibits the activity of SOCS1, a natural negative regulator of the interferon signaling pathway. Given that a cardinal feature of presence of SLE as well as disease activity and severity is increased interferon signaling, dysregulation of interferon-regulating lncRNAs may play an even greater role in SLE than has yet been demonstrated.[22] Further studies will be required to determine whether these and other lncRNAs that control the interferon signaling pathway influence other autoimmune diseases that also exhibit an excessive IFN signature.

MIRNAS

MicroRNAs (abbreviated *miRNAs*) are another important class of ncRNAs.[10] They are transcribed from genes into a long precursor miRNA that is less than 200 nucleotides long. A specific enzyme complex, the Drosha-Dgcr8 micoprocessor complex, digests the precursor miRNA to produce a hairpin intermediate. This miRNA intermediate is exported to the cytoplasm where Drosha, a RNase III—like nuclease, processes the miRNA intermediate into its mature form. Mature miRNAs are bound by argonaute (Ago) proteins and act as guide sequences that bind to typically short 6 to 8 complementary nucleotide sequences on target mRNAs, usually located in the 3′ untranslated region (3′ UTR). If complementarity is perfect between the miRNA and the target mRNA sequences, this triggers what is termed *RNA interference*, resulting in digestion of the mRNA by the RNase H-like activity present in one of the Ago domains. However, this appears to occur only rarely in mammals. The current general view is that miRNAs predominantly cause deadenylation and decay of mRNA.

Initiation and elongation of translation of proteins may also be inhibited by miRNAs. Approximately 2000 distinct miRNA genes have been identified in the human genome. However, individual miRNAs can target multiple protein-coding or lncRNA genes, so their impacts on cell identity and phenotype are quite large. Studies have demonstrated important roles for miRNAs in multiple aspects of development, differentiation, and function of both the innate and adaptive arms of the immune system.

In addition to the intracellular function of miRNAs, recent studies have found the presence of miRNAs in many different extracellular fluids. Originally it was thought that these extracellular miRNAs were simply released from dying or dead cells and tissues. Thus presence of mRNAs in extracellular fluids were considered to lack any real biologic significance. However, more recently it has been found that these extracellular miRNAs are actually packaged in small lipid bilayer vesicles termed *exosomes*. This has led to the view that miRNAs are packaged into exosomes inside cells, released into the extracellular space, and then transported to and taken up by distant recipient cells. As such, these exosomal miRNAs may play a kind of messenger role enabling cell-to-cell communication.

As with lncRNA investigations, implementation of high- dimensional techniques such as RNA-seq or miRNA-specific microarrays to identify expression levels of miRNAs in case/control studies has demonstrated substantial dysregulation of multiple miRNAs in immune-mediated disease.[23] What is notable regarding these studies is that miRNAs dysregulated in a given autoimmune disease are known important regulators of critical molecular and cellular pathways that are thought to contribute to pathology associated with a given autoimmune disease. For example, as with lncRNAs described above, miRNAs that stimulate the type 1 interferon signaling pathway are upregulated in SLE, and miRNAs that inhibit the type 1 interferon signaling pathway are downregulated in SLE. A similar case is observed in RA. Expression levels of specific miRNAs whose natural function is to inhibit cell cycle progression and reduce expression of inflammatory cytokines are depressed in RA, and this might contribute to the hyperproliferation and increased levels of activation exhibited by synovial cells in RA. Expression levels of other miRNAs whose normal functions are to promote activity of the proinflammatory transcription factor NF-κB and stimulate expression of proinflammatory cytokines, are elevated in RA and this is also likely to contribute to elevated inflammatory processes that drive RA pathogenesis. Similar observations have been made in other autoimmune diseases including psoriasis, multiple sclerosis, and systemic sclerosis. Thus, miRNAs play critical roles in determining expression levels of protein-coding genes and, as with lncRNAs, contribute to development of the many diverse cell lineages present in humans. A cost may be that dysregulation of miRNA expression may make significant contributions to the many complex diseases found in humans, including immune-mediated diseases.

Enhancer RNAs

Enhancers are conserved DNA elements whose function is to serve as binding sites for transcription factors and enhance transcription of target genes.[11] In the genome, enhancers can be near to their target genes, many kilobases away, or even on separate chromosomes. They can influence transcription of one or multiple genes. Transcriptional enhancers have conserved DNA motifs that recruit transcription factors, which in turn recruit the epigenetic machinery to enhancer loci to write the epigenetic code culminating in recruitment of RNA polymerase II to the enhancer locus. The general view is that enhancers function by changing the conformation of the chromosome and recruiting bound RNA polymerase II to their target gene loci. This recruitment will enhance transcription. RNA polymerase II can also use enhancer DNA as the template to produce eRNAs.

One class of eRNAs, termed unidirectional eRNAs, are greater than 200 nt in length, transcribed from one DNA strand, polyadenylated and fairly stable, making them, by definition, lncRNAs. Functions of unidirectional eRNAs are incompletely understood. However, these unidirectional eRNAs seem to re-enforce enhancer function by promoting recruitment of transcription factors to enhancers, recruiting histone acetyltransferases to enhancers to write the epigenetic code, increasing enhancer chromatin accessibility, and looping distal enhancers to promoters. All of these effects can serve to stimulate transcription of target genes.

Expression of unidirectional eRNAs may be inferred from RNA-seq experiments using such criteria as (1) the RNA transcript is greater than 200 nt in length and nonspliced, (2) the transcript does not map to a known lncRNA or mRNA gene, and (3) the transcript maps to a known transcriptional enhancer. Studies examining unidirectional eRNAs in immune-mediated diseases are quite limited, but they have shown that the level of dysregulation of unidirectional eRNAs in multiple autoimmune diseases far exceeds the level of dysregulation of either annotated lncRNAs or mRNAs, both in terms of degree of differences in expression as well as the numbers of this class of ncRNA that are dysregulated.[19] This raises the possibility that this dysregulation may contribute to the origins and pathogenesis of multiple

immune-mediated diseases. However, interpreting the significance of these results is difficult. It is not yet known if the differential expression of a given unidirectional eRNA may directly contribute to disease pathogenesis or if differences in eRNA expression simply reflect differences in underlying enhancer function at a different level. These factors could include the sequence of the enhancer DNA, differences in the binding of one or more transcription factors, or reading and writing the epigenetic code. It must be recognized that our lack of knowledge is due in part to the fact that we do not fully understand the contribution of unidirectional eRNAs to enhancer function and to the stimulation of target gene transcription.

GENOME-WIDE ASSOCIATION STUDIES

With the goal of identifying genetic variants, single nucleotide polymorphisms (SNPs) and GWAS have been performed for almost all immune-mediated diseases (Chapter 18). Several unexpected results have emerged from these studies.[9,24] First, many SNPs (100 s), rather than a single monogenic variant, are associated with risk for developing any given immune-mediated disease. Second, less than 10% of GWAS-identified SNPs are found in protein-coding genes in the genome, which suggests that altered protein amino acid sequence and protein function is not a major contributor to genetic risk for developing immune-mediated diseases. Instead, many of these genetic variants have been found to reside in noncoding regions of the genome. This has hindered progress in understanding how these genetic variants confer disease risk. Studies are now emerging that demonstrate that many of these genetic variants are located in lncRNA genes. Expression differences of many of these encoded lncRNAs have been implicated in the pathogenesis of immune-mediated diseases.[20–24]

One example is IFNG-AS1. The IFNG-AS1 gene is located in a region that is rich with genetic variants linked to IBD risk. Expression of IFNG-AS1 is elevated in this disease at sites of inflammation, and this seems to be associated with increased expression of IFN-γ. These and other studies suggest a general rule that genetic variants identified by GWAS that confer disease risk for developing immune-mediated diseases influence expression levels of lncRNAs in the genome, and altered expression of these lncRNAs contributes to disease risk and disease pathogenesis.

A second group of genetic variants identified by GWAS include those that influence functions of miRNAs.[25] One area of attention has been to examine how genetic variants located in the 3′ untranslated regions (3′ UTR) of protein-coding genes alter miRNA target sites. In this regard, genetic variants in 3′ UTRs may produce new miRNA target sites, resulting in increased binding of miRNAs to the 3′ UTR in question and reduced expression of target protein-coding genes. The alternative is also seen; genetic variants in the 3′ UTR of certain protein-coding genes results in loss of miRNA binding sites, leading to increased expression of these target protein-coding genes. Besides genetic variants in 3′ UTRs of protein-coding genes, SNPs linked to immune-mediated diseases have also been found in miRNA genes that change the ability of the encoded and processed miRNA to bind to its target protein-coding gene 3′ UTR and guide mRNA degradation. In summary, these variants produce altered expression levels of proteins that can alter cell traits and confer disease risk.

A third group of immune-mediated diseases' susceptibility to genetic variants includes transcriptional enhancer regions and associated transcribed eRNAs.[9,11,19] Variants in these genomic regions potentially impact expression of target protein-coding genes by altering enhancer functions. Transcriptional enhancers probably represent the most common site of genetic variation linked to immune-mediated diseases as well as other complex traits. Enhancer function is a very complex process and the roles of eRNAs in transcriptional activation and silencing are incompletely understood. Thus, it is not possible to conclude with certainty that altered enhancer function endowed by these genetic variants actually results from loss or gain of eRNA expression or by other mechanisms. However, it has been clearly demonstrated that the presence or absence of disease-associated SNPs at target enhancer loci is associated with differences in eRNA expression and expression of their target protein-coding genes.

One goal of GWAS has been to identify genetic variants that confer risk of developing a given disease, including immune-mediated diseases. Several common themes emerge.[24] One is that multiple genetic variants confer disease risk and effects of a single genetic variant are relatively small so there is no "smoking gun." Perhaps more importantly, the majority of the genetic variants thus far identified by GWAS function primarily by regulating expression levels of target protein-coding genes, rather than by producing proteins with altered amino acid sequence and functions. Most genetic variants identified by GWAS are localized in the noncoding region of the genome responsible for producing the known array of ncRNAs.

The general view is that immune-mediated diseases arise from a combination of genetics, or family traits, and the environment (Fig. 19.3). Genetic variants thus far identified that contribute to immune-mediated disease are predominantly localized in the enhancer elements that give rise to eRNAs and in genes that are transcribed into lncRNAs and miRNAs. Many of these ncRNAs may be stably expressed in certain cell lineages. Therefore the ncRNA transcriptome determines cellular phenotypes by impacting critical mRNA and protein biology. In this way, the ncRNA transcriptome will define how a given cell lineage responds to environmental cues, causing these responses to be predetermined. If this is considered in the context of immune-mediated disease, one might propose that the ncRNA transcriptome shifts the balance toward heightened immune activity by both loss of ncRNAs that inhibit, and gain

FIG. 19.3 Model describing contributions of regulatory RNAs to immune-mediated disease. *eRNA*, Enhancer RNA; *lncRNA*, long noncoding RNA; *miRNA*, microRNA.

of ncRNAs that stimulate, immune activity, thereby shifting the balance to a proinflammatory state. As such, this would lower the threshold at which an individual responds to both external and internal environmental cues. Since these altered responses are predetermined by the ncRNA transcriptome, they may become permanently imprinted in the host and produce lifelong idiopathic disease.

An added consideration is that certain lncRNAs act *in cis* to regulate neighboring target protein-coding genes, while others act *in trans* to regulate expression of large gene sets that seem able to define the inflammatory state. Similarly, activity of critical immune pathways can be positively or negatively regulated by multiple lncRNAs at multiple levels, including the level of gene transcription, mRNA translation, and protein function. Dysregulation of any of these lncRNAs may significantly impact overall immune function. By extension, lncRNAs may represent targets for therapeutic intervention. As discussed earlier, these targets may include lncRNAs that promote the IFN signature in SLE, NF-κB activity in RA, or heightened Th2 responses in allergy. Currently, use of antisense oligonucleotides to silence overexpressed or restore underexpressed lncRNAs via nanoparticle technologies represents a possible therapeutic approach. Undoubtedly, a better and more complete understanding of how lncRNAs regulate key aspects of both innate and adaptive arms of the immune system, mechanisms of dysregulation in immune-mediated disease, and general principles of how lncRNAs interact with their DNA, RNA, and protein binding partners will offer new therapeutic insights that may be applied to this constellation of disease processes.

ON THE HORIZON

- The search for a unifying logical explanation for the role of noncoding RNA (ncRNA) dysregulation in immune-mediated diseases, i.e., identification of a "smoking gun,"
- Elucidation of the contribution of genetic or familial traits and environment to ncRNA dysregulation.
- The development of ncRNAs as therapeutic targets.

REFERENCES

1. Clark MB, Choudhary A, Smith MA, et al. The dark matter rises: the expanding world of regulatory RNAs. *Essays Biochem*. 2013;54:1–16.
2. Liu GQ, Mattick JS, Taft RJ. A meta-analysis of the genomic and transcriptomic composition of complex life. *Cell Cycle*. 2013;12(13):2061–2072.
3. Morris KV, Mattick JS. The rise of regulatory RNA. *Nat Rev Genet*. 2014;15(6):423–437.
4. Aune TM, Crooke PS, Spurlock CF. Long noncoding RNAs in T lymphocytes. *J Leukocyte Biol*. 2016;99(1):31–44.
5. Rinn JL, Chang HY. Genome regulation by long noncoding RNAs. *Annu Rev Biochem*. 2012;81:145–166.
6. Fatica A, Bozzoni I. Long non-coding RNAs: new players in cell differentiation and development. *Nat Rev Genet*. 2014;15(1):7–21.
7. Guttman M, Rinn JL. Modular regulatory principles of large non-coding RNAs. *Nature*. 2012;482(7385):339–346.
8. Ulitsky I, Bartel DP. lincRNAs: genomics, evolution, and mechanisms. *Cell*. 2013;154(1):26–46.
9. Hon CC, Ramilowski JA, Harshbarger J, et al. An atlas of human long non-coding RNAs with accurate 5′ ends. *Nature*. 2017;543(7644):199–204.
10. Gebert LFR, MacRae IJ. Regulation of microRNA function in animals. *Nat Rev Mol Cell Biol*. 2019;20:21–37.
11. Lam MTY, Li WB, Rosenfeld MG, et al. Enhancer RNAs and regulated transcriptional programs. *Trends Biochem Sci*. 2014;39(4):170–182.
12. Chen YG, Satpathy AT, Chang HY. Gene regulation in the immune system by long noncoding RNAs. *Nat Immunol*. 2017;18(9):962–972.
13. Spurlock CF, Tossberg JT, Guo Y, et al. Expression and functions of long noncoding RNAs during human T helper cell differentiation. *Nat Commun*. 2015;6:6932.
14. Kwon GJ, Henao-Mejia J, Williams A. Editorial: Long non-coding RNAs and immunity. *Front Immunol*. 2019;10:2378.
15. Collier SP, Collins PL, Williams CL, et al. Cutting edge: influence of Tmevpg1, a long intergenic noncoding RNA, on the expression of Ifng by Th1 cells. *J Immunol*. 2012;189(5):2084–2088.
16. Gibbons HR, Shaginurova G, Kim LC, et al. Divergent lncRNA GATA3-AS1 regulates GATA3 transcription in T-helper 2 cells. *Front Immunol*. 2018;9:2512.
17. Aune TM, Spurlock CF. Long non-coding RNAs in innate and adaptive immunity. *Virus Res*. 2016;212:146–160.
18. Salehi S, Taheri MN, Azarpira N, et al. State of the art technologies to explore long non-coding RNAs in cancer. *J Cell Mol Med*. 2017;21(12):3120–3140.
19. Aune TM, Crooke PS, Patrick AE, et al. Expression of long non-coding RNAs in autoimmunity and linkage to enhancer function and autoimmune disease risk genetic variants. *J Autoimmun*. 2017;81:99–109.
20. Castellanos-Rubio A, Ghosh S. Disease-associated SNPs in inflammation-related lncRNAs. *Front Immunol*. 2019;10:420.
21. Pearson MJ, Jones SW. Review: Long noncoding RNAs in the regulation of inflammatory pathways in rheumatoid arthritis and osteoarthritis. *Arthritis Rheumatol*. 2016;68(11):2575–2583.
22. Xue ZX, Cui CJ, Liao ZJ, et al. Identification of LncRNA Linc00513 containing lupus-associated genetic variants as a novel regulator of interferon signaling pathway. *Front Immunol*. 2018;9:2967.
23. Chandan K, Gupta M, Sarwat M. Role of host and pathogen-derived microRNAs in immune regulation during infectious and inflammatory diseases. *Front Immunol*. 2020;10:3081.
24. Ricano-Ponce I, Wijmenga C. Mapping of immune-mediated disease genes. *Annu Rev Genom Hum G*. 2013;14:325–353.
25. Giral H, Landmesser U, Kratzer A. Into the wild: GWAS exploration of non-coding RNAs. *Front Cardiovasc Med*. 2018;5:181.

Immunometabolism

Michihito Kono, Amir Sharabi, and George C. Tsokos

Cell metabolism (Fig. 20.1) is crucial for cell survival, proliferation, and differentiation, and aberrations are involved in the pathophysiology in many diseases.[1] In recent years it has become clear that cell metabolism not only provides energy but also regulates cell phenotype, fate, and function. During the last 5 years, evidence has demonstrated that modulation of immune cell metabolism can be exploited as a therapeutic target for autoimmune diseases. In this chapter, we describe current insights into the immune cell metabolism in autoimmune diseases by focusing on systemic lupus erythematosus (SLE), rheumatoid arthritis (RA), and multiple sclerosis (MS).

GLYCOLYSIS AND PENTOSE PHOSPHATE PATHWAY

Glycolysis is the metabolic pathway that converts glucose into pyruvate. Ten enzymes (including hexokinase, glyceraldehyde 3-phosphate dehydrogenase [GAPDH], and pyruvate kinase) are involved in this pathway. Execution of the pathway generates two pyruvate, two adenosine triphosphate (ATP), and two reduced nicotinamide adenine dinucleotide (NADH) molecules from each glucose molecule. The generated pyruvate has two options: either to form acetylcoenzyme A (acetyl-CoA) by pyruvate dehydrogenase (PDH) and enter the tricarboxylic acid (TCA) cycle in the mitochondria for further energy production, or to be converted into lactate by lactate dehydrogenase (LDH). The pentose phosphate pathway (PPP) is a metabolic pathway parallel to glycolysis. PPP generates nicotinamide adenine dinucleotide phosphate (NADPH), pentose, and ribose 5-phosphate, a precursor for the synthesis of nucleotides.[2,3]

GLUTAMINOLYSIS

Glutaminolysis is crucial in the production of energy in proliferating cells, including lymphocytes and cancer cells. Glutamine is transported in the cells by specialized transporters and is converted to glutamate by glutaminase. Glutaminase has two isoforms: glutaminase 1 (kidney isoform) and glutaminase 2 (liver isoform). Glutaminase 1 has more and faster enzymatic activity than glutaminase 2. Glutamate is sequentially converted to α-ketoglutarate through the action of glutamate dehydrogenase and enters the TCA cycle.[2,3]

FATTY ACID OXIDATION AND FATTY ACID SYNTHESIS

Fatty acid oxidation (FAO) is a mitochondrial process that generates acetyl-CoA from fatty acids. It involves a repeated action of four enzymes that generate NADH and flavin adenine dinucleotide ($FADH_2$), which are used by the electron transport chain (ETC) to produce ATP. In contrast, fatty acid synthesis is a metabolic process whereby end products of glucose catabolism are converted to fatty acids.[2,3]

TRICARBOXYLIC ACID CYCLE AND ELECTRON TRANSPORT CHAIN

The TCA cycle utilizes acetyl-CoA from glycolysis and FAO, and α-ketoglutarate from glutaminolysis. To generate ATP, NADH produced by the TCA cycle is used in the oxidative phosphorylation (OXPHOS) complexes, which are present in the inner membrane of mitochondria. These complexes are known as NADH: ubiquinone oxidoreductase (complex I), succinate dehydrogenase (complex II), cytochrome bc1 complex (complex III), cytochrome c oxidase (complex IV), and ATP synthase (complex V).[2,3] Metformin, which is widely used for treating patients with type 2 diabetes, inhibits the mitochondrial complex I.[4]

KEY CONCEPTS

Main Cell Metabolic Pathways

Glycolysis
- Used by many proliferating cells (e.g., cancer cells and lymphocytes)
- Converts glucose to pyruvate
- Ten enzymes (including hexokinase and pyruvate kinase) are involved
- Pyruvate is converted to acetylcoenzyme A (acetyl-CoA) to enter the TCA cycle or to lactate

Glutaminolysis
- Used by proliferating cells
- Glutamine is converted to glutamate and then α-ketoglutarate and enters the TCA cycle

Fatty Acid Oxidation (FAO)
- Mitochondrial process
- Generates acetyl-CoA from fatty acids

TCA Cycle
- TCA cycle utilizes acetyl-CoA from glycolysis and FAO, and α-ketoglutarate from glutaminolysis
- To generate ATP, NADH produced by the TCA cycle is used in the oxidative phosphorylation complexes

Immune Cell Metabolism

Adaptive Immune Cell Metabolism

The adaptive immune system, also known as the acquired immune system, is highly involved in eliminating infectious pathogens and cancer cells. This immune response generates

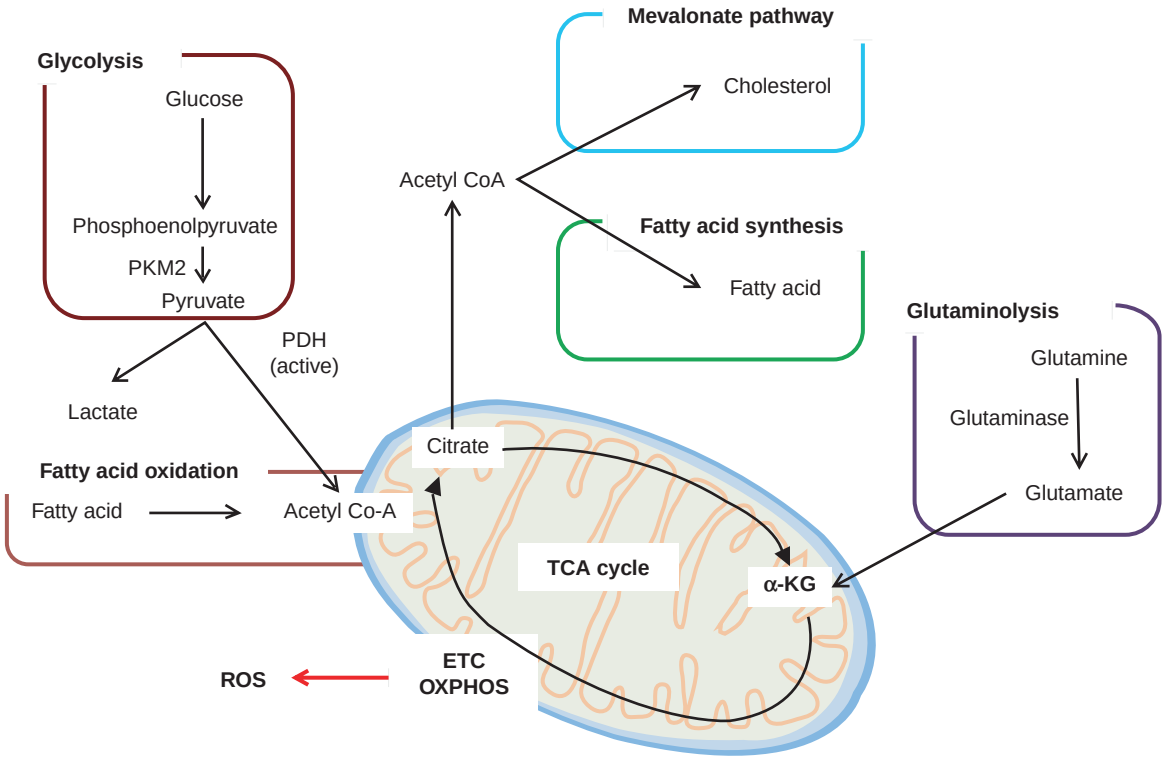


FIG. 20.1 Main Cell Metabolic Pathways. α-*KG*, α-Ketoglutarate; *Acetyl Co-A*, acetyl coenzyme A; *ETC*, electron transport chain; *OXPHOS*, oxidative phosphorylation; *PDH*, pyruvate dehydrogenase; *PKM2*, pyruvate kinase muscle isozyme 2; *ROS*, reactive oxygen species; *TCA*, tricarboxylic acid.

immunological memory which enables a prompt response upon repeat exposure to the same pathogen. Both T cells and B cells are essential components of the adaptive immune system.

T-Cell Metabolism

1. T-cell activation
   Following T-cell receptor (TCR) engagement, naïve T cells undergo a rapid shifting from quiescence to activation and differentiation. Phosphoinositide 3-kinase (PI3K)–AKT–mammalian target of rapamycin (mTOR) pathway is a central signaling pathway in this process that enables a shift from FAO and pyruvate oxidation towards glycolysis and glutaminolysis.[5] Myc is a transcription factor downstream of mammalian target of rapamycin complex 1 (mTORC1), which regulates enzymes involved in glycolysis and glutaminolysis, wherein the levels of Myc, elevated or low, dictate cell differentiation into effector T cells or memory-like T cells, respectively.[5]
2. Effector T-cell metabolism
   Naive CD4 T cells can differentiate into various effector T-cell subsets, and each cell subset utilizes metabolic pathways downstream of the TCR in a differential manner. mTORC1 activity is essential for Th1, IL-17 producing CD4 T (Th17), and CD8 T-cell differentiation, whereas combined mTORC1–mTORC2 activity is required for Th2 and T follicular helper (Tfh) cell differentiation.[1]
   a. Glucose metabolism in effector T cells
      Th1, Th17, and Tfh cells, as well as in CD8 T cells, require glycolysis for cell differentiation, whereas glycolysis deprivation promotes regulatory T (Treg)-cell differentiation (Fig. 20.2).[1] Hypoxia-inducible factor 1α (Hif-1α) and PDH control the differentiation of Th1 and Th17 cells,[2,6] and the Ras homologue gene family member A (RhoA) controls glycolysis during Th2 cell differentiation.[1] The last step of glycolysis results in the generation of pyruvate, which can be converted to lactate in the cytoplasm or to acetyl-CoA in the mitochondria. Th17 cells also express increased levels of pyruvate dehydrogenase kinase 1 (PDHK1) and its inhibition prevents Th17 cell development and promotes Th1 cell differentiation.[6] The transcriptional-factor-inducible cAMP early repressor (ICER), which is present in Th17 cells, suppresses pyruvate dehydrogenase phosphatase catalytic subunit 2 (PDP2) and promotes glutaminolysis.[2,3] Such findings indicate that cell function determination is defined through the control of distinct metabolic processes.
   b. Amino acid metabolism in effector T cells
      The most abundant amino acid in the circulation is the nonessential amino acid glutamine. Following T-cell activation, glutamine consumption is increased.[1,5] Glutamine is hydrolyzed in a process called glutaminolysis into metabolites that support the TCA cycle, as well as the metabolic pathways affecting the synthesis of polyamine, glutathione, and serine.[1,5] Glutaminase is the first enzyme in the glutaminolysis pathway, which is induced by ICER, and glutaminase 1 inhibition impairs mTORC1-dependent Th17 differentiation.[2,3] Glutamine restriction enables Th2 polarization; however, under Th1-polarizing conditions T cells differentiate into Treg cells.[1] Notably, glutamine deprivation promotes Treg cell induction.[1] Leucine is an essential amino acid that stimulates mTOR signaling, and leucine deficiency limits T-cell activation and differentiation into effector T cells.[5] Serine, a nonessential amino acid synthesized from a glycolysis intermediate, is

FIG. 20.2 Main Metabolic Pathway in CD4 T-cell Subsets. *AMPK*, AMP-activated protein kinase; *FAO*, fatty acid oxidation; *HIF-1α*, hypoxia-inducible factor 1α; *mTORC*, mammalian target of rapamycin complex; *OXPHOS*, oxidative phosphorylation; *PPP*, pentose phosphate pathway.

essential for the purine synthesis pathway as well as CD4 and CD8 effector T-cell proliferation.[5]

c. Lipid metabolism in effector T cells
Upon T-cell activation, β-oxidation that catabolizes fatty acids towards the TCA cycle is replaced with glycolysis and glutaminolysis. De novo fatty acid synthesis is important for Th17 cell differentiation. Catalysis of acetyl-CoA is carried out by acetyl-CoA carboxylase 1 (ACC1), and ACC1 deficiency in T cells suppresses Th17 cell differentiation and promotes Treg-cell generation.[5]

3. Memory T-cell metabolism
There are three types of memory CD4 and CD8 T cells: central, effector, and tissue-resident memory T cells. Central memory T cells depend on endogenous lipids and exogenous glycerol for lipid biosynthesis.[1] Upon re-stimulation, a switch to glycolysis takes place to enable the maturation of memory T cells into effector T cells.[5] Tissue-resident memory T cells depend on the presence of exogenous fatty acids for their survival and maintenance within tissues.[1]

4. Treg-cell metabolism
Treg cells utilize mainly fatty acid and pyruvate oxidation (mitochondrial oxidative metabolism) to generate energy. In Treg cells, restraint of mTORC1 signaling is achieved by the high levels of AMP-activated protein kinase (AMPK) and the capacity of forkhead box P3 (FOXP3) to increase ceramide that activates serine/threonine protein phosphatase 2 A (PP2A).[1] Liver kinase B1 (LKB1) is the upstream kinase of AMPK, which is considered to serve as a metabolic sensor in Treg cells that stabilizes FOXP3 expression.[1]

a. Glucose metabolism in Treg cells
During Treg-cell activation and proliferation, glycolysis becomes prominent[4]; however, because mTORC1 prevents

FOXP3 expression, Treg-cell function is interrupted.[1] In contrast, when FOXP3 is increased, the cell preferentially metabolizes lipids over glucose as FOXP3 represses the promoter activity of *MYC*.[1]

b. Lipid metabolism in Treg cells
Peripheral Treg cells use both the mevalonate pathway and end products of glycolysis for the synthesis of fatty acids for their proliferation and function.[1] Tissue-resident Treg cells utilize exogenous fatty acids for their induction and maintenance.[1]

c. Amino acid metabolism in Treg cells
Generally, Treg cells do not require amino acids to exert their suppressive function.[1] However, the induction of peripheral Treg cells is promoted by metabolic intermediates of tryptophan such as kynurenine.[1] Furthermore, Treg cells express amino acid–consuming enzymes, such as indoleamine 2, 3-dioxygenase (IDO) and arginase 1, that reduce the availability of tryptophan and arginine to surrounding T cells, which in turn inhibit mTOR signaling and effector T-cell proliferation and promote Treg-cell induction.[1]

B Cell Metabolism. B cells play a pivotal role in the adaptive immune system by presenting antigens to T cells and by producing pathogen-specific antibodies, as well as cytokines. On the other side, autoreactive B cells are involved in the development of autoimmune diseases. Compared with naïve B cells, activated B cells use more glycolysis and OXPHOS.[7] Activated B cells show an increase in glucose uptake and lactate production, while resting B cells maintain metabolic quiescence through signaling molecules, including TNF-receptor-associated factor 3 (TRAF3) and glycogen synthase kinase 3 (GSK3). Inhibition of glycolysis and mitochondrial respiration or cytosolic acetyl-CoA synthesis reduce activated B-cell proliferation and survival. Several factors (including mTORC1, c-myc, and protein kinase Cβ) are requisite for B-cell activation.[7] After antigen exposure, B cells initiate the formation of germinal centers (GCs) using the help of Tfh cells. During this process, random mutations are introduced into the B-cell antigen receptor (BCR) to increase diversity. GC B cells also consume more glucose than naïve B cells. GC B cells express more genes associated with glycolysis. The expression of HIF-1α, a transcription factor which drives glycolysis, is also enhanced in GC B cells.[7] The serine/threonine phosphatase PP2A is required for GC formation and purine/pyrimidine metabolism.[8] B cells exit GC and differentiate to memory B cells or plasma cells. Since plasma cells produce large amounts of antibodies, they need more glucose and amino acid than naïve B cells. Glucose is also used not only for energy generation but also antibody glycosylation, while OXPHOS is requisite for antibody secretion in plasma cells.[7]

KEY CONCEPTS

B-Cell Metabolism

- Activated B cells use more glycolysis and OXPHOS.
- Germinal center B cells favor glycolysis.
- Plasma cells utilize glucose for not only energy generation but also antibody glycosylation.
- OXPHOS is also requisite for antibody secretion in plasma cells.

Innate Immune Cell Metabolism

The innate immune system (Chapter 3) represents the first response to infections and consists of several primitive components that are recruited and activated before the adaptive immune system. The innate leukocytes include natural killer (NK) cells, mast cells, eosinophils, basophils, macrophages, neutrophils, and dendritic cells (DCs).

Natural Killer Cell Metabolism. NK cells are essential lymphocytes that have antiviral and antitumor activities (Chapter 12). They can produce interferon-gamma (IFN-γ) and directly kill target cells through various cytotoxic mechanisms. Resting NK cells have low levels of glycolysis and OXPHOS.[9] Overnight stimulation of NK cells with cytokines results in substantial increases in both glycolysis and OXPHOS. Activated NK cells metabolize glucose to pyruvate and then lactate through aerobic glycolysis. Inhibition of glycolysis by 2-deoxyglucose (2-DG) or OXPHOS by oligomycin impairs IFN-γ production, while inhibition of glutaminase, the first enzyme in glutaminolysis, or inhibition of FAO by etomoxir does not inhibit NK-cell functions. NK-cell metabolism is impaired in the tumor microenvironment due to the restriction of glucose and amino acids such as arginine.[10] This represents an attractive target to improve NK-cell-based immunotherapy for cancer.

Mast Cell Metabolism. Mast cells are tissue-resident myeloid cells characterized by their effector function in Th2 immunity and are essential in allergic diseases and host defense responses to parasitic diseases (Chapter 44). IgE primes mast cells to release granules and chemical mediators such as histamine and cytokines. IgE-mediated activation of mast cells rapidly increases glycolysis. OXPHOS is also increased after the activation within 2 hours. IL-33 is a cytokine mediator released by endothelial, epithelial, fibroblast cells, and mast cells. IL-33 also activates mast cells and is linked to allergic diseases. IL-33 increases glycolysis and OXPHOS.[11] Inhibition of glycolysis by 2-DG suppresses antigen-induced histamine and inflammatory cytokine release.[11] Suppression of glycolysis with dichloroacetate and inhibition of complex I of the ETC with rotenone results in suppression of cytokine production and degranulation, while inhibiting FAO with etomoxir has no effect.

Eosinophil Metabolism. Eosinophils are important in host defense against parasites and in promoting allergic inflammation (Chapter 45). Eosinophils release their granule proteins, leukotrienes, chemokines, and cytokines. Eosinophils have similar levels of glycolysis and more OXPHOS compared to neutrophils.[12] IL-3, IL-5, or granulocyte–macrophage colony-stimulating factor (GM-CSF) increase glycolysis, lactate production, and OXPHOS.[13] Polyunsaturated fatty acid (PUFA)-derived bioactive metabolites are formed in vivo by enzymatic oxidation through the activation of cyclooxygenases (COX), lipoxygenases (LOX), and cytochrome P450 monooxygenases (CYP). COX-derived prostaglandins and 5-LOX-derived leukotriene play a vital role in the initiation of inflammatory responses. Eosinophil PUVA metabolism represents a target to control allergic diseases.

Basophil Metabolism. Basophils have multiple functions and migrate into organs affected by allergic inflammation and parasitic diseases. Although basophil numbers are low compared with mast cells, the sensitivity to IgE-mediated activation is higher than that of mast cells. IgE-mediated activation of basophils enhances HIF-1α expression, which is essential for glycolysis.[14]

Neutrophil Metabolism. Neutrophils are the most abundant of terminally differentiated leukocytes. During acute inflammation, neutrophils are the first responders to migrate to the sites of inflammation. During differentiation, neutrophils use mainly FAO and OXPHOS while glycolysis is limited.[15] On the other hand,

glycolysis is essential for phagocytosis. PPP potentiates reactive oxygen species (ROS) by providing NADPH, while mitochondrial complex I and III inhibition leads to increased production of mitochondrial ROS. Neutrophil extracellular traps (NETs) are complex extracellular structures composed of DNA from neutrophils and specific proteins from neutrophilic granules. NETs support neutrophil killing of extracellular pathogens but also contribute to the pathogenesis in autoimmune diseases, including ANCA-associated vasculitis and SLE. NETosis increases glucose uptake and glycolysis inhibition by 2-DG, while PPP inhibition by 6-amino-nicotinamide reduces NETs formation.[15]

Macrophage Metabolism. Macrophages are phagocytes that detect, engulf, and destroy pathogens and apoptotic cells. In addition, macrophages are important in tissue homeostasis, repair, pathology, and development. Although several activated forms of macrophages have been reported, there are two main groups: LPS (+IFN-γ)-activated inflammatory macrophages (M1) and IL-4-induced alternatively activated macrophages (M2). Glycolysis is strongly induced in M1 macrophages.[16] The glycolytic enzyme pyruvate kinase M2 (PKM2) is vital to induce IL-1β expression through the activity of HIF-1α. PKM2 also promotes macrophage activation by the inflammasome through the activation of eukaryotic translation initiation factor 2 alpha kinase 2 (EIF2AK2). The glycolytic activator 6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 3 (PFKFB3) promotes the antiviral capacity of macrophages. GAPDH, another enzyme in the glycolytic pathway, regulates TNF secretion in macrophages. Glycolysis is also induced in M2 macrophages and is crucial for IL-4-induced macrophage activation. In M1 macrophages, arginine is used to produce nitric oxide (NO). In contrast, arginine is metabolized by arginase-1 in M2 macrophages. OXPHOS is impaired by NO in M1 macrophages, while in M2 macrophages OXPHOS is induced and supports M2 macrophage development. M1 macrophages also utilize PPP, while M2 macrophages utilize FAO.

Dendritic Cell Metabolism. DCs are professional antigen-presenting cells, which can recognize pathogens and induce naïve T-cell activation and effector differentiation. There are two types of DCs: conventional DCs (cDCs) and plasmacytoid DCs (pDCs). The ETC inhibitor rotenone inhibits the differentiation of human monocytes into DCs by GM-CSF and IL-4.[17] Immature DCs use mainly FAO, and activation of DCs by Toll-like receptor (TLR) agonists increases glycolysis, PPP, and fatty acid synthesis.

Aberrations in Immunometabolism in Immune-Mediated Diseases

a. SLE (Chapter 52).

Patients with SLE exhibit oxidative stress, as reflected by the low levels of glutathione and cysteine in their serum and peripheral blood mononuclear cells.[18] T cells from mice with lupus and patients with SLE are chronically active as a result of TCR signaling hyperactivity and mitochondrial hyperpolarization, which culminate in increased production of ROS.[1] TCR rewiring in SLE favors hyperactivation because the native CD3ζ is replaced with the FcεRIγ chain that increases both the sensitivity and the downstream signaling of the receptor.[1] Increased TCR signaling in SLE can also be due to aberrant lipid raft formation. Lipid rafts are the domains on the plasma membrane where signaling-related molecules accumulate to enable cell activation, and the composition of lipid rafts in resting T cells from patients with SLE resembles that of activated cells.[19] Furthermore, patients with active SLE have a higher synthesis of lipid rafts in CD4 T cells.[19] This can be explained in part by the higher expression of a nuclear receptor: for example, liver X receptor β (LXRβ), which regulates cellular lipid metabolism (including glycosphingolipid synthesis), and the recycling of surface receptors.[1] The transcription factor Friend leukemia integration 1 (FLI1) has been associated with glycosphingolipid synthesis and T-cell activation, and disease severity in lupus-prone mice.[1] FLI1 may also promote lupus nephritis as it increases the expression of chemokines in the kidneys wherein CXCR3-expressing T cells migrate.[1] Passer-by T cells in the kidneys of patients with lupus nephritis are also prone to activation and cell differentiation.[1] Oxidative stress and the increase in PP2A levels in T cells of SLE patients lead to hypomethylation of genes that are involved in SLE pathogenesis.[1] In lupus-prone mice, targeting DNA methylation in CD4 T cells expands Treg cells and in CD8 T cells promotes cytotoxicity.[1] mTOR signaling is increased in T cells from patients with SLE via several pathways. ROS can promote mTOR signaling and CD4 CD8 T-cell differentiation.[1] mTORC1 activity is the highest in CD^{4-} $CD8^{-}$ double-negative T cells from patients with SLE.[18] In SLE, the activity of the T-cell-restricted, serine/threonine protein phosphatase calcium/calmodulin-dependent protein kinase IV (CaMKIV) is upregulated, which promotes mTOR activity.[1] Additionally, the PPP (branches out from glycolysis), which is upregulated in T cells in SLE, as well as amino acid breakdown products, such as kynurenine (a metabolite of tryptophan), can promote mTORC1 activity in these cells.[18] Increase in mTORC1 activity in SLE can also be determined genetically as reported in patients with tuberous sclerosis, who carry mutations of the *TSC1* or *TSC2* genes, resulting in uncontrolled activation of mTORC1 and the development of fulminant

KEY CONCEPTS

Innate Immune Cell Metabolism

Natural Killer (NK) Cells
- Resting NK cells have low levels of glycolysis and OXPHOS.
- Stimulation with cytokines increases glycolysis and OXPHOS.
- Glycolysis and OXPHOS are requisite for NK cell functions.

Mast Cells
- IgE-mediated activation increase glycolysis and OXPHOS.
- IL-33 also increases glycolysis and OXPHOS.

Eosinophils
- Use same levels of glycolysis and more OXPHOS than neutrophils.
- IL-3, IL-5, and GM-CSF increase glycolysis and OXPHOS.

Neutrophils
- Use fatty acid oxidation (FAO) and OXPHOS for differentiation.
- Use glycolysis for phagocytosis.
- Use glycolysis and PPP for neutrophil extracellular trap formation.

Macrophages
- M1 macrophages favor glycolysis, PPP, and use of arginine.
- M2 macrophages also use glycolysis and OXPHOS.

Dendritic Cells (DC)
- Immature DC mainly use FAO.
- Toll-like receptor (TLR) agonists increase glycolysis PPP, and fatty acid synthesis.

manifestations of SLE.[1] The chronic activation of T cells in murine and human SLE is associated with increased rates of glycolysis.[1] Studies in lupus-prone mice have shown that increased glycolysis predisposes to SLE. For example, overexpression of Glut1 in mice leads to lupus-like manifestations, including the production of autoantibodies and deposition of immune complexes in glomeruli.[1] Indeed, patients with active SLE display high levels of GLUT1 in their T cells.[1] The rate of glutaminolysis in SLE is undetermined. Yet, decreasing the rate of glutaminolysis in lupus-prone MRL/*lpr* mice has led to reduced Th17 cell differentiation and lowered disease activity in a mechanism that included HIF-1α and glycolysis downregulation.[2] Th17 cells in comparison with other T-cell subsets in patients with SLE display increased glycolysis and decreased glucose-mediated OXPHOS.[1] These lupus-derived Th17 cells have increased ICER activity, which in turn decreases the levels of PDP2 and diverts pyruvate to lactate.[2,3] In agreement, the ICER/CREMα-suppressive splice variants of CREM are also highly expressed in CD4 T cells from patients with SLE and function as transcriptional enhancers of glutaminase, which promotes glutaminolysis and Th17 cell differentiation.[2,3]

b. RA (Chapter 53).
Many studies have shown the importance of T cells in the pathogenesis of RA.[20] The strong association of RA with the human leukocyte antigen (HLA)-DRB1 locus points to the significance of T-cell selection and antigen presentation in the pathogenesis of RA.[20] Other genetic risk alleles, including variants of the genes *CD28*, *PTPN22*, *IL2RA*, *CD40*, *CCL21*, and *CCR6*, also suggest a role for T cells in the development of RA.[20] Although RA pathogenesis was previously considered to be Th1-dependent, recent evidence indicates the importance of Th17 cells. Th17/Treg balance is also involved in RA pathogenesis.[21] Th1 and Th17 cells use mainly glycolysis, which is initiated within minutes following TCR activation.[22] Glut1, a glucose transporter, is highly expressed in the surface of Th1 and Th17 cells, while Treg cells express low levels of this transporter.[6] A glucose transporter inhibitor CG-5 reduces Th1 and Th17 cell differentiation and promotes Treg cell induction.[4] Inhibition of hexokinase, the enzyme catalyzing the first step in glycolysis, by 2-DG reduces Th1[4] and Th17 cell differentiation.[22] Pyruvate, the end product of glycolysis, is converted to lactate by LDH or converted to acetyl-CoA by PDH to fuel the TCA cycle. In the absence of oxygen and even under normoxic conditions in Th17 cells, pyruvate tends to be converted to lactate by LDH.[22] Lactate itself has an important role in autoimmune diseases and is one of the most enriched products of cellular metabolism in inflamed tissues, including synovial tissue.[23] Lactate induces expression of its transporter, SLCA12, on human CD4 T cells in the inflamed tissue and promotes IL-17 production via PKM2/signal transducer and activator of transcription 3 (STAT3) signaling. Furthermore, SLCA12 blockade ameliorates the disease severity in mice induced to develop arthritis.[23] Sirolimus, an mTOR inhibitor, reduces disease activity in RA.[24]

PPP enables cells to generate products that are crucial for T-cell function and expansion, as they fuel biomass generation.[21] Naïve T cells from patients with RA have a defect in glycolytic flux due to the upregulation of glucose-6-phosphate dehydrogenase (G6PD) and downregulation of 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase 3 (PFKFB3) after stimulation.[21] Excess G6PD shunts glucose into the PPP, resulting in NADPH accumulation and ROS consumption (see Fig. 20.1). T cells from patients with RA do not sufficiently activate the redox-sensitive kinase ataxia telangiectasia mutated (ATM), which in turn shifts naïve CD4 T-cell differentiation towards Th1 and Th17 cell lineages.[21]

c. MS (Chapter 66).
MS is a demyelinating disease characterized by a chronic inflammatory process in the brain and spine. Loss of myelinating oligodendrocytes is considered the beginning of MS pathogenesis, and both Th1 and Th17 cells are involved. Inhibition of glycolysis by 2-DG ameliorates murine MS (e.g., experimental autoimmune encephalomyelitis [EAE]).[22] PKM2 inhibitor, which inhibits the last enzyme of the glycolysis pathway, also ameliorates EAE.[2] Lack of HIF-1α, which mediates glycolysis, diminished Th17 cell development and enhanced Treg cell differentiation in mice, thus protecting them from autoimmune neuroinflammation.[22] Inhibition of glutaminase 1, glutamate oxaloacetate transaminase (GOT) 1, or amino acid transporter ASCT2 reduces Th17 cell differentiation and improves disease activity in EAE.[2] The inhibition of ACC1, the rate-limiting enzyme for fatty acid synthesis, attenuates EAE[5]; C75, which is another inhibitor of this synthesis, also ameliorates EAE.[25]

Immunometabolism-Based Targeted Treatment

Disease-modifying anti-rheumatic drugs (DMARDs) can affect several metabolic pathways. Glucocorticosteroids promote gluconeogenesis and inhibit glycolysis. Methotrexate inhibits folate metabolism, thereby inhibiting the one-carbon metabolism pathway, and it can also increase the adenosine pool and AMP production, which in turn activates AMPK, which inhibits mTOR signaling.[1] Mycophenolate mofetil (MMF) interferes with DNA synthesis and limits T-cell expansion, inhibiting the expression of *MYC* and *HIF1A* and the signaling via the PI3K–AKT–mTOR pathway; MMF also reduces glycolysis and oxidative stress in CD4 T cells in vitro.[1] In a phase I–II clinical trial in patients with SLE, sirolimus, an mTORC1 inhibitor, reduced disease activity and restored the balance of T-cell lineages.[18] Inhibition of mTORC1 signaling suppressed T-cell hyperactivation, the development of Th17, and double-negative T cells, and promoted Treg cell differentiation and function in patients with SLE.[18] Metformin, an antidiabetic medication, when coadministered with 2-DG (an inhibitor of glucose metabolism), normalized T-cell metabolism, decreased the numbers of Tfh cells, and reversed kidney disease in lupus-prone mice.[4] In a proof-of-concept trial studying patients with mild-to-moderate SLE, the addition of metformin to standard-of-care treatment decreased the frequency of clinical flares and permitted prednisone dose reduction.[1]

Pioglitazone, an insulin-sensitizing drug that can rapidly activate AMPK, also inhibits mTORC1 and promotes the expansion of Treg cells from patients with SLE in vitro.[1] Bz-423 is a benzodiazepine compound that can inhibit an ATP synthase, which is increased in autoreactive lymphocytes from mice with lupus; and treatment of the mice with Bz-423 led to autoreactive lymphocyte apoptosis and clinical improvement.[1] Cysteine is depleted in T cells from patients with SLE. *N*-acetyl cysteine is a cysteine analogue that, in patients with SLE, led to inhibition of mTORC1 activity in double-negative T cells and decreased disease activity.[18] NB-DNJ (*N*-butyldeoxynojirimycin) inhibits the synthesis of glycosphingolipids, and it is used to treat type

TABLE 20.1 Drug Treatments Affecting Immunometabolic Targets

Drug	Therapeutic Target	Effects on Disease Model
2-Deoxy-D glucose and metformin	Hexokinase and mitochondrial complex I	Reduces disease activity, and improves kidney disease in lupus model
BPTES, CB-839, and 968	Glutaminase 1	Reduces disease activity in EAE and arthritis model, and improves kidney disease in lupus model
Bz-423	Mitochondrial metabolism	Reduces disease activity in lupus model
NB-DNJ	Glucosylceramide synthetase	Reduces disease activity in lupus model
N-acetyl cysteine	Cysteine metabolism	Reduces disease activity in arthritis model, and improves kidney disease in lupus model
Sirolimus	mTOR signaling	Reduces disease activity in EAE and lupus model, and rheumatoid arthritis
Pioglitazone (agonist)	PPARγ	Improves disease activity in EAE and arthritis model, and nephritis in lupus model

BPTES, Bis-2-(5-phenylacetamido-1, 3, 4-thiadiazol-2-yl)ethyl sulfide; *mTOR,* mammalian target of rapamycin; *NB-DNJ, N*-butyldeoxynojirimycin; *PPARγ,* peroxisome proliferator-activated receptor γ.

I Gaucher disease. In patients with SLE it had beneficial effects on T cells in vitro, and in lupus-prone mice it restored lipid metabolism in T cells and B- and T-cell associated antigen (BTLA) function (Table 20.1).[1]

CONCLUSIONS AND FUTURE DIRECTIONS

Immune cell metabolism changes dramatically during the immune response. It is now evident that cell survival, fate, and differentiation can be determined by the predominance and dynamics of metabolic pathways within the responding cell. Therefore, manipulation of metabolic pathways can be utilized to modulate and rebalance specific subsets of cells within the immune system as a novel approach in the treatment of autoimmune diseases. Future studies will decipher further the pathogenic role of immunometabolism in the development of autoimmune diseases and ways to readjust these metabolic pathways to advance treatment of autoimmune diseases.

REFERENCES

1. Sharabi A, Tsokos GC. T cell metabolism: new insights in systemic lupus erythematosus pathogenesis and therapy. *Nat Rev Rheumatol.* 2020;16(2):100–112.
2. Kono M, Yoshida N, Tsokos GC. Metabolic control of T cells in autoimmunity. *Curr Opin Rheumatol.* 2019;32(2):192–199.
3. Vukelic M, Kono M, Tsokos GC. T cell metabolism in lupus. *Immunometabolism.* 2020;2(2):e200009.
4. Teng X, Cornaby C, Li W, Morel L. Metabolic regulation of pathogenic autoimmunity: therapeutic targeting. *Curr Opin Immunol.* 2019;61:10–16.
5. Bantug GR, Galluzzi L, Kroemer G, Hess C. The spectrum of T cell metabolism in health and disease. *Nat Rev Immunol.* 2018;18(1):19–34.
6. Makowski L, Chaib M, Rathmell JC. Immunometabolism: from basic mechanisms to translation. *Immunol Rev.* 2020;295(1):5–14.
7. Jellusova J. The role of metabolic checkpoint regulators in B cell survival and transformation. *Immunol Rev.* 2020;295(1):39–53.
8. Meidan E, Li H, Pan W, et al. Serine/threonine phosphatase PP2A is essential for optimal B cell function. *JCI Insight.* 2020;5(5):e130655.
9. O'Brien KL, Finlay DK. Immunometabolism and natural killer cell responses. *Nat Rev Immunol.* 2019;19(5):282–290.
10. Terren I, Orrantia A, Vitalle J, et al. NK cell metabolism and tumor microenvironment. *Front Immunol.* 2019;10:2278.
11. Caslin HL, Taruselli MT, Haque T, et al. Inhibiting glycolysis and ATP production attenuates IL-33-mediated mast cell function and peritonitis. *Front Immunol.* 2018;9:3026.
12. Porter L, Toepfner N, Bashant KR, Guck J, Ashcroft M, Farahi N, et al. Metabolic profiling of human eosinophils. *Front Immunol.* 2018;9:1404.
13. Jones N, Vincent EE, Felix LC, et al. Interleukin-5 drives glycolysis and reactive oxygen species-dependent citric acid cycling by eosinophils. *Allergy.* 2020;75(6):1361–1370.
14. Sumbayev VV, Nicholas SA, Streatfield CL, Gibbs BF. Involvement of hypoxia-inducible factor-1 HiF(1alpha) in IgE-mediated primary human basophil responses. *Eur J Immunol.* 2009;39(12):3511–3519.
15. Kumar S, Dikshit M. Metabolic insight of neutrophils in health and disease. *Front Immunol.* 2019;10:2099.
16. Van den Bossche J, O'Neill LA, Menon D. Macrophage immunometabolism: where are we (going)? *Trends Immunol.* 2017;38(6):395–406.
17. Pearce EJ, Everts B. Dendritic cell metabolism. *Nat Rev Immunol.* 2015;15(1):18–29.
18. Huang N, Perl A. Metabolism as a target for modulation in autoimmune diseases. *Trends Immunol.* 2018;39(7):562–576.
19. Robinson GA, Waddington KE, Pineda-Torra I, Jury EC. Transcriptional regulation of T-cell lipid metabolism: implications for plasma membrane lipid rafts and T-cell function. *Front Immunol.* 1636;2017:8.
20. Okada Y, Wu D, Trynka G, et al. Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature.* 2014;506(7488):376–381.
21. Weyand CM, Goronzy JJ. Immunometabolism in early and late stages of rheumatoid arthritis. *Nat Rev Rheumatol.* 2017;13(5):291–301.
22. Chapman NM, Boothby MR, Chi H. Metabolic coordination of T cell quiescence and activation. *Nat Rev Immunol.* 2019
23. Pucino V, Certo M, Bulusu V, et al. Lactate buildup at the site of chronic inflammation promotes disease by inducing CD4(+) T cell metabolic rewiring. *Cell Metab.* 2019;30(6):1055–1074. e8.
24. Wen HY, Wang J, Zhang SX, et al. Low-dose sirolimus immunoregulation therapy in patients with active rheumatoid arthritis: a 24-week follow-up of the randomized, open-label, parallel-controlled trial. *J Immunol Res.* 2019;2019:7684352.
25. Young KE, Flaherty S, Woodman KM, et al. Fatty acid synthase regulates the pathogenicity of Th17 cells. *J Leukoc Biol.* 2017;102(5):1229–1235.

21

Immune Deficiencies at the Extremes of Age

Claire E. Gustafson, Cornelia M. Weyand, and Jörg J. Goronzy

Age is a major factor determining the quality and quantity of an immune response, with alterations most obvious at the extremes of age. Both infants and older adults exhibit impaired immune responses to infection and vaccination. However, the underlying mechanisms for immune dysfunction are distinct. Infants are born with limited antigen exposure and a tolerogenic-inclined immune system, which progressively develops throughout infancy and childhood. In older age, the mechanisms underlying immune decline are broadly termed *immunosenescence*. The term is reminiscent of *cellular senescence*, which is defined as an irreversible cell cycle arrest. However, it is misleading to interpret immune aging only as the accumulation of senescent cells, and, similar to aging in general, numerous pathways are involved.[1] The age-related decline in immune competence is not linear. As early as age 40 years, immune responses to selected vaccines (e.g., hepatitis B) start to decline. The incidence rate of herpes zoster, a reactivation of latent varicella-zoster virus (VZV), starts to increase at age 50 years, as does morbidity and mortality from influenza infection. A more abrupt immunological decline appears to occur in the eighth decade of life.

In terms of public health, infections are major causes of morbidity in the very young and the very old. Although child mortality rates have dropped by almost 50% between 1990 and 2013, pathogenic infections are still one of the largest causes of infant mortality worldwide, accounting for more than 1.6 million deaths per year. Vaccinations have helped change the infectious landscape in children and young adults, but certain vaccines, such as those against *Streptococcus pneumoniae*, still have limited protective capacity. Susceptibility to pathogenic infection and ineffectiveness of vaccination is even greater in older adults than in young infants. Moreover, during aging, a functional immune system is important for tissue repair, such as in degenerative diseases, and it is vital for cancer surveillance.

INFANCY AND THE GENERATION OF AN IMMUNE SYSTEM

Infants start to develop their immune system before birth, maintaining a fine balance between immunological tolerance that helps prevent pro-inflammatory responses in utero while preserving the ability to respond to foreign antigen exposure upon birth. Regulatory pathways utilized by the fetus for protection against possible infections and maternal–fetal rejection in utero are still reflected in the newborn infant immune system, characterized by immune-suppressive and anti-inflammatory cellular responses to foreign antigens. Although newborns and young infants lack the ability to mount effective immune responses against many pathogens, they acquire initial immune protection through the passive transfer of maternal immunoglobulin G (IgG) in utero via the placenta (termed "passive immunity") and from secretory IgA and antimicrobial factors present in maternal breast milk.[2] Alterations in passive immunity (e.g., preterm birth) can lead to increased susceptibility to infections and breakdown of immune tolerance in the infant. Passive immunity can be enhanced by vaccination of pregnant mothers to promote the transfer of vaccine-specific IgG. However, within the first months of life, as maternal IgG wanes and breastfeeding is discontinued, infants must actively develop their own immune responses for protection.

INNATE IMMUNE DEVELOPMENT

The innate immune system is typically considered the first line of defense against infection. The development of the innate immune system begins in utero, with all the classic innate cell types present by the end of the first trimester; monocytes and dendritic cells (DCs) are the earliest cells observed at 4 weeks' gestation (WG) followed by granulocytes and natural killer (NK) cells at 8 WG.[3] These innate immune cells expand in number and mature in utero; however, at birth the functionality of innate immune cells is still diminished compared with later in life.

During early infancy, all cell types of the innate immune system demonstrate some impairment of function, ranging from reduced mobility to skewed cytokine production in response to stimulation.[4] The most significant functional limitation of neonatal innate immune cells is their collective inability to kill pathogens. Limited chemotaxis in neutrophils is accompanied by a reduced ability to effectively kill pathogens as a consequence of poor phagocytosis and decreased secretion of neutrophil extracellular traps. Moreover, NK cells and plasmacytoid DCs have reduced capacity to prevent infection through reduced cytotoxic function and lower secretion of interferon-α (IFN-α), respectively. In addition, antigen-presenting cells (APCs) are unable to provide effective help to T cells because of reduced expression of costimulatory molecules and production

KEY CONCEPTS

Characteristics of the Developing Innate Immune System

- Reduced antibacterial responses (i.e., phagocytosis, secretion of neutrophil extracellular traps [NETs]) by neutrophils and monocytes
- Impaired neutrophil chemokinesis ("directed movement")
- Diminished capacity of antigen-presenting cells (APCs; e.g., dendritic cells [DCs]) to provide proper costimulatory help to T cells
- Reduced antiviral responses (i.e., interferon [IFN]-α secretion) by plasmacytoid dendritic cells (pDCs)

of more anti-inflammatory cytokines. These combined limitations during early infancy lead to increased susceptibility to viral and bacterial infections and also contribute to the reduced functionality of adaptive immune cells.

ADAPTIVE IMMUNE DEVELOPMENT

Adaptive immune cells (T and B cells) begin to develop in utero around the start of the second trimester. It is important to note that although lymphocytes initially develop within the fetal liver (~7 WG) and then in the bone marrow (~12 WG), T-cell maturation occurs in the thymus, whereas B cells continue maturation in bone marrow. Over the course of gestation, both naïve T- and B-cell compartments expand, reaching absolute cell concentrations at birth greater than that of adults. Thus, a defect in cellular generation cannot account for the immune limitations observed in infants. However, the composition of lymphocyte populations within young infants is distinct from that of adults (Fig. 21.1). At birth, the infant T- and B-cell compartments consist mainly of naïve and transitional cells recently migrated out of the thymus or bone marrow (>90% of the population). Infant T- and B-cell repertoires have less diversity within their receptors' antigen-binding regions compared with those of adults. T-cell receptors (TCRs) display reduced V-J complexity and fewer amino acid additions. B cells have decreased B-cell receptor diversity because of less affinity maturation (i.e., somatic hypermutation; Chapter 7) compared with adult populations.

The increased levels of recent thymic emigrants and naïve T cells are maintained both in the periphery and in tissues during early infancy. Effector memory T cells can be found selectively in the lungs and gastrointestinal tract of infants, likely because these mucosal tissues are the earliest sites of antigen exposure. Class-switch antibodies produced by B cells are required for mucosal protection (primarily IgA) and systemic protection (primarily IgG) through opsonization, neutralization, or antibody-dependent cellular cytotoxicity. However, infants demonstrate significantly lower levels of IgG and IgA compared with adults throughout the first year of life, in both the periphery and mucosal tissues.

INFANCY AND FUNCTIONAL DIFFERENTIATION OF ADAPTIVE IMMUNE CELLS

The main feature of a "mature" adaptive immune system is the development of memory T and B cells that recognize specific foreign (or in some context, self) antigens. Not only are infants born with limited antigen exposure and thus little memory, infant responses to antigenic stimulation in the T- and B-cell compartments display altered functionality compared to adults. Some of the most striking alterations are seen within the CD4 T-cell compartment, characterized by significantly increased frequency of immunosuppressive T regulatory cells (Tregs; 30 to 40% in infants vs. 1 to 10% in adults).[5] These changes are accompanied by increased T-helper 2 (Th2) frequencies. This bias is due to preferential dif-

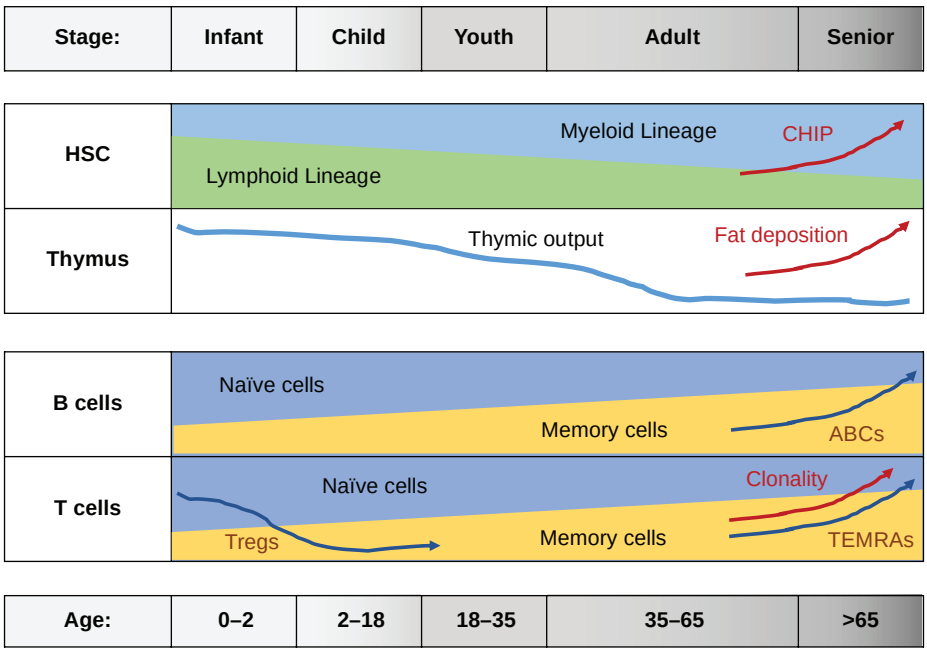


FIG. 21.1 Distinct Changes in Human Immune Cells Across Age. From infancy to old age, the immune system undergoes numerous changes. Hematopoietic stem cells *(HSCs)* progressively become biased towards development into the myeloid lineage, in conjunction with DNA mutations, causing increased clonal hematopoiesis of indeterminate potential *(CHIP)* in old age. Thymic output decreases with time, diminished to almost negligible levels during adulthood. Adipose tissue also accumulates in the thymus with age. In the adaptive immune system, B cells remain relatively constant in absolute numbers, whereas T cells decline. However, B and T cells exhibit alternations in subset composition, highlighted by a decrease in naïve cell frequencies, with a relative gradual increase in memory cells and significant expansion of pro-inflammatory memory subsets (i.e., age-associated B cells *[ABCs]* and terminally differentiated effector memory T cells re-expressing CD45RA *[TEMRA]* that combine features of the innate and the adaptive immune system) with old age. This expansion is in tandem with clonal expansion in the T-cell compartment. Conversely, infants have high levels of the anti-inflammatory regulatory T-cell subset that dramatically decreases with age.

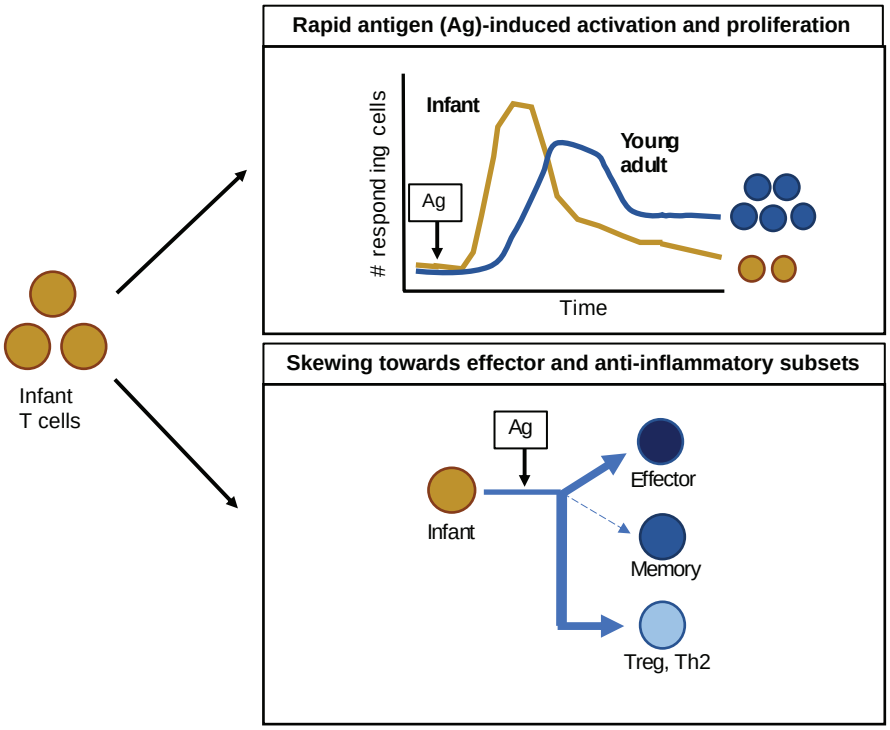


FIG. 21.2 T-Cell Functionality During Infancy. Infant T cells display unique properties compared with those in adults, highlighted by more rapid antigen *(Ag)*-driven effector responses. Additionally, infant T cells show a propensity to differentiate into effector cells, regulatory T cells, or Th2 subsets, instead of memory cells.

ferentiation of neonatal CD4 T cells into Treg or Th2 cells, instead of Th1, Th17, or T follicular helper (Tfh) cells. Altered differentiation of CD4 Th subsets promotes a tolerance-inducing environment good for controlling unnecessary responses to foreign antigen exposure after birth. However, it also leads to inhibition of effector T cells and limits B-cell antibody responses against infectious pathogens and to vaccination (Fig. 21.2).

Until very recently, neonatal T cells were considered "immature" versions of adult cells; however, new studies have elegantly demonstrated that neonatal T cells are actually a functionally distinct population of cells specifically tailored to respond to the needs of the neonatal environment.[6] Neonatal T cells, particularly $CD8^+$ cells, are epigenetically poised to develop into effector over memory T cells and display enhanced ability to respond to danger signals (e.g., pathogen-associated molecular patterns [PAMPs]) and inflammation (e.g., complement) (see Fig. 21.2). microRNA regulation of transcriptional networks contributes to a preferential effector-like state of neonatal T cells, which includes lower TCR activation thresholds caused indirectly via increased miR-181a. The innate-like features of neonatal T cells allow for a more rapid effector response to environmental cues and increase the protective capacity of the infant, albeit at the expense of the development of memory cells. In addition, the more broadly reactive (promiscuous) TCR repertoires, from utilization of more germ-line encoded sequences, may permit a wider responsiveness to multiple types of antigens (i.e., self, commensals, and pathogens).

Similar to T-cell responses, B cells also display less effective immune responses during infancy, with preferential development of short-lived effector cells. Neonatal B cells are less able to undergo class-switching from IgM to IgA or IgG upon antigen stimulation and differentiated antibody-secreting plasma cells within the infant bone marrow have poor survival.[7] Moreover, infant B cells have lower rates of somatic hypermutation compared with those of adults, which is important for the generation of high-affinity antibodies. Although these defects can be partially attributed to extrinsic T-cell defects, such as poor Tfh cell generation, intrinsic defects in infant B cells (e.g., lack of costimulatory molecules) also prevent effective antibody responses. Collectively, these neonatal B-cell defects lead to less effective and shorter-lived antibody responses against infectious pathogens and vaccines during infancy.

KEY CONCEPTS

Characteristics of the Developing Adaptive Immune System

- High frequencies of antigen-naïve T and B cells
- Skewing of naïve CD4 T-cell differentiation toward immunosuppressive regulatory T cells (Tregs) and T-helper 2 (Th2) cells, but not Th1 or Tfh cells
- Skewing of naïve CD8 T cells toward effector cell, but not memory cell, development
- Limited ability of B cells to produce high-affinity, class-switched antibodies in response to both T-cell–dependent and T-cell–independent antigens

INFANT IMMUNE DEVELOPMENT AND THE MICROBIOME

One of the first environmental exposures that infants experience is with colonizing bacteria. Within hours after birth, the

infant intestinal tract is colonized by nonpathogenic bacteria (termed "microbiome"; Chapter 22). The microbiome can reach concentrations of 10^{10} bacteria per gram of stool within 7 days but mirrors an adult-like composition only after 2 years of age. Early animal studies demonstrated that microbial colonization is essential for the normal development of the immune system during early infancy and throughout life.[8] In particular, IgA production requires the presence of a microbiome for its development and long-term maintenance. Moreover, the specific composition of the microbiome influences the generation of innate lymphoid cells and CD4 T-cell subsets (e.g., Treg, Th17, Tfh). Perturbations in the infant microbiome caused by such factors as mode of delivery and antibiotic treatment may significantly alter developing immune responses. For example, babies delivered via cesarean section (C-section) have a microbiome composition distinct from that of babies delivered vaginally, with C-section delivered babies having an intestinal microbiome more dominated by skin bacteria. These differences are correlated with an increased incidence of allergy and asthma during childhood in infants delivered by C-section. Recent work has now demonstrated microbiome-derived metabolites influence Treg development in early infancy, providing a causal link between the microbiome and infant immune tolerance.

CLINICAL CONSEQUENCES FOR CHILDHOOD VACCINATION

Infants are highly susceptible to pathogenic infection during the first year of life. Indeed, infections are one of the major causes of mortality, accounting for more than 24% of global infant mortality. More than 1 million deaths per year are caused by respiratory tract infections (e.g., *S. pneumoniae*, respiratory syncytial virus [RSV]) and almost the same number by intestinal infections (e.g., rotavirus, *Escherichia coli*). The development and distribution of vaccines targeting these pathogens has significantly reduced infant infection rates and consequent mortality; however, vaccine efficacy can vary depending on the infant's gestational age at birth, mode of vaccine delivery (oral vs. intramuscular), interference by maternal antibodies, and tolerogenic immune responses, as described above. Additionally, infants require multiple vaccine boosters to elicit and maintain robust protective immunological memory against infectious pathogens (*Chapter 87*). Better understanding of the mechanisms of these limitations of the infant immune responses, particularly at mucosal sites where these pathogens initially infect, would facilitate the generation of better vaccines designed to induce more effective and long-term immune protection during infancy.

CLINICAL RELEVANCE

Consequences of "Immaturity" of the Neonatal Immune System

- Increased morbidity and mortality from bacterial infections (e.g., pneumococci)
- Increased morbidity and mortality from viral infection (e.g., influenza virus, respiratory syncytial virus, rotavirus)
- Ineffective primary vaccinations; multiple boosters required for most vaccines to elicit effective protection
- Susceptibility to allergy and asthma when microbiome colonization is disrupted during early infancy

OLDER AGE AND IMMUNE CELL GENERATION

The immune system is in constant demand for cellular replenishment to compensate for peripheral losses and cell death. For neutrophils, which have a short half-life, the body needs to produce ~10^{10} cells/kg/day; for lymphocytes, which are more long-lived but are also more numerous because of their wider tissue distribution, the daily need is in the order of several billion cells. Studies in the 1960s showed that the hematopoietic tissue in bone marrow decreases with age. A similar numerical decline is seen for peripheral hematopoietic stem cells (HSCs). Telomerase expression in peripheral HSCs is not fully protective, and telomeric length shortens with age, similar to differentiated mononuclear cells. The functionality of HSCs also changes with age, exhibiting reduced regenerative capacity and a differentiation bias towards myeloid versus lymphoid precursors (see Fig. 21.1).[9] This bias is driven by DNA damage and epigenetic changes and may contribute to the clinical observation that HSC-derived leukemia preferentially has a lymphoid phenotype in the young and a myeloid one in older adults. Additionally, DNA damage causing mutation in genes mediating epigenetic regulation (e.g., TET2, *DNMT3A*) can give certain HCSs fitness advantages, which in turn propagates to all of their descents, in particular the myeloid lineage. This expansion, termed clonal hematopoiesis of indeterminate potential (CHIP), is highly age-dependent and is closely associated with the development of hematological malignancies and cardiovascular disease.[10] CHIP may also contribute to the low-grade inflammation that is characteristic of the aging host.

Collectively, lymphocyte replenishment is more affected by age than that of myeloid lineages. Peripheral neutrophil numbers do not decline and are able to recover after therapy-induced depletion (e.g., chemotherapy). Furthermore, there is no loss in the ability to generate robust neutrophilia in response to infection or other stressors.

KEY CONCEPTS

Aging Influences Immune Cell Generation and Population Homeostasis

- Hematopoietic stem cells (HSCs) are reduced in frequency and biased toward myeloid lineages over lymphoid lineages.
- HSC mutations cause HSC clonality predisposing to hematological malignancies and cardiovascular disease.
- Myeloid cell generation is largely intact.
- B-cell generation is relatively maintained but B-cell repertoire diversity is disturbed.
- As a result of thymic involution, naïve T cells are generated from homeostatic proliferation of peripheral T cells.
- Ability to rebuild a T-cell repertoire after lymphocyte-depleting interventions is compromised after mid-adulthood.
- Naïve T cells acquire memory-like features, and preferentially differentiate into effector, but not memory cell populations. $CD4^+$ cells favor Th1, but not Tfh development.
- Virus-specific terminally differentiated effector T-cell populations accumulate.

Although HSC-intrinsic alterations skew lineage commitments away from lymphoid cell generation, the peripheral B-cell population is relatively maintained across aging, as is the development of effector and memory B cells.[11] However, an expansion of nonclassical memory B cells termed "age-associated

B cells" is observed. These cells are pro-inflammatory, more autoreactive, and secrete antibody in response to innate (i.e., Toll-like receptor [TLR]) but not antigen-specific stimulation.[12] In addition, aged B cells display a reduced ability to undergo class switch recombination and somatic hypermutation, which likely accounts for lower antibody affinities and reduced antibody functionality observed in older individuals. Age-associated differences in B-cell selection and expansion may also contribute to monoclonal gammopathies (in >5% of individuals older than 70 years; Chapter 79) and the presence of autoantibodies in the absence of disease with increasing age.

T-cell generation is more affected by age than any other myeloid or lymphoid lineage because of the involution of the thymus. The thymus undergoes dramatic structural changes that begin during childhood.[13] Thymopoietin niches disappear, and the numbers of thymic epithelial cells and thymocytes decline in what appears to be a developmentally regulated process. Thymic resection in children undergoing cardiac surgery reproduces many of the T-cell repertoire changes in 20-year-old individuals that are otherwise seen in 70 to 80 year olds, documenting the importance of thymic production during the human growth period. In contrast, throughout adulthood, homeostatic proliferation of T cells accounts for the bulk of T-cell generation.[14] This process is more robust for naïve CD4 compared to naïve CD8 T cells, the latter of which are clearly lost with age. Overall repertoire diversity (i.e., the number of different TCRs) contracts with age but still remains highly diverse for both naïve CD4 and CD8 T-cell subsets, suggesting that thymic activity is not necessary to maintain repertoire diversity once a repertoire has formed.

T-CELL POPULATION HOMEOSTASIS

The adaptive system responds to antigenic challenges with clonal expansion and differentiation of naïve cells into effector cells, followed by clonal downsizing and persistence of long-lived memory T cells (i.e., immunological memory). Infections therefore leave a permanent imprint on the immune system, a mechanism on which vaccinations capitalize. Similar to neonatal responses, naïve CD8 T cells in older individuals preferentially differentiate into effector cells and fail to establish memory cells (Fig. 21.3). Naïve CD4 T cells display a similar bias, preferentially developing into short-lived effector T cells at the expense of Tfh and memory cell development. These changes are in tandem with the loss of vaccine-specific memory T cells over time, and likely account, in part, for reduced vaccine efficacy as well as increased susceptibility to infections in older individuals.[15]

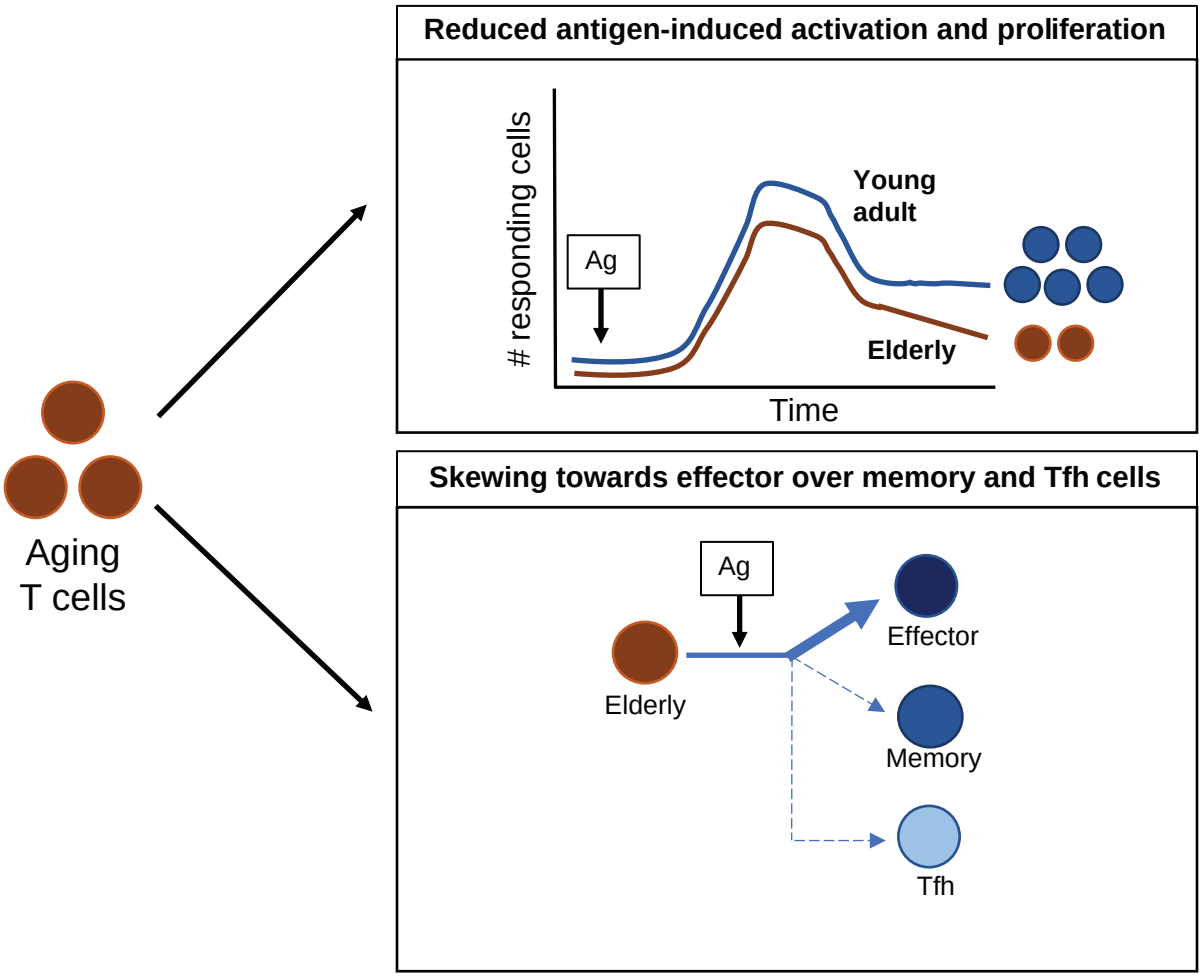


FIG. 21.3 T-Cell Functionality During Advanced Age. In older individuals, T cells show a diminished ability to respond to antigen. This reduction has been linked with preferential development into effector cells over memory or Tfh subsets, which are two subsets essential for long-term immune protection.

Pathogen-induced clonal expansion represents a challenge to homeostatic mechanisms that are supposed to maintain a balance among naïve, memory, and effector cells.[16] This is particularly evident in persistent infections where the offending pathogen cannot be cleared (e.g., herpes viruses). Herpes viruses establish latency following infection, making them highly prevalent in apparently healthy populations without causing active disease. Classic examples VZV, Epstein-Barr virus (EBV), and cytomegalovirus (CMV). The effects of these herpes viruses with immune aging differ greatly. VZV tends to reactivate with age, presenting as shingles. A decrease in the frequency of VZV-specific CD4 memory T cells has been postulated to explain this lack in viral control mechanisms. In contrast, EBV and CMV infections relapse only in severely immunocompromised individuals, not during normal immune aging. The immune system commits extraordinary resources to controlling CMV, and CMV-specific CD8 T cells can make up a large fraction of the entire T-cell repertoire. Whether this memory inflation has broader implications for immune health remains a matter of controversy. Expansion of the CMV-specific T cells may compromise the size of naïve and central memory T-cell repertoires; however, many of the CMV-specific CD8 T cells have the phenotype of end-differentiated effector T cells and are unlikely to compete for the same space as naïve cells.

End-differentiated T cells also express negative regulatory receptors of the killer lectin-like receptor (KLR), killer immunoglobulin-like receptor (KIR), and immunoglobulin-like transcript (ILT) families that appear to constrain their otherwise unopposed expansion. In spite of these inhibitory receptors, these cells are competent effector T cells capable of producing inflammatory cytokines and therefore may contribute to inflammation in older adults. Fig. 21.4 provides a comparison of these cells with age-associated B cells. End-differentiated T cells should be distinguished from exhausted CD8 T cells. T-cell exhaustion is seen with chronic stimulation by highly replicating viruses or tumor cells and is characterized by the expression of inhibitory receptors programmed death 1 (PD-1), T-cell immunoglobulin and mucin-domain containing-3 (TIM-3), and lymphocyte activation gene 3 (LAG3).[17] T-cell exhaustion is not a general feature per se of T-cell aging.

INFLAMMATION, AGING, AND THE AGING HOST ENVIRONMENT

The aging host environment is characterized by the continuous presence of inflammatory mediators (Fig. 21.5).[18] Low-level systemic inflammation plays an important role in the progression of several age-related diseases, including Alzheimer disease, atherosclerosis, and cancer. Moreover, inflammatory markers are associated with several conditions that are characteristic of older adults. IL-6 serum concentrations have been correlated with loss of mobility and advent of disability; increased mortality of older individuals has been shown among those who have higher levels of tumor necrosis factor-α (TNF-α). Increased IL-6 and cross-reactive protein (CRP) serum levels predispose to, and are associated with, frailty. A causative relationship may exist between the increased production of IL-6 or TNF-α and the age-associated loss in muscle mass, eventually presenting as sarcopenia. Moreover, long-lived individuals, such as centenarians, tend to have lower levels of pro-inflammatory cytokines and increased levels of anti-inflammatory mediators, such as cortisol and IL-10.

Production of inflammatory cytokines is driven by several mechanisms. Failure of the adaptive immune system leads to a less effective control of chronic viral infections as well as incomplete response to exogenous challenges, resulting in increased and prolonged innate immune activation. Defective epithelial barrier function, as well as decline in the mucosa-associated lymphoreticular tissue, results in increased leakage, increased systemic levels of lipopolysaccharide, and innate immune activation. Failure in maintaining T-cell population homeostasis and accumulation of effector T cells also favor an inflammatory response.

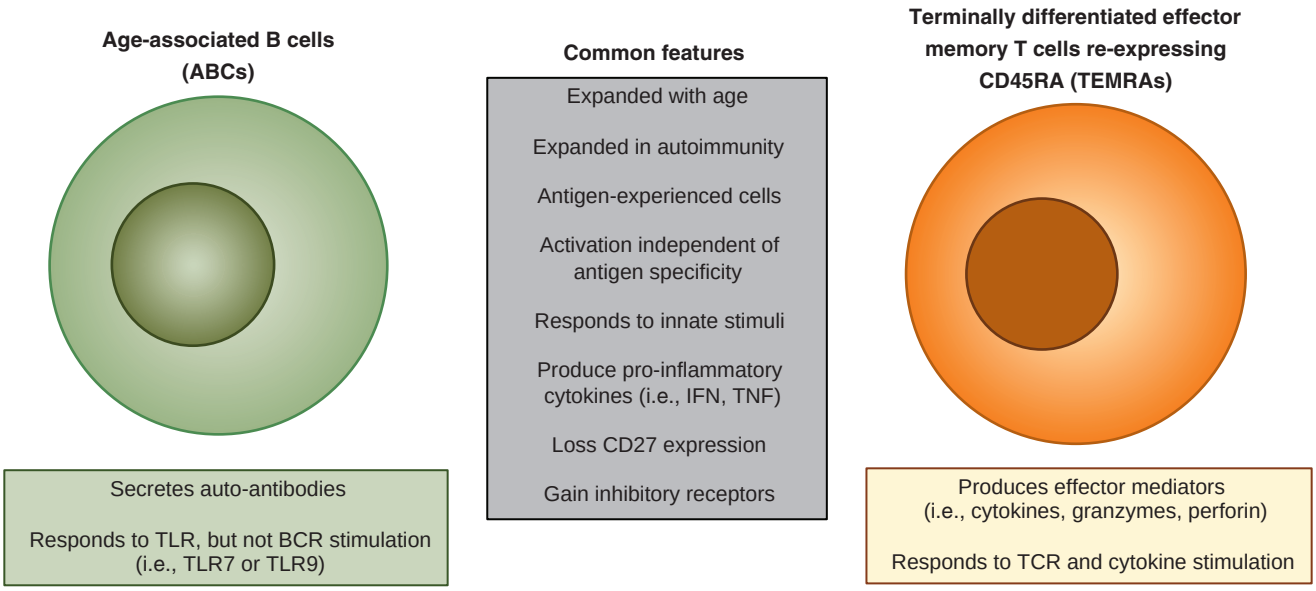


FIG. 21.4 Comparison of B- and T-Cell Subsets Expanded with Age. During aging, both the T- and B-cell compartments demonstrate expansion of pro-inflammatory memory populations with many shared features. These features include the ability to become active independent of antigen, the production of inflammatory cytokines, and alteration in stimulatory and inhibitory receptor profiles. *BCR*, B-cell receptors; *IFN*, interferon; *TNF*, tumor necrosis factor; *TCR*, T-cell receptors; *TLR*, Toll-like receptor.

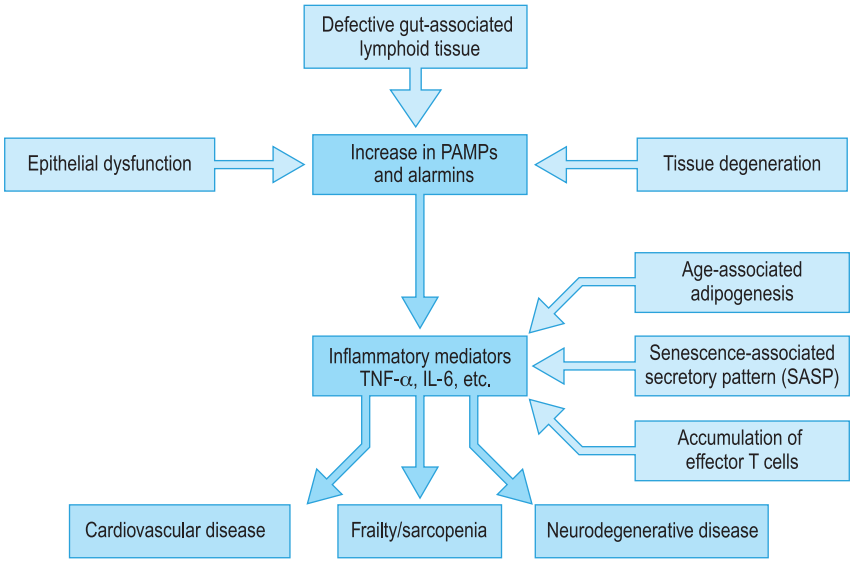


FIG. 21.5 Inflammation in Older Adults. The schematic diagram depicts possible mechanisms that account for the increased production of inflammatory mediators with age. These mediators contribute to many of the age-associated diseases. *IL-6*, Interleukin-6; *PAMPs*, pathogen-associated molecular patterns; *TNF-α*, tumor necrosis factor-α.

KEY CONCEPTS

Causes for Increased Constitutive Production of Inflammatory Mediators With Age (Inflammation in Older Adults)

- Activation of the innate immune system as a result of defective epithelial barrier function
- Activation of the innate immune system as a result of defective adaptive immunity
- Accumulation of adipocytes producing inflammatory mediators
- Accumulation and activation of effector T-cell populations
- DNA damage-induced transcription of inflammatory cytokine genes in senescent cells

The immune system, however, is not the only source of inflammatory cytokines. Adipocytes, in part replacing muscle cells in older adults, produce inflammatory mediators. Additionally, cellular senescence is associated with increased production of pro-inflammatory cytokines, where persistent DNA damage response signaling not only induces an irreversible cell cycle block that is characteristic of cellular senescence but also initiates a transcriptional program to secrete numerous growth factors, proteases, and inflammatory cytokines, termed the *senescence-associated secretory phenotype* (SASP).[19]

CELLULAR DEFECTS IN IMMUNE AGING

As described so far, immune aging occurs at the system level with organizational restructuring. Equally important are changes at the single-cell level, which are partly cell-intrinsic and partly caused by the host environment. The increased cytokine concentrations in older adults not only activate but also attenuate signaling pathways. Low responsiveness to cytokine stimuli is frequently seen in those cells that have increased baseline activation of a signaling pathway (e.g., cells that constitutively have increased signal transducer and activator of transcription 3 [STAT3] or STAT1 phosphorylation respond less to IL-6/granulocyte macrophage–colony-stimulating factor [GM-CSF] or type I/II interferons). Attenuation of signaling pathways by induction of negative feedback loops explains, in part, the reduced responsiveness and functionality of innate immune cells.

Although neutrophil and monocyte/macrophage numbers remain normal, many of their functions decline with age.[20] Decreased chemotaxis in neutrophils delays tissue infiltration; reduced phagocytosis and respiratory burst compromise the ability to control bacterial infections; and TLR-induced monocyte/macrophage activation is dampened in older adults. Declines in responsiveness, for example, to TLR stimulation, are partly reversible *in vitro*, suggesting that they are not intrinsic. Adaptive immune cells are also directly affected by the pro-inflammatory environment in the aging host.

Equally important are cell-intrinsic changes that appear to be a consequence of the replicative history and failures in cellular processes, such as DNA repair and autophagy. The program most obviously influenced by age is cellular senescence. In all hematopoietic cell lineages, including stem cells, telomeric lengths decline with age. Telomeric erosion results not only from cumulative replicative history and DNA damage but also from a decline in the ability to express telomerase and repair

KEY CONCEPTS

Cellular Dysfunction With Advanced Age

- Exposure to aging host environment (e.g., inflammatory cytokines) activates negative regulatory signaling loops.
- Telomeric erosion impairs proliferative competence and restraints clonal expansion.
- End-differentiation reduces functional plasticity.
- Activation of specific gene programs modifies cell function:
  - Gene programs associated with differentiation (e.g., microRNA [miRNA])
  - Loss of CD28 on T cells
  - Gain in natural killer (NK) cell–associated regulatory receptors on T cells (e.g., killer immunoglobulin-like receptor [(KIR], killer lectin-like receptor [KLR], immunoglobulin-like transcript [ILT])
- Senescence-associated gene activation (e.g., inflammatory mediators).

telomeric ends. This is of particular importance for T cells because much of their response depends on their ability to proliferate and clonally expand.

Lymphocytes in older adults are more differentiated than those in the young. Although most obvious for CD8 T cells, increasing differentiation can also be noted for B cells and CD4 T cells. Differentiation is generally driven by antigen recognition but could also occur in the absence of exogenous antigen under the influence of cytokines. Proliferation alone may be sufficient to drive initial steps of differentiation. A classic example is the acquisition of memory-like and effector phenotypes with lymphopenia-induced homeostatic proliferation. So-called virtual memory cells, which presumably have never seen an exogenous antigen, have been identified in mouse models. Some gene expression changes found in aged naïve T cells may represent partial differentiation, such as declines in miR-181a and changes in the expression of phosphatases and other signaling molecules. The downmodulation of miR-181a leads to reduced TCR-induced activation in these cells. Moreover, naïve CD8 T cells in the elderly display an epigenetic landscape more similar to that of memory T cells and a global upregulation of miR-146a, indicating underlying cellular activation.[14]

Within the memory T-cell compartment, changes in cell surface molecules that are seen with terminal differentiation, such as the gain in CD57 and the loss of CD27 and CD28 expression, are the most striking. Of functional importance, predominantly for CD8 T cells, is the gain in expression of cell surface receptors that are usually only found in NK cells. Most of these receptors have inhibitory function, but some of them also stimulate. Since expression of these receptors on individual cells is stochastic, the consequences can range from immunosuppression to autoreactivity.

CLINICAL CONSEQUENCES OF IMMUNE AGING—IMMUNODEFICIENCY, AUTOIMMUNITY, AND ACCELERATED DEGENERATIVE DISEASES

The most profound and most noted consequence of immune aging is the increased susceptibility to infections. Upper respiratory tract and urinary tract bacterial infections are frequent in the older population. Despite annual vaccination, influenza infections continue to be associated with high morbidity and mortality in older individuals. Pneumonia caused by RSV, usually infecting young children, is also common with advancing age, as is the loss of immune competence to combat chronic infections, as described above. Not surprisingly, the immune system of an older adult is not able to induce a protective response to new antigens to which the individual has not been exposed in the past. Clinically important examples are the COVID-19 pandemic, severe acute respiratory syndrome (SARS) epidemic, and West Nile fever virus infection, all of which severely affected the older population. Moreover, first-time vaccinations with live viruses (e.g., yellow fever virus) are associated with increased morbidity and even mortality in older adults.

Immune aging also predisposes for autoimmune manifestations and a breakdown in self-tolerance.[21] Autoantibodies are a common finding in healthy older adults; many of these autoantibodies are specific for common autoantigens, such as IgG Fc or nuclear components. The risk for several autoimmune diseases, most notably polymyalgia rheumatica and giant cell arteritis, increases with age. Although polymyalgia rheumatica predominantly presents as an activation of innate immunity, giant cell arteritis is clearly a disease of the adaptive immune system with T cell–dependent granulomatous inflammation in the vascular wall of midsized and large arteries.

The low-grade inflammation in the aging host has direct clinical consequences in promoting frailty and sarcopenia as well as accelerating degenerative diseases, such as coronary artery disease, osteopenia, and Alzheimer disease. Accelerated immune aging may be one of the reasons that autoimmune diseases, such as rheumatoid arthritis (RA), are associated with a shorter life span and increased risk for cardiovascular morbidity. Inflammation as a manifestation of accelerated aging has also been implicated in the increased morbidity and mortality of patients with HIV infection in spite of highly active antiretroviral therapy (HAART).[22]

CLINICAL RELEVANCE

Consequences of Immune Aging

- Increased morbidity and mortality from bacterial infections (e.g., pneumococci)
- Increased morbidity and mortality from viral infections (e.g., SARS-CoV-2, influenza, West Nile fever)
- Reactivation of latent virus (e.g., varicella-zoster virus)
- Ineffective primary and booster vaccinations
- Acceleration of degenerative diseases as a result of the production of inflammatory mediators (e.g., atherosclerotic disease, Alzheimer disease, osteoarthritis)
- Increased incidence of autoimmune disease (e.g., polymyalgia rheumatica, giant cell arteritis, rheumatoid arthritis)

STRATEGIES AND INTERVENTIONS ON THE HORIZON

Vaccinations hold the promise for reducing susceptibility to infections in young children and older adults but improving vaccine responses has proven to be a challenge. Current strategies for targeting more elusive pediatric pathogens (e.g., RSV) include maternal vaccination and the development of infant-specific vaccine adjuvants utilizing newer understanding of the developing infant's immune system. Further investigations into the role of microbiome and immune regulation may open novel strategies to influence immune system development to reduce immune deficiencies and prevent inappropriate hyperactivity, such as allergies and asthma. Newly developed high-dimensional techniques, such as mass cytometry and single-cell RNAseq, will provide a more comprehensive and mechanistic understanding of immune system development to improve vaccine design and subsequent efficacy in infants.

In older adults, a proper T-cell response to infectious organisms or to vaccines depends on the availability of a T-cell repertoire that includes antigen-receptor specificities that respond to vaccination. Repertoire contraction resulting from thymic involution or memory inflation appears to be less severe than originally thought, and currently explored interventions to restore thymic activity may be meaningful only for selected populations, such as bone marrow recipients. Deficiency in the antigen-presenting system and in costimulatory signals may be overcome by identifying new adjuvants. Increasing the vaccine dose (by utilizing higher antigen doses for older individuals, for example) is another promising approach. Live vaccines or

self-replicating constructs that also accomplish higher antigen loads may not have a sufficient safety profile in older adults. Direct targeting of signaling defects in older T cells resulting from increased expression of cytoplasmic phosphatases or inhibitory cell surface receptors will be feasible, as shown for checkpoint inhibitors in oncology; however, such approaches need to demonstrate a much better safety profile.

Interventions to influence inflammation in older adults at present needs to be nonspecific, given the multitude of underlying mechanisms. Immunomodulatory therapy is obviously standard practice in patients with autoimmune disease who exhibit accelerated aging and increased all-cause mortality; however, such therapy is associated with infections even in young adults. Calorie restriction to slow immune aging to an extent that it is effective is generally not well accepted. Statins and aspirin are being routinely used to prevent cardiovascular disease; their effect may be mostly anti-inflammatory. Future interventions will require the development of mild and low-toxicity medications to reduce low-grade inflammation while not impairing the ability to prevent immune activation in response to harmful antigens. One interesting concept is to deplete senescent cells that may entertain the low-grade inflammation in the elderly through the (SASP) mechanism. Several such senolytic drugs have entered clinical trials. An alternative approach is to restore the ability of the immune system to clear senescent cells.

ON THE HORIZON

- Improved vaccination strategies specifically tailored to the infant or aged immune system (novel adjuvants, novel vaccine delivery systems)
- New vaccines for pregnant females to confer passive immunity
- Manipulation of the microbiome composition to influence immune system development
- Thymic rejuvenation (*e.g.* with KGF, IL-7 and other mediators)
- Prevention of chronic infection that accelerate immune aging (*e.g.* immunization for CMV)
- Pharmacological approaches to improve T- and B-cell activation, clonal expansion and differentiation
- Treatment of inflamm-aging

REFERENCES

1. Lopez-Otin C, Blasco MA, Partridge L, Serrano M, Kroemer G. The hallmarks of aging. *Cell*. 2013;153(6):1194–1217.
2. Gollwitzer ES, Marsland BJ. Impact of early-life exposures on immune maturation and susceptibility to disease. *Trends Immunol.* 2015;36(11):684–696.
3. Basha S, Surendran N, Pichichero M. Immune responses in neonates. *Expert Rev Clin Immunol.* 2014;10(9):1171–1184.
4. Tsafaras GP, Ntontsi P, Xanthou G. Advantages and limitations of the neonatal immune system. *Front Pediatr.* 2020;8:5.
5. Simon AK, Hollander GA, McMichael A. Evolution of the immune system in humans from infancy to old age. *Proc Biol Sci.* 2015;282(1821):20143085.
6. Rudd BD. Neonatal T cells: a reinterpretation. *Annu Rev Immunol.* 2020;38:229–247.
7. Siegrist CA, Aspinall R. B-cell responses to vaccination at the extremes of age. *Nat Rev Immunol.* 2009;9(3):185–194.
8. Gensollen T, Iyer SS, Kasper DL, Blumberg RS. How colonization by microbiota in early life shapes the immune system. *Science.* 2016;352(6285):539–544.
9. Geiger H, de Haan G, Florian MC. The ageing haematopoietic stem cell compartment. *Nat Rev Immunol.* 2013;13(5):376–389.
10. Jaiswal S, Ebert BL. Clonal hematopoiesis in human aging and disease. *Science.* 2019;366:6465.
11. Hagen M, Derudder E. Inflammation and the alteration of B-cell physiology in aging. *Gerontology.* 2020;66(2):105–113.
12. Cancro MP. Age-associated B cells. *Annu Rev Immunol.* 2020;38:315–340.
13. Thomas R, Wang W, Su DM. Contributions of age-related thymic involution to immunosenescence and inflammaging. *Immun Ageing.* 2020;17:2.
14. Goronzy JJ, Weyand CM. Mechanisms underlying T cell ageing. *Nat Rev Immunol.* 2019;19(9):573–583.
15. Gustafson CE, Kim C, Weyand CM, Goronzy JJ. Influence of immune aging on vaccine responses. *J Allergy Clin Immunol.* 2020;145(5):1309–1321.
16. Nikolich-Zugich J. The twilight of immunity: emerging concepts in aging of the immune system. *Nat Immunol*. 2018;19(1):10–19.
17. Wherry EJ, Kurachi M. Molecular and cellular insights into T cell exhaustion. *Nat Rev Immunol.* 2015;15(8):486–499.
18. Furman D, Campisi J, Verdin E, et al. Chronic inflammation in the etiology of disease across the life span. *Nat Med.* 2019;25(12):1822–1832.
19. Khosla S, Farr JN, Tchkonia T, et al. The role of cellular senescence in ageing and endocrine disease. *Nat Rev Endocrinol.* 2020;16(5):263–275.
20. Goldberg EL, Shaw AC, Montgomery RR. How Inflammation blunts innate immunity in aging. *Interdiscip Top Gerontol Geriatr.* 2020;43:1–17.
21. Goronzy JJ, Li G, Yang Z, et al. The janus head of T cell aging—autoimmunity and immunodeficiency. *Front Immunol.* 2013;4:131.
22. Kaplan-Lewis E, Aberg JA, Lee M. Aging with HIV in the ART era. *Semin Diagn Pathol.* 2017;34(4):384–397.

22

The Microbiota in Immunity and Inflammation

Craig L. Maynard

Humans and other species of mammals are hosts to an array of microbial communities collectively referred to as the *microbiota*. The microbiota comprises prokaryotes (bacteria and archaea), viruses (bacteriophages as well as eukaryotic viruses), and eukarya or the meiofauna (mainly fungi and protozoa). As revealed by the Human Microbiome Project, specialized subcommunities colonize barrier surfaces of the digestive, respiratory, and urogenital tracts and skin. Our immune system is viewed as having evolved to ensure peaceful coexistence with these microorganisms that aid in immune homeostasis, pathogen resistance, and digestion in exchange for a nutrient-rich habitat. There is a perpetual crosstalk between the microbiota and the immune system throughout an individual's life. Specific modulation of this microbiota, particularly in infancy, has important long-term health consequences. Addition of specific symbionts to the microbiota can provide tangible health benefits. Reconstitution of a dysbiotic microbiota continues to be utilized or explored as therapy for inflammatory diseases.

OVERVIEW OF OUR NONMAMMALIAN "SELF"

Exposure to the microbiota and its products occurs via a gradual, ordered process. It begins in utero and accelerates during and after birth. The overall composition of an individual's microbiota is impacted by such phenomena as the method of birth delivery, diet, treatment with antibiotics, and environmental exposures. All of these modifying factors can have lasting effects on immune health.

Prokaryotes

The bacterial component of our microbiota is the most widely studied by far. This has been made possible, in part, by a fairly recent explosion of genomic and bioinformatics capabilities that has enabled taxonomic identification, and even enumeration, of the members of microbial communities without the need for bacterial culture. Two phyla—Bacteroidetes and Firmicutes—account for almost 90% of intestinal bacteria. The remainder comprises organisms belonging to the phyla Proteobacteria (the third largest), Cyanobacteria, Proteobacteria, Actinobacteria, Fusobacteria, and Verrucomicrobia.[1] It is estimated that the average human being plays host to approximately 100 trillion bacteria. Most of these cells inhabit the lumen of the gastrointestinal (GI) tract. The density of bacteria increases from $\approx 10^2$ to 10^3 bacterial cells per mL in the stomach to 10^{12} cells/mL in the distal large intestine (Fig. 22.1). The sheer number of bacteria sets the stage for a very complex relationship between the mammalian immune system and the microscopic inhabitants of the various barrier surfaces.

Viruses

The viruses that inhabit mammalian hosts can be subdivided into bacteriophages, which infect prokaryotic cells; eukaryotic viruses, which infect host and other eukaryotic cells; and virus-derived genetic elements, which can incorporate into host chromosomes and result in the generation of infectious virus at a later date. It has been difficult to quantify the exact size of the *virome*. In the case of bacteriophages, it is generally accepted that they can exist in 10-fold greater numbers than prokaryotes. Bacteriophages can have profound effects on the structure and functions of intestinal prokaryotic communities via viral gene transfer of virulence factors and antibiotic-resistant genes between prokaryotic organisms or through predator–prey relationships. Indeed, the infectivity of some enteric viruses requires the microbiota.[2] Eukaryotic viruses include a vast array of viruses that permanently infect the host and can exist for decades in asymptomatic individuals. These viruses can persist locally or systemically. They can directly impact tissue-specific immunity, including in the GI tract.

Fungi

Collectively referred to as the *mycobiota*, fungal communities represent a considerably smaller proportion of the total microbes in the human body. Commensal fungi can be detected in the mouth, lungs, intestines, vagina, and skin (Fig. 22.2). Advances in our knowledge and understanding of the size and functions of the mycobiota have been hampered by relatively limited genomics and bioinformatics capabilities when compared with the study of bacteria or viruses. However, studies utilizing broad-spectrum antifungal agents have begun to highlight the possible roles played by fungi in protecting against disease processes, such as inflammatory bowel disease (IBD) (Chapter 75).

The microbes that inhabit the human body are often acceptably referred to as the *commensal microbiota*, which in the strictest sense refers to organisms that derive benefit from their host without negatively or positively affecting the host. Although this is true of some members of the microbiota, our relationship with other microbes is one of *mutualism*, whereby each organism performs unique and necessary functions that benefit the other. For example, certain bacteria find a home in the anaerobic environment of the cecum and proximal colon, where they receive a rich source of nutrients in the form of insoluble carbohydrates, which our own digestive enzymes are unable to process. By the process of anaerobic fermentation, these carbohydrates are broken down to generate short-chain fatty acids including butyrate, which is utilized preferentially by colonocytes and also impacts host immunity and metabolism.

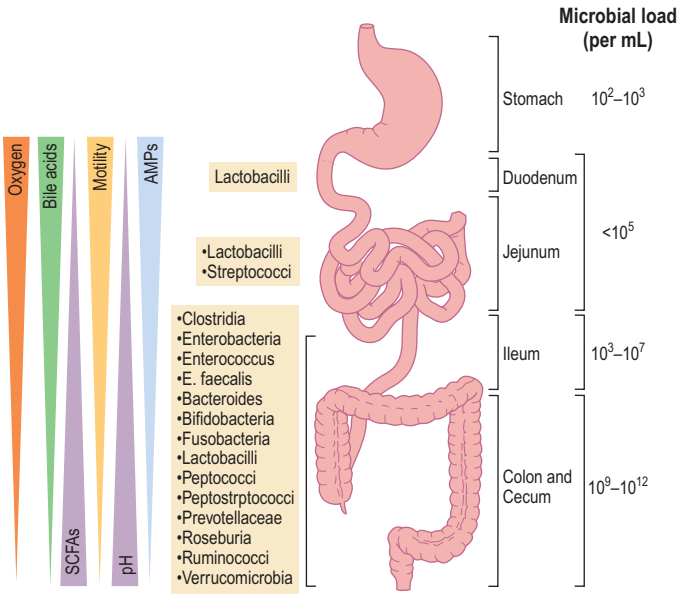


FIG. 22.1 Spatial Organization of Microbial Communities and Physiological Gradients Along the Mammalian Gastrointestinal (GI) Tract. The numbers and types of bacterial communities, as well as physiological factors, vary along the length of the GI tract. It is well-appreciated that the oxygen levels, bile acid concentrations, intestinal motility, antimicrobial peptides *(AMPs)*, and luminal pH in the proximal portion of the GI tract (stomach, duodenum, jejunum) play major roles in restricting the numbers and types of microorganisms. In general, aerobic and facultative anaerobic bacteria are found almost exclusively in the proximal portion of the GI tract. The hypoxic nature and more physiological pH of the distal small intestine (ileum) and colon, coupled with overall reductions of bile acids, AMPs, and gut motility, allows for unfettered growth of large numbers of obligate anaerobic bacteria. These oxygen-sensitive microbes are capable of producing large quantities of short-chain fatty acids (*SCFAs*; acetate, propionate, butyrate) from complex carbohydrates (fiber) to be used for important colonic and immunological processes. (From Reinoso Webb C, Koboziev I, Furr KL, Grisham MB. Protective and pro-inflammatory roles of intestinal bacteria. *Pathophysiology*. 2016;23:67–80, Fig. 2.)

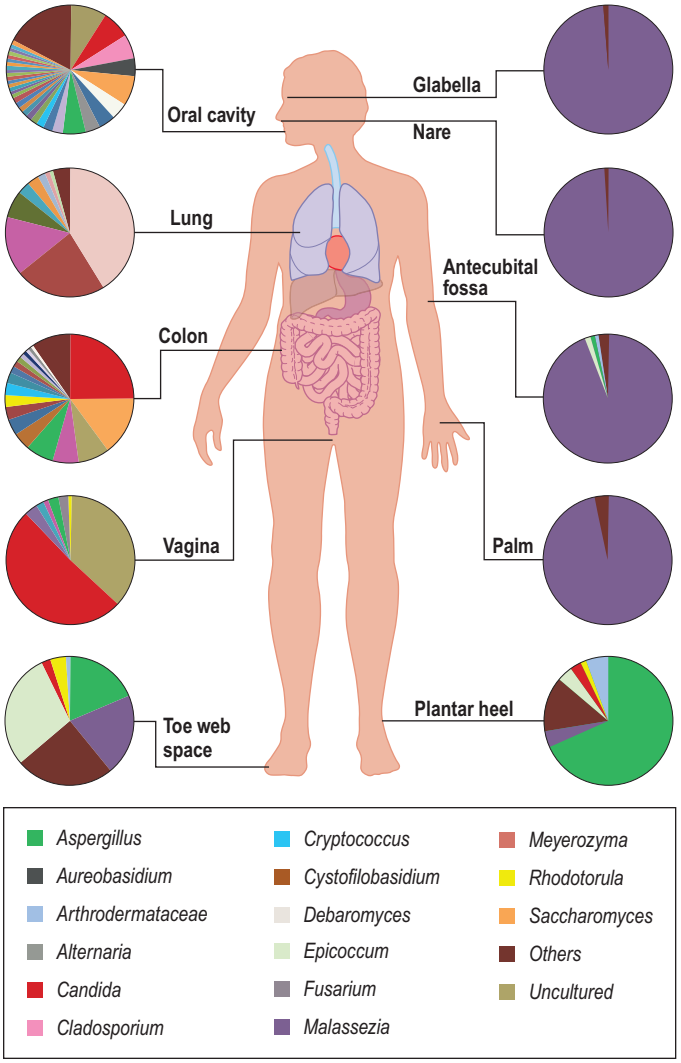


FIG. 22.2 The Human Mycobiota. Complex populations of fungi have been found associated with the skin and all mucosal surfaces of the healthy human body. The *pie charts* indicate the relative proportions of fungal genera that are reported to be associated with the respective sites in representative fungal deep-sequencing studies. The fungal populations that are found on mucosal surfaces tend to be more diverse than those on the skin. The healthy lung probably reflects mostly environmental fungi, which are not included in the key. *Others* refers to sequences that represent <1% of the total recovered sequences at each site. *Uncultured* refers to sequences identified in the National Center for Biotechnology Information (NCBI) GenBank database as fungal, but of uncharacterized, origin. Data for pie charts were derived from studies of the fungal genera that are present in the oral cavity, lungs, colon, vagina, and skin. (From Underhill DM, Iliev ID. The mycobiota: interactions between commensal fungi and the host immune system. *Nat Rev Immunol*. 2014;14:405–416, Fig. 1.)

Obviously, parasitic or harmful microbes are not considered part of the commensal microbiota, although organisms that fit this definition can often coexist with the microbiota without driving overt disease under homeostatic conditions. The term *pathobiont* was previously coined to refer to any microbe that peacefully colonizes its host but can evoke severe inflammatory responses under specific genetic and/or environmental conditions.

THE IMMUNE SYSTEM FACILITATES MICROBIAL COLONIZATION

The developing fetus may be exposed in utero to a microbiota or microbial products acquired via the placenta and/or maternal circulation.[3] The acquisition of microbial communities continues with the passage of the fetus through the birth canal and culminates in the first 2 to 5 years of life in humans. Despite the vast array of microbes that take up residence in the host, colonization is usually an ordered process that—in most cases—is well tolerated. This is because of a robust antimicrobial defense system that is initiated prenatally and fortified postnatally, simultaneous with rapid microbial colonization.

KEY CONCEPTS

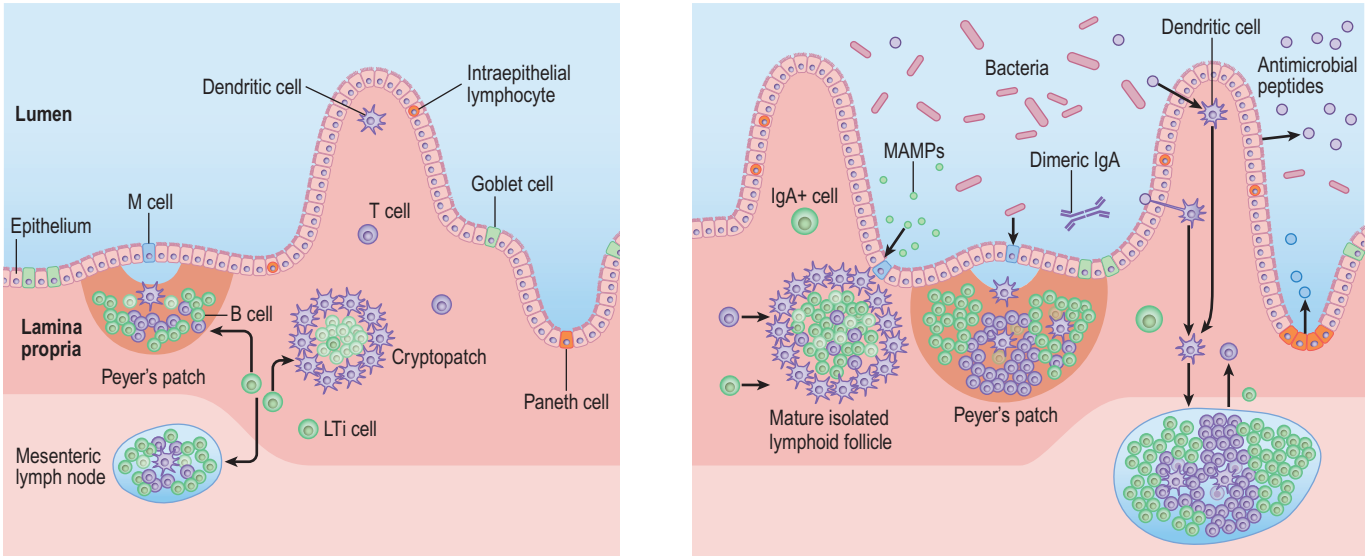
Definitions

- **Microbiota**: A collective term for all the microscopic organisms that reside on or in the human body.
- **Microbiome**: The combined genomes of all the organisms that constitute the microbiota.
- **Mycobiota**: That subset of the microbiota that includes fungi alone.
- **Virome**: The collection of all viruses, including viruses integrated into the human genome, found in or on humans.
- **Dysbiosis**: A condition in which there is disequilibrium of the microbial communities that constitute the microbiota at a given body site.
- **Germ-free**: Experimental animals birthed and raised in a sterile environment, devoid of microbes.
- **Gnotobiotic** ("known life"): Describes animals in which the full complement of colonizing microbes is known.

Prenatal Development of the Immune System

In the fetal liver, some common lymphoid progenitor cells—ancestral to all lymphocytes—develop into a specialized subset of innate lymphoid cells (ILCs) referred to as *lymphoid tissue inducer (LTi) cells*.[4] As their name implies, LTi cells are essential for the development of all secondary lymphoid structures throughout the body. These structures will eventually become the sites for initiation of immune responses to the commensal microbiota, pathogenic invaders, and self-antigens. In the developing fetus, LTi cells promote the development of mesenteric lymph node (MLN) and of the Peyer patches (PPs) (Chapters 2 and 24) in the distal ileum (Fig. 22.3). They also recruit B and T lymphocytes to these tissues and facilitate their organization into distinct B-cell follicles and T-cell zones, respectively. Throughout life, the MLN provides a so-called mucosal firewall that prevents systemic dissemination of gut bacteria.

Other local mechanisms are also initiated to limit host collateral damage to the neonatal intestine by early microbial encroachment. For example, Toll-like receptor 4 (TLR4) (Chapter 3), the receptor for lipopolysaccharide (LPS) that is derived from gram-negative bacteria, is highly expressed by intestinal epithelial cells (IECs) prior to birth, but its expression and signaling are rapidly downregulated following onset of colonization. In addition, a diverse array of lymphocytes



(A) Prenatal (B) Postnatal

FIG. 22.3 The Gut-Associated Lymphoid Tissue (GALT) Establishes Perinatal Host–Microbiota Mutualism in the Intestine. (A) Prenatally, secondary lymphoid tissues (Peyer patches and mesenteric lymph nodes) and cryptopatches develop by the spatiotemporal recruitment of lymphoid tissue inducer *(LTi)* cells to sites of the developing intestine and supporting neurovascular structures. This, in turn, stimulates the recruitment of dendritic cells (DCs), T cells, and B cells in preparation for the immune response to the microbiota. Intraepithelial lymphocytes (IELs) seed the epithelium before birth. (B) Postnatally, bacteria colonize the neonatal intestine immediately, initiating multiple events that affect the development or functional maturation of the mucosa and GALTs. *Shown from left to right:* Microbe-associated molecular patterns *(MAMPs)* sensed by pattern-recognition receptors on intestinal epithelial cells and DCs adjacent to cryptopatches stimulate the further recruitment of B cells and T cells, causing the cryptopatches to develop into mature isolated lymphoid follicles. The isolated lymphoid follicles release immunoglobulin A *(IgA)*–producing plasma cells, which are formed through T-cell–dependent and –independent interactions, into the lamina propria. Microbes also cross the epithelium and enter the Peyer patch through M cells, from which they are endocytosed by the DCs in the subepithelial dome. Antigen-loaded DCs in the Peyer patch interact with local lymphocytes to induce T-cell differentiation and T-cell–dependent B-cell maturation in the germinal center (GC) to induce the development of IgA-producing plasma cells that home to the lamina propria, where they release dimeric IgA for transport into the intestinal lumen. DC-mediated luminal sampling of microbial products or transcytosis of bacteria across the epithelium results in antigen loading of the lamina propria DCs, which then migrate through the afferent lymphatics vessels (not shown) to a draining mesenteric lymph node to induce differentiation of effector T cells that traffic to the lamina propria. *Shown on the far right*: Sensing of MAMPs stimulates the proliferation of intestinal epithelial cells in crypts, resulting in their increased depth and, in the small intestine, increased density of Paneth cells. This sensing also arms the intestinal epithelial cells for release of antimicrobial peptides. (From Maynard CL, Elson CO, Hatton RD, Weaver CT. Reciprocal interactions of the intestinal microbiota and immune system. *Nature*. 2012;489:231–241, Fig. 1.)

collectively referred to as intraepithelial lymphocytes (IELs) can be found intercalated between the IECs. IELs display multiple features of activated cells and participate in the maintenance of epithelial barrier integrity by limiting bacterial translocation and promoting epithelial repair following injury.[5]

Reproductive Tract Microbiota and the Immune System in Perpetuation of the Species

As viviparous organisms, the microbiota of the human genital tract appears to be a critical component of a healthy reproductive system. Dysbiosis in the genital tract is implicated in susceptibility to infection, infertility, and recurrent miscarriage as well as premature delivery and placental malfunction.[6] The human vagina harbors a stable community of microorganisms which, similar to the microbiota at other sites, figures prominently in the maintenance of immune homeostasis and resistance to pathogen colonization.

The vaginal microbiota exists in a state of dynamic equilibrium with minor fluctuations induced by such factors as hormones associated with the menstrual cycle, sexual behavior, and personal hygiene. This community is composed primarily of several species of the genus *Lactobacillus*. In non-pregnant women of reproductive age, the vaginal microbiota is relatively dynamic in comparison to the more stable composition present during pregnancy. The dominance of a few species is believed to be central to the relative stability and colonization resistance of the vaginal microbiota, especially during development of the fetus in the adjacent uterus. Though still dominated by *Lactobacillus*, the dominant species detectable appears to vary with gestational age, history of pregnancy, and ethnicity. For example, *L. crispatus*, *L. iners*, and *L. acidophilus* were the most dominant in Caucasian, African American, and Hispanic cohorts, respectively. Historically, the uterus, and the placenta have been considered sterile. However, due primarily to the emergence of culture-independent methods for microbiota detection, it has been suggested that there exist low biomass microbial communities in both the endometrium and placenta, but this concept remains somewhat controversial. Nevertheless, dysbiosis of the vaginal reproductive tract microbiota is believed to contribute to reproductive difficulties and pregnancy complications that adversely affect the health of the mother, offspring, or both.

Lactobacillus spp. in the upper genital tract are believed to protect this region from infection via the production of lactic acid, maintaining a low pH of less than 4.5, which is prohibitive to the growth of pathogens. The concept of an essential role for *Lactobacillus* reproductive health is supported by an increase in the relative abundance of such species during puberty in response to the hormone estradiol. In general, a beneficial vaginal microbiota promotes robust expression of defensins and specific vaginal antimicrobial peptides (AMPs). In contrast, increased expression of vaginal AMPs such as the secretory leukocyte protease inhibitor and human epididymis protein 4 have been associated with a less beneficial vaginal microbiota. This microbiota must also facilitate the induction of an immune milieu characterized by reduced proinflammatory cytokines and sustained immune regulatory pathways.

Passive Acquisition of Antimicrobial Immunity

The microbiota of the neonate is acquired from his or her mother, who has already established a tolerogenic relationship with the same microbiota. The composition of the transmitted microbiota varies initially with the mode of delivery—vaginal versus cesarean section. The mother's "mucosal memory" also gets transmitted to her offspring. During transvaginal delivery, this mucosal memory is seeded with the mother's native microbiota. Mothers produce antibodies to bacteria-derived antigens. These bacteria-responsive antibodies enter the maternal circulation and ultimately get passed to the offspring in breast milk.

Immunoglobulin A (IgA) is the major antibody isotype (Chapter 8) generated by mammary glands. IgA inhibits bacterial translocation across the neonatal intestinal epithelium, thereby limiting collateral damage by an encroaching microbiota and providing passive immunity to pathogenic infection.[7] The offspring also acquires antibodies of the IgG isotype both in utero and via the breast milk. These antibodies serve to limit enteric infection in the newborn[8] and dampen neonatal mucosal T-cell and germinal center (GC) B-cell responses to commensal antigens.[9]

Maternally acquired anti-commensal antibodies can transfer bound microbial molecules to the offspring during gestation and via the breast milk. This transfer of microbial products contributes to the earliest education of the immune system and limits deleterious postnatal inflammatory responses.[3] Breast milk is also a rich source of immunosuppressive transforming growth factor-β (TGF-β) and interleukin-10 (IL-10) (Chapter 14), which also help promote tolerogenic responses to the microbiota. Recent studies suggest that microbes are also passed to the newborn via the breast milk.

MICROBIOTA-DEPENDENT MATURATION OF THE INTESTINAL IMMUNE SYSTEM

Microbial colonization prompts a rapid organization of immune structures that are quickly seeded with immune cells. This process helps avoid overexuberant responses to the microbiota and sets the stage for continued tolerance of commensals and for defense against pathogens in the future. The intestinal epithelium limits direct encounter between luminal microbes and the immune cells in the underlying lamina propria by forming a physical barrier and by the production and/or transport of immune and antimicrobial factors (Fig. 22.4).

Gut-Associated Lymphoid Tissues

Although MLN and Peyer patches begin to develop before birth, complete maturation does not occur until after birth. Germ-free or "germ-reduced" mice display reduced size and cellularity and altered numbers and distribution of immune cells both in the gut and gut-associated lymphoid tissues (GALTs). Thus, maturation of the mucosal immune system is contingent on the acquisition of microbiota. A third type of secondary lymphoid tissue—isolated lymphoid follicles (ILFs), which are also induced by LTi cells—is also completely dependent on colonization with the microbiota and thus develops postnatally.

LTi cells cluster at the base of the crypts in structures referred to as *cryptopatches*. Stimulation of cryptopatches by peptidoglycan derived from gram-negative bacteria induces recruitment of B cells, thereby forming ILFs. The importance of ILFs in the direct control of bacterial growth is demonstrated by the fact that mice devoid of mature ILFs display an overrepresentation of gram-negative bacteria.[10]

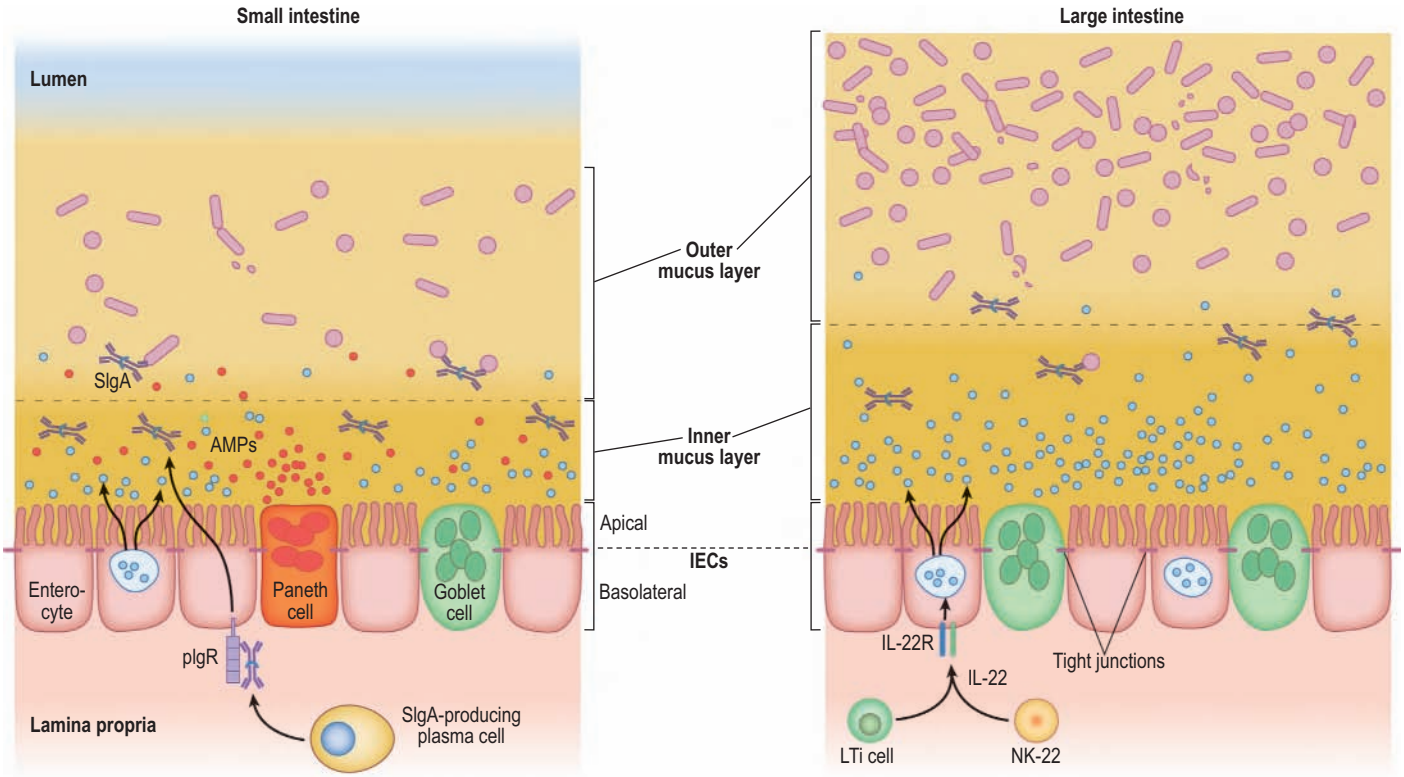


FIG. 22.4 Active Barrier Functions of the Intestinal Epithelium. The intestinal epithelium comprises a single layer of polarized columnar epithelial cells *(IECs)*, which is sealed by tight junctions. Specialized cells known as *goblet cells* secrete mucins that form a bilayered mucus sheath, which maintains separation between the luminal bacteria and the epithelium. The thicker *inner mucus layer* is sparsely populated with bacteria, whereas the *outer mucus layer* is more loosely structured and contains ≈10-fold more bacteria. Paneth cells within the epithelium secrete defensins that, along with epithelial-derived antimicrobial peptides, further help with bacterial containment. Epithelial cells also transport antibodies into the lumen and transmit cytokine signals that help maintain barrier integrity. *AMP*, Antimicrobial peptide; *Ig*, immunoglobulin; *IT*, interleukin; *LTi*, lymphoid tissue inducer; *NK*, natural killer; *SIga*, secretory IgA; *pIgR*, polymeric immunoglobulin receptor.

Innate Lymphoid Cells

ILCs represent an early line of defense at mucosal surfaces. ILCs are classified as LTi or "LTi-like" (described above) cells or as "helper-like" cells. The progenitor of helper-like ILCs is distinct from that of LTi cells or natural killer (NK) cells. "Helper-like" ILCs are classified into three main subcategories based on their expression of particular sets of surface receptors, transcription factors, and secreted proteins. Unlike their adaptive, thymus-derived counterparts, ILCs do not express antigen receptors and are therefore activated by cytokine signals. This enables ILCs to launch an initial rapid response to microbial challenge and facilitate the development of an adaptive lymphocyte-mediated immune response.

ILC1s, such as T-helper type 1 (Th1) cells, express the transcription factor Tbet and secrete interferon-γ (IFN-γ) and tumor necrosis factor-α (TNF-α) in response to intracellular bacteria. Similar to Th2 cells, ILC2s express GATA3 and secrete IL-13 and IL-5. ILC3s are arguably the most complex subset of ILCs and can be subdivided into "LTi-like" ILC3s, which populate the intestinal lymphoid follicles, and the more plastic NKp46$^+$ ILC3s, which are distributed throughout the intestinal lamina propria. Both subsets express retinoic acid–related orphan receptor (ROR)γt and produce IL-22 and IL-17, although NKp46$^+$ ILC3s can also express Tbet and secrete IFN-γ.[11]

The impact of the microbiota on ILC3 accumulation appears to commence in utero, as transient colonization of pregnant germ-free mice with nonpathogenic *Escherichia coli* increases accumulation of NKp46$^+$ ILC3s and F4/80+CD11c+ mononuclear cells in newborn pups.[3] In the normal intestine, ILC3s are the dominant source of IL-22.[11] The microbiota, acting on innate cells, including macrophages and dendritic cells (DCs), induces expression of IL-23, which, in turn, stimulates the production of IL-22 and/or IL-17 by ILC3. IL-22 can induce the secretion of antimicrobial peptides, such as RegIIIγ, by intestinal epithelial cells to limit bacterial invasion and within GALT to prevent systemic dissemination of resident microbes. Throughout life, ILC3 expressing major histocompatibility complex (MHC) class II molecules (Chapter 5) on their cell surface help regulate the pool size and function of commensal-specific CD4 T cells to prevent spontaneous inflammation.[12]

Invariant Natural Killer T Cells

A unique subset of thymus-derived cells, invariant NK T (iNKT) cells (Chapter 12) express an invariant T-cell receptor (TCR) (Chapter 4) that recognizes lipid antigens presented by the nonclassic MHC class I molecule CD1d (Chapter 5). iNKT cells are essential for the ulcerative colitis–like phenotype that develops in experimental mice treated with the chemical compound oxazolone.

The role of the microbiota in iNKT cell expansion and function provides an example of how early postnatal colonization can impact immune maturation and long-term function.

Germ-free mice harbor increased numbers of iNKT cells in both the colon and the lung and correspondingly demonstrate increased severity of experimental colitis and asthma. The disease phenotype can be reduced to that of conventionally raised mice if germ-free mice are colonized with commensals in the neonatal period, but not during adulthood.[13]

Regulatory T Cells

Regulatory T cells (Tregs) (Chapter 13) are crucial to the establishment and maintenance of immune homeostasis. Mice or humans with absent or defective expression of the signature Treg transcription factor Foxp3 (FOXP3 in humans) succumb to spontaneous multiorgan autoimmune disease very early in life. Commencing in the first few days of life, thymic FOXP3$^+$ Tregs seed every lymphoid and nonlymphoid tissue. Most Tregs in the intestine express Foxp3, and most of these coexpress IL-10, TGF-β, and IL-35. However, other FOXP3$^-$ subsets also exist, including IL-10-producing Tr1 cells and human lamina propria CD4CD8αα T cells.[14]

Commensal-specific Foxp3-expressing Tregs in the intestinal lamina propria are viewed as a mixture of thymic-derived Tregs (tTregs) and peripherally induced Tregs (pTregs).[15] The inability to generate pTreg cells results in spontaneously intestinal pathology late in life due to their essential role in tolerizing the host to colonization by epithelial "border-dwelling" organisms.[16] Colonization of germ-free mice with a benign cocktail of eight commensal microbes, collectively referred to as *altered Schaedler flora*, is sufficient to induce expansion and accumulation of Foxp3$^+$ cells in the large intestine, approaching the levels seen in conventionally raised mice. However, a case has been made for the ability of specific bacteria and/or bacterial components to uniquely promote Treg induction. For example, a cocktail of 46 mouse *Clostridium* strains potently induced expansion of colonic Foxp3$^+$ cells.[17] This cocktail, which has since been reduced to 17 strains derived from a single human donor, recapitulates the phenotype in germ-free mice.

Several factors produced or induced by gut microbes have been shown to be essential for microbiota-dependent Treg accumulation, particularly in the colonic lamina propria. The capsular polysaccharide A (PSA) moiety of *Bacteroides fragilis*, which mediates the interaction between the bacterium and the colonic mucosa, can also act via the TLR–MyD88 pathway to promote expansion of IL-10–producing colonic Tregs.[18] Clostridial strains stimulate production of TGF-β by intestinal immune cells and are very effective at limiting pathogen colonization of the GI tract.[19] This, in turn, enhances Foxp3$^+$ Treg induction. Clostridia are also very adept at anaerobic fermentation of indigestible fiber. Their production of butyrate may also enhance accumulation of Foxp3$^+$ Tregs in the colonic lamina propria.[20]

CD4 T-Helper Cells

In the intestines of both mice and humans, both IFN-γ–producing Th1 and T-helper "IL-17–producing" (Th17) cells are present (Chapter 11). These populations, which are reactive to microbial antigens, are largely held in check by the intestinal immunoregulatory system. In some cases, expression of the Th cell signature (transcription factors and cytokines) of distinct lineages can overlap with each other or with the Treg transcription factor Foxp3, denoting either a common progenitor or the dynamic lineage transitions that can ensue in response to competing immune signals. In the absence of regulatory pathways, for example, in IL-10 deficiency, the numbers and frequencies increase gradually, coincident with the onset of chronic inflammation. The same is true during GI infection (see below), where the ability of invasive bacteria or viruses to enter a cell, or the physical interaction of the bacteria with the intestinal epithelium, culminates in induction and expansion of Th1 and Th17 cells, respectively, in the intestinal lamina propria.

Mucosal Antibody-Secreting Cells (ASC)

With increasing colonization comes an increased likelihood of epithelial breach, particularly by bacteria capable of penetrating the inner mucus layer and gain access to the intestinal epithelium. By intercalating dendrites between epithelial cells and into the lumen, DCs in the intestinal lamina propria are able to sample the luminal bacteria. Antigen-loaded DCs migrate to the mesenteric lymph nodes, where they interact with B and T cells to induce production of anti-commensal IgA. These IgA-producing plasma cells migrate to the intestinal lamina propria, where they secrete IgA dimers that migrate across the epithelium and into the lumen. These predominantly polyreactive IgA antibodies bind conserved moieties on the surface of diverse bacterial species, helping to sequester these organisms away from the epithelial layer. Commensal organisms, most notably *Bacteroides fragilis*, can harnesses this anti-commensal IgA to facilitate their own colonization of the GI tract. One study showed that the IgA-bound fraction of commensal bacteria supports the development of more severe disease in an animal model of ulcerative colitis.

In germ-free mice, the absence of the microbiota and resultant absence of IgA production is seemingly replaced by IgE class switching (Chapter 4) in mucosal lymphoid tissues with a corresponding increase in susceptibility to oral antigen-induced systemic anaphylaxis.[21] Therefore, the microbiota itself helps limit hyperreactivity to allergens and parasite challenge throughout an individual's life.

IMMUNE CONSEQUENCES OF EARLY MICROBIAL MANIPULATION

Because the microbiota is so essential to the early postnatal maturation of the immune system, the consequences of manipulation or insufficiency of the microbiota can impact host immunity throughout life. Microbial disruption, particularly during the neonatal–infancy period, is being associated with increased risk of autoimmunity and chronic inflammation later in life.

The increasing incidence of autoimmune disease in industrialized countries has been linked to use of antibiotics,[22] improved sanitation, and consumption of processed foods rich in fat and carbohydrates but negligibly low in fiber.[23] These practices limit microbiota diversity and deplete the bacteria that "educate" the developing immune system, leaving it prone to overreaction to subsequent challenges.

There is a growing body of experimental findings in support of this concept. For example, vancomycin treatment in young mice increases their susceptibility to asthma as well as food allergy, in large part as a result of depletion of clostridial strains known to promote colonic Treg induction and expansion. Antibiotic treatment in early life can limit the success of subsequently administered vaccines. Early life treatment with low-dose antibiotics can also result in increased susceptibility to obesity and corresponding alterations in immune gene expression in

the ileum. Individuals who had limited sanitary amenities during childhood or were raised around livestock are at reduced risk of developing IBD (Chapter 75) during adulthood. This latter phenomenon is commonly attributed to the acquisition of a diverse microbiota in these microbially enriched environments. However, such living conditions also pose an increased risk of parasitic helminth infection. This concept is supported by experimental data showing that treatment with *Trichuris muris* prevents development of experimental colitis in mice deficient for the IBD-related gene *Nod2*.

IMMUNE SYSTEM—MICROBIOTA CROSSTALK IN INTESTINAL INFLAMMATION

The immune system and the microbiota are in constant dialogue at steady state. Throughout an individual's life, the microbiota undergoes transient shifts in response to external influences. These include infections, medications (e.g., antibiotics), and dietary changes. The timing, magnitude, and targets of these perturbations can result in immune responses aimed at resetting this balance or limiting host collateral damage. Furthermore, feces of patients with certain extraintestinal chronic inflammatory diseases also display reduced microbial abundance and diversity relative to their healthy counterparts. This may reflect the role of the intestinal microbiota in the etiology of diseases, the impact of tissue-specific inflammation on the microbiota, or both.

Gastrointestinal Infection

The microbiota helps provide resistance to pathogenic invasion at mucosal sites. This can occur either indirectly by enhancing barrier defenses or directly by competing with harmful microbes. Nevertheless, several organisms still manage to breach these defenses. Such pathogen invasion induces production of proinflammatory cytokines by innate cells that can appropriately expand the magnitude of the immune response by signaling the differentiation of Th cells. The soluble products of these cells recruit other immune cells and collectively aid in pathogen eradication and sometimes in repairing any physical damage to the intestinal barrier. Unimpeded inflammatory responses can be destructive to host tissue, and thus, immunosuppressive mechanisms are induced that help to restore immune homeostasis.

In a healthy host, GI infections are largely self-limiting. However, the temporary disruption of microbial homeostasis (*dysbiosis*) can have lasting effects on host health. In the cases of *Salmonella enterica* serovar typhimurium (*S. typhimurium*) and *E. coli* infection in mice, the associated inflammation can drive production of substances that favor pathogen growth or enable the pathogen to outcompete resident microbes. The inflammatory response to the enteric pathogen *Citrobacter rodentium* promotes a restructuring of the microbiota that can predispose to chronic inflammation. Similarly, *Yersinia pseudotuberculosis* infection results in chronic inflammation and long-term defective lymphatic communication between the gut and mesenteric lymph nodes. These changes are supported by the dysbiotic microbiota.[24]

Infection of mice with the protozoan parasite *Toxoplasma gondii* (*T. gondii*) induces production of IL-12, which, in turn, promotes the differentiation of IFN-γ–secreting Th1 cells. IFN-γ is essential for pathogen control and promotes dysbiosis via targeted destruction of Paneth cells. The production of IL-10 by a subset of the Th1 cells helps overcome the inflammation, whereas IL-10–deficient mice succumb to the infection. However, even when the infection is controlled, the brief disruption in Treg homeostasis enables systemic dissemination of commensal bacteria and a temporary disruption in tolerance to the microbiota.[25]

Certain microbes also induce specific immune responses simply by their physical interaction with host cells. The ability of certain bacteria to bind directly to the intestinal epithelium is central to their ability to provoke a Th17 cell response that culminates in the accumulation of these cells in the underlying lamina propria. In mice, this is vividly displayed following infection with bacterial species, including the *Clostridium* spp., *Candidatus arthromitus* (segmented filamentous bacteria [SFB]), *Citrobacter rodentium*, and enterohemorrhagic *E. coli* (EHEC) O157:H7, which can breach the mucous barrier and adhere directly to intestinal epithelial cells. In the case of *C. rodentium* and EHEC, this results in a temporary effacing of the epithelial cell layer, local inflammation, diarrhea, and weight loss and in robust induction of Th17 cells. The same phenomenon can be reproduced if germ-free mice are colonized with *E. coli*–adherent bacteria derived from patients with ulcerative colitis.[26] Production of IL-17 as a result of the presence of SFB in the terminal ileum provides colonization resistance to *C. rodentium*. Induction of this immune pathway is stimulated by a microbial breach and helps restore epithelial integrity and limit further invasion. This provides an explanation for the therapeutic blockade of IL-17 being ineffective in the treatment of IBD—and even exacerbating the disease—even though it was successful in reversing symptoms of psoriasis (Chapter 64).[27]

Inflammatory Bowel Disease

IBD is a collective term that refers to a group of chronic relapsing–remitting inflammatory disorders that can occur anywhere along the GI tract. The two main forms of IBD, Crohn disease and ulcerative colitis, have similar clinical presentations but can differ in terms of histopathological features, affected sites, and risk of malignancy. Although disease onset can be impacted by a vast array of genetic, environmental, and immune factors, IBD is characterized by dysregulated immune responses to microbial antigens.

Although no single causative microbe or microbial cluster has been identified, there is strong correlative evidence in support of a dominant role for the microbiota in disease development and function. First, antibiotic therapy continues to be very effective in patients with inflammation in the lower bowel.[28] Second, in treatment-naïve patients with IBD, there is increased abundance of mucosa-associated pathobionts, including Enterobacteriaceae, Pasteurellaceae, Veillonellaceae, and Fusobacteriaceae. Conversely, a decrease in "beneficial" microbes, including Erysipelotrichales, Bacteroidales, and Clostridiales, has been observed. Reduced abundance of *Faecalibacterium prausnitzii*, a member of *Clostridium* cluster IV that induces an anti-inflammatory phenotype in immune cells, has also been associated with IBD and risk of relapse. Whether dysbiosis is a cause or consequence of disease is debatable. Dysbiosis can occur in disease-free relatives of patients with IBD,[29] which is consistent with a genetic influence on microbiota composition. However, several inflammatory mediators very effectively promote dysbiosis in experimental systems, supporting the notion that inflammation precedes microbial disruption.

Genome-wide association studies (GWAS; Chapter 18) have identified over 160 genetic loci that segregate with clinical disease. Several genes encode immune-related proteins involved

in detecting and/or responding directly to microbial products, inducing or amplifying the immune response to microbial challenge, or restoring and/or maintaining immune homeostasis with intestinal microbes.[30]

The first gene to be causally linked to IBD is *NOD2*, which encodes the nucleotide-binding oligomerization domain–containing protein 2, a microbial sensor that enables the immune system to recognize and respond to intracellular fragments of bacterial peptidoglycan.[31] Loss-of-function (LOF) mutations in *NOD2* (i.e., defective microbial recognition) predisposes to intestinal inflammation—but only in response to infection or injury. This is consistent with the "multiple hit" model of IBD pathogenesis. *NOD2* deficiency is associated with reduced numbers of intestinal goblet cells and, therefore, reduced mucus production, hyperactive IELs in the small intestine, and increased expansion of the commensal *Bacteroides vulgatus*.[32] All of these abnormalities can be prevented in mice infected with the helminth *Trichuris muris*. However, helminth infection as a therapy for human IBD has been largely unsuccessful.

Another immune-related IBD risk allele identified by GWAS is IL10, which encodes the immunosuppressive cytokine IL-10. In addition, rare LOF mutations in IL-10RA or IL-10RB, which encode the IL-10 receptor α and β chains, respectively, are found in a subset of patients with very-early-onset IBD. Mice with global or even CD4- or Foxp3-specific deletion of *IL-10* develop spontaneous colitis, which is dependent on the presence of the microbiota. The importance of the microbiota is highlighted by the varying kinetics and severity (ranging from complete protection to severe inflammation) of disease in different animal colonies or even in the same colony at different points in time. Disease is more consistent and severe and can even progress to colorectal cancer in mice colonized with *Helicobacter* spp. Thus, IL-10, which can be produced by multiple hematopoietic and nonhematopoietic cells, is critical for maintaining tolerance to the microbiota, particularly in the presence of species that can progressively disrupt the homeostasis of the microbiota.

Experimental studies have suggested a role for the nonprokaryotic inhabitants of the intestines in maintaining the status quo, thus limiting susceptibility to IBD. Fungi interact with the immune system via the receptor dectin-1. A single polymorphism in *CLEC7A*, which encodes dectin-1, has been linked to a severe form of ulcerative colitis. Accordingly, dectin-1 deficiency in mice precipitates increased susceptibility to chemical-induced colitis, and long-term antifungal treatment results in increased severity of acute and chronic experimental colitis.[33]

Members of the virome can either promote or prevent development of IBD-like symptoms via their interactions with intestinal bacteria. Murine norovirus induces inflammation in mice with disruption in the IBD susceptibility gene *Atg16L1*. Antibiotic administration reverses this viral-induced disease.[34] Conversely, antiviral pretreatment of otherwise healthy mice can exacerbate acute experimental colitis by disrupting the production of anti-inflammatory IFN-β.[35]

EXTRAINTESTINAL MANIFESTATIONS OF GUT MICROBIOTA–IMMUNE SYSTEM INTERACTION

Obesity develops when energy intake exceeds expenditure and culminates in deposition of excess adipose tissue. The associated chronic complications, collectively and clinically referred to as *metabolic syndrome*, include hyperglycemia, hypertriglyceridemia, dyslipidemia, and hypertension. Studies in mice and humans have revealed that there are alterations in the gut microbiota in obesity and associated metabolic diseases brought on by increased gut permeability, immune responsiveness, and aberrant bacterial translocation. For example, mice transplanted with human fecal microbiota from obese or lean cotwins displayed increased weight gain and adipose tissue relative to those transplanted with the microbiota of the lean cotwins.[36]

Obesity has been associated with an increased abundance of Firmicutes relative to Bacteroidetes, and differential responses of the lean and obese microbiota to the caloric content of the diet.[37] The altered composition can also lead to a reduction in microbial gene richness, which, in turn, affects the "inflammatory tone" of the microbiota and likely contributes the chronic low-grade inflammation characteristic of obesity. Lipopolysaccharides (LPSs), also known as endotoxins, derived from the outer cell membrane of gram-negative bacteria, are found at low concentrations in the circulation of healthy individuals, but are dramatically increased in obese individuals and even more so in those who develop type 2 diabetes. LPS infiltrates tissues, including the liver and adipose tissues, and the activation of macrophages via TLR4 initiates an innate immune response characterized by secretion of cytokines, including IL-6 and TNF-α. Obesity-related inflammation also contributes to defective insulin signaling or insulin resistance—a major player in the transition from metabolic syndrome to diseases, including type 2 diabetes, hepatic steatosis, and cardiovascular disease.[38]

Although the microbiota (usually fecal) is distinct in certain non-GI diseases, including ankylosing spondylitis (Chapter 58), multiple sclerosis (Chapter 66), and asthma (Chapter 43), whether these changes precede or follow disease development is not clear. One measure might be the effect of microbiota-induced immune mediators on disease pathology in extraintestinal tissues. For example, although gut microbes that can adhere to the epithelial cells induce robust expression of IL-17, their functions in the gut are largely protective. In contrast, microbiota-induced IL-17 can be proinflammatory in extraintestinal tissues. In germ-free mice colonized with SFB and subjected to experimental models of arthritis or multiple sclerosis, severe IL-17–dependent disease develops.[39] Elevated induction of RORγt and IL-17 in the central nervous system is also a hallmark of virus-induced maternal immune activation (MIA), which can lead to autism spectrum disorder (ASD)–like symptoms in a murine model.

CANCER AND THE MICROBIOTA

The tissue-specific immune inflammatory response that is necessary for pathogen eradication or to restore host–microbiota homeostasis can be deleterious to the host if allowed to reach chronicity. In addition to causing permanent tissue damage (scarring), such a chronic inflammatory response predisposes to tumor development and the production of neo-self-antigens that are now recognized as foreign to the host. The antitumor immunity that ensues possesses characteristics similar to those of the antipathogen response and ultimately has the same goal—eradication of a foreign body. Likewise, anti-inflammatory or regulatory mechanisms are deployed to limit the magnitude and duration of the inflammatory response in an effort to restore immune homeostasis (Fig. 22.5). However, continued tumor growth means that the "foreign" antigens have not been eradicated, and/or new ones are continually being generated.

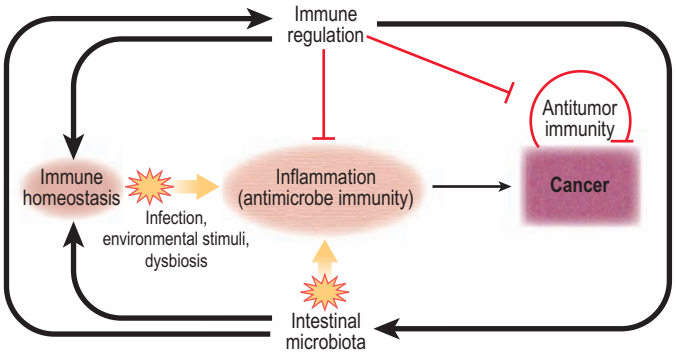


FIG. 22.5 Microbiota–Immunity–Cancer Triad. The reciprocal interaction of intestinal microbiota and the immune system induces default regulatory pathways that help maintain intestinal immune homeostasis. Factors that promote microbial dysbiosis, such as infection and other environmental insults, can disrupt immune homeostasis. This can result in proinflammatory responses targeting the causative agent that accompany immune regulatory responses that aim to reset immune balance. The perpetuation of the inflammatory response enhances the likelihood of progression to cancer. In response to the novel antigens generated in the tumor, an immune response is initiated with the goal of eradicating the tumor. As this response bears similar characteristics to an antipathogen response, it is subject to the same immune suppressive mechanisms.

In response to cancer, the host immune response needs to persist, rather than being opposed by the host's anti-inflammatory arsenal. Thus, similar nondiscriminatory immune regulatory mechanisms that are beneficial in curbing inflammation can impede antitumor immunity, thereby enabling tumor growth and eventual dissemination.[40]

There are several immune-related mechanisms whereby microbes directly impact the inflammation–cancer continuum. Some pathogens can promote an inflammatory milieu that encourages tumor development, whereas others can directly transform the eventual tumor-initiating cells. Oncogenic bacteria, including certain strains of *Enterococcus faecalis*, produce carcinogenic reactive oxygen species (ROS) capable of inducing DNA-damaging compounds or can induce production of carcinogenic compounds by activated immune cells. Inflammation-induced cell turnover directly increases the likelihood of introducing mutations in replicating DNA.

Microbiota–Immune System Interactions in Cancer Susceptibility and Development

Not surprisingly, animal models in which deletions of immune genes favor the emergence of a dysbiotic microbiota tend to develop spontaneous nonremitting intestinal inflammation. These include mice deficient in IL-10, Nod1, Nod2, Tbet, or Rag1. In several cases, microbiota reduction using antibiotics, rederivation in an axenic (germ-free) environment, or colonization with the microbiota derived from a wild-type animal is sufficient to significantly inhibit the development of inflammation, as well as the severity of cancer. Therefore, as a community, the microbiota has the potential to drive both gut inflammation and the eventual progression to cancer.

The gut microbial community has also been implicated in modulating carcinogenesis outside the intestines. Infection with *Helicobacter hepaticus* enhances mammary carcinoma in mice via mechanisms dependent on innate immune activation and TNF production. In addition, TLR5 signaling promotes progression of sarcomas in mice deficient for the tumor suppressor p53 and in which the protooncogene *Kras* has been activated. This disease phenotype can be abrogated by antibiotic-mediated reduction in commensal bacterial load.

In humans, approximately one of every six cancers develop downstream of a pathogenic infection. Notable pathogen–cancer axes include *Helicobacter pylori* and gastric carcinoma, human papillomavirus (HPV) and cervical cancer, and hepatitis B and C viruses and hepatocellular carcinoma.[41] As with intestinal inflammatory diseases, the composition of the fecal and mucosal microbiota in patients with colorectal cancer (CRC) is distinct from that of healthy individuals. The differences between mucosal bacterial populations present "on" and "off" the tumors in the same patient suggest a role for site-specific bacterial community structure in disease development and/or production.[42] More specifically, correlations have been made between CRC and the presence of colonic microbes, including enterotoxigenic *Bacteroides fragilis*, *Fusobacterium nucleatum, Enterococcus faecalis*, and *E. coli*. Animal studies support a role for all these strains in Wnt signaling and myeloid cell activation of nuclear factor-κB (NF-κB)–dependent inflammatory pathways.

The Microbiota in Cancer Immunotherapy

Because of its ability to educate the immune system to be constantly poised to respond to challenge, the microbiota is now also recognized as an important ally in the fight against cancer. Seminal discoveries have been made recently in delineating the pathways whereby the microbiota-primed immune system is essential for the success of common anticancer therapeutic agents and the gut microbiota is being targeted in efforts to enhance cancer immunotherapy.[1,43] Commensal bacteria support the potency of anti–IL-10R/CpG ODN treatment (a form of immunotherapy) as well as oxaliplatin (a form of platinum salt chemotherapy) in treating colon carcinoma by enhancing the production of myeloid-derived proinflammatory cytokines and ROS, respectively. The efficacy of the alkylating drug cyclophosphamide (CTX) is reduced in germ-free mice or mice treated with vancomycin, which depletes gram-positive bacteria, as a result of diminished antitumor adaptive immune responses. The microbiota is also critical for the antitumor effects of immune checkpoint inhibitors anti–programmed death ligand 1 (PD-L1) and anti–cytotoxic T lymphocyte antigen-4 (CTLA-4).[44]

In humans, anti–CTLA-4 treatment induces mucosal damage and microbiota modification, partly as a result of partial depletion of gut Tregs. The modified microbiota and the

consequent Th1-like immune response are critical for anti–CTLA-4 antitumor functions. In this study, the microbiota of patients treated with anti–CTLA-4 was enriched with bacterial species, including *Bacteroides thetaiotaomicron, Bacteroides fragilis*, and *Burkholderia cepacia*. Transplantation of germ-free mice with *B. fragilis* and *B. cepacia* partially rescued the efficacy of anti–CTLA-4 and prevented the mucosal toxicity of the antibody.

THE SKIN MICROBIOTA AND THE IMMUNE SYSTEM

Skin Microbes Maintain Barrier Integrity in the Steady State

With a surface area of approximately 1.8 m^2, skin is the largest organ in the body. Skin functions as a physical barrier against foreign agents (Chapter 23) and also participates in thermoregulation. Unlike the warm, nutrient-rich intestinal tract, skin is cool, desiccates, and is limited in available nutrients for microbial species. Thus, skin is populated by microbial communities capable of tolerating its diverse physiology (Fig. 22.6).[45] The total bacterial content of healthy human skin averages about 1 million/cm^2, for upward of 10^{10} total cells covering a single individual (or ≈1% of the number of bacterial cells per milliliter in the distal colon). The skin microbiota is generally acquired in concert with the colonization of other barrier surfaces in infancy. However, coincident with the individual's sexual maturation during adolescence, his or her skin bacterial communities undergo a major shift.[46]

Unlike the gut microbiota, skin commensals are dispensable for the maturation of the immune compartment of the tissue. However, their involvement in resistance to infection and collateral enhancement of skin disease risk are clear.[47] In response to pathogen challenge, skin-resident microbes mount a robust innate immune response characterized by some of the more primitive and evolutionarily conserved immune system messengers—antimicrobial peptides (AMPs), including cathelicidins and β defensins, components of the complement system (Chapter 40), and IL-1. Epithelial cells constitutively express some AMPs, which can target a vast array of skin pathogens, including bacteria, fungi, viruses, and parasites. Other AMPs are induced in a microbiota-specific manner and are expressed secondary to activation of the complement system.

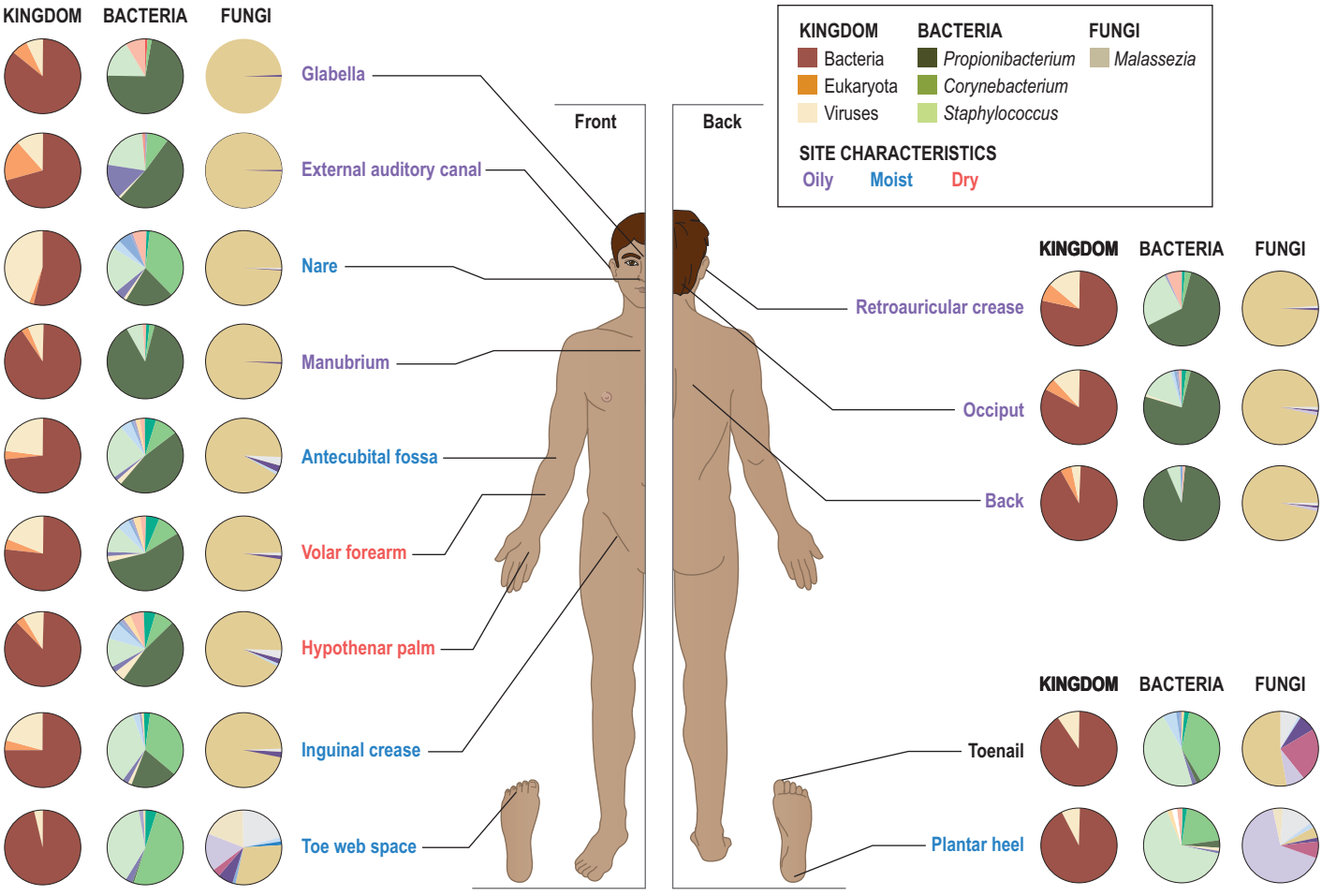


FIG. 22.6 Relative Abundance of Viral, Bacterial, and Fungal Components of The Microbial Community of Skin. Sites represent three microenvironments: sebaceous *(blue)*, dry *(red)*, and moist *(green)*. The toenail *(black)* is a site that does not fall under these major microenvironments and is treated separately. *Pie charts* represent consensus relative abundance of the different categories' kingdom, bacteria, and fungi. For bacteria and fungi, major taxa colors are identified in the legend. The minor taxa are colored to represent their relative proportion. (From Belkaid Y, Segre JA. Dialogue between skin microbiota and immunity. *Science*. 2014;346(6212):954–959, Fig. 1.)

In addition to being part of the innate response to microbial encroachment of skin, IL-1 stimulates robust adaptive immune responses that are essential for containment of both pathogens and commensals. In the skin of germ-free mice, there is reduced expression of IL-1, impaired induction of skin Th1, Th17, and IL-17-producing γδ T cells, and an elevated frequency of Foxp3$^+$ Tregs. Consequently, these mice also mount a suboptimal response to skin infection. Additionally, in germ-free mice, the reduction in adaptive immunity results in impaired containment of skin commensals and their dissemination to the draining lymph nodes. In contrast, in immune-replete mice, introduction of commensal organisms triggers the induction of CD8 T-cell responses, including the production of IL-17—which is further reinforced by DC-derived IL-1—and helps preserve the integrity of the barrier.

Skin Microbes in Chronic Inflammatory Disease

Skin dysbiosis and resultant dysregulated immune responses have been associated with inflammatory disorders, including acne vulgaris, psoriasis (Chapter 64), and atopic dermatitis (AD) (Chapter 48). Modulation of AMP production is critical, although the specific AMP, and whether it is upregulated or downregulated, can vary from one condition to the next.

In acne vulgaris, sebaceous hyperplasia and the release of lipids into the follicular lumen ultimately clogs the follicle and promotes a self-perpetuating outgrowth of *Propionibacterium acnes*. The follicular wall is breached, triggering an influx of inflammatory neutrophils and pustule formation. The expansion of *P. acnes*, as well as *Staphylococcus epidermidis*, leads to dysregulated immune responses, including elevated expression of AMPs and TLR expression by keratinocytes. These factors sustain the inflammatory response.[48]

Psoriatic lesions in human skin have been found to contain a reduced abundance of Actinobacteria, including the genus *Propionibacterium*, but an overrepresentation of Firmicutes. Overexpression of antimicrobial peptides, particularly IL-37 produced by stressed cells, is also detectable in diseased tissue.[49] IL-37 primes the innate immune response that subsequently induces development of Th1 and Th17 cells in the draining lymph nodes. The culpability of IL-17 expression in psoriasis has been confirmed by early clinical trials in which antibody-based targeting of IL-17 or the IL-17 receptor resulted in improvements in Psoriasis Area Severity Index (PASI) in at least 80% of patients after 12 weeks.[49]

Atopic dermatitis is characterized by dry skin (xerosis), which, along with the associated change in skin pH, favors the colonization and expansion of some microbes more than others. Colonization with *Staphylococcus aureus* has been linked to development of AD. *S. aureus* is detectable in the skin lesions of over 90% of patients with AD.[50] *S. aureus* infection induces both innate and adaptive immune activation. The expression of IL-37 transcript is markedly reduced in AD lesions, whereas other AMPs, including psoriasin, human β-defensin-2 and RNase 7, are overexpressed.

THE RESPIRATORY TRACT MICROBIOTA IN HEALTH AND DISEASE

Inhaled air contains bacteria, viruses, and fungi, and the respiratory tract serves as the main entry portal for these airborne microbes. However, for over a century, the healthy lung was considered a sterile environment, free of culturable and/or resident reproducing microbes. The detection of microbes in samples collected using instruments that had to traverse the mouth or nasal cavities was often assumed to have resulted from contamination with microbes from these sites. The emergence of culture-independent techniques for detection of microbial communities has precipitated an appreciation of the presence and diversity of microbial communities along the respiratory tract even in the absence of overt disease.[51] Although the airway and lung microbiota resembles the bacterial populations of the upper respiratory tract, there are differences that point to the existence of a specific lung microbiota.

The concept of a lung microbiota during health is relatively new, but the role of microbial agents in the pathogenesis of chronic lung diseases has been widely examined. In the healthy lung, the immune system eradicates potential pathogens and overcomes environmental disturbances that threaten to impair lung function. Direct impairment or intrinsic failure of the pulmonary defense mechanisms can lead to infection and/or chronic lung diseases, including asthma, chronic obstructive pulmonary disease (COPD), cystic fibrosis, and bronchiectasis.

Although microbial expansion in the lung is a characteristic of many lung diseases, it is not always clear whether the dysbiotic microbiota is the cause, the consequence, or both. Longitudinal studies have established a strong correlation between the early development of the respiratory tract microbiota and respiratory health both in infancy and adulthood. For example, the nasopharyngeal cavity is initially colonized with primarily *S. aureus* in the first few weeks of life. This is rapidly followed by an expansion of *Corynebacterium* and *Dolosigranulum*, with a subsequent transition to predominantly *Moraxella* spp. between 6 and 12 weeks of age. However, a rapid transition from *Staphylococcus* to *Moraxella*, essentially bypassing the *Corynebacterium* phase, occurred in infants who also experienced more frequent respiratory tract infections in the first year of life.

The Shaping of the Healthy Respiratory Microbiota

The bacterial density of inhaled air is approximately 10^4 to 10^6 cells/m^3. Thus, with each breath, mammals are constantly exposing their lungs to airborne bacteria. There are system-intrinsic mechanisms that serve to regulate entry and colonization of the lungs by microbes. The lung, like skin, and in stark contrast to the GI tract, is a very low nutrient resource. The thin mucus layer, which might seem to represent reduced barrier protection relative to the intestine, also means the absence of a valuable source of nutrients for certain mucophilic microbes. Furthermore, the combination of oxygen tension, temperature, regional pH, the diverse architecture of the respiratory tract, and the proximity of inflammatory cells all help keep the bacterial biomass of the lungs extremely low relative to that of the gut.

In a healthy individual, community composition at any given time is largely determined by the relative amount of microbial immigration and elimination. The major routes of immigration are microaspiration, air inhalation, and direct dispersion along the mucosal surface. Elimination is a continuous process mediated by the ciliated epithelial cells, by coughing, and by the actions of the pulmonary immune system. Colonization and growth contribute minimally to microbiota composition during homeostasis but are favored by changes in regional conditions that promote and/or sustain chronic inflammatory diseases of the lung.

Similar to the intestinal microbiota, the microbiota of the upper respiratory tract stabilizes in early childhood and can be

influenced by external factors, including breastfeeding, use of antibiotics, and social interactions as occurs in daycare and in multi-sibling households.[52] In a healthy lung, there is little spatial variation within the same individual. This lack of variation supports the theory that the composition of the lung microbiota is more heavily influenced by immigration and elimination than by local growth.

The most abundant bacterial phyla in the lung are Bacteroidetes and Firmicutes. At the genus level, *Prevotella, Veillonella*, and *Streptococcus* predominate. There are also prominent fungal communities, with dramatic differences in composition observed in different healthy cohorts. The oral cavity is home to several genera of fungi, including *Candida, Cladosporium*, and *Aspergillus;* but the fungal colonization of a healthy lung is unclear. Considering the constant exposure of the lung to oral and inhaled fungi, it is likely that the immune apparatus of the lung does, indeed, encounter fungal antigens in the steady state.

The involvement of the bacterial communities of other body sites, particularly the intestines, in lung immune homeostasis is still being actively explored. This phenomenon is commonly referred to as the "gut–lung axis" and involves the action of soluble mediators produced and/or induced by the gut microbiota that can enter the systemic circulation. By this rationale, this phenomenon is more a reflection of the intestinal microbiota on the organism as a whole and not representative of a unique relationship between these two organs. Nevertheless, there is compelling evidence that modulation of the gut microbiota, particularly during infancy, can have lifelong effects on lung immunity and susceptibility to chronic diseases.

MICROBES AS THERAPY

Certain infections or chronic inflammatory disorders are associated with a severely damaged microbiota. Transplantation of healthy donor microbiota has emerged as a successful therapeutic approach to repair and/or restore microbial communities. Fecal microbiota therapy (FMT), or fecal transplantation, has been safely and effectively utilized as a last resort to treat chronic *Clostridium difficile* infection. FMT presents the risk of adverse effects because the transplanted microbes, which are dormant in the donor, experience propathogenic conditions in the recipient. Because of the successes to date, empirical information regarding the long-term stability and resilience of one individual's microbiota transplanted into another should emerge over time. Other, more focused approaches that employ distinct microbes, groups of microbes, or microbial products known to have immune cell-specific effects to treat inflammatory diseases, such as IBD, are also being considered.

In Western, industrialized societies, there has been a consistent upward trend in the incidence of autoimmune diseases, and this has been linked to improved sanitation and diets low in fiber but rich in unhealthy fats and carbohydrates. These factors have contributed to dramatic shifts in microbial diversity away from that of our ancestral microbiota. Restoration of the richness of the microbiota using *probiotics*—live microorganisms that, when administered in adequate amounts, provide a measurable health benefit to the host—may be helpful in slowing and possibly reversing this trend. Long-term persistence of probiotic organisms within the complex microbiota of the host is essential if these organisms are to induce durable shifts in microbiota dynamics. Complementary *prebiotic* regimens that support the growth and survival of the introduced microbes without detrimental alterations in the distribution of functions of other species will likely be helpful in this regard.

SUMMARY

ON THE HORIZON

- Improved understanding of the structure and function of the microbiota in health and disease:
  - Microbial reconstruction in neonatal life to limit the immediate and long-term effects of cesarean delivery, preterm birth, and oral antibiotics on immune function.
  - Commensal microbes as a rich source of biological agents that can be used to regulate community development and resistance.
  - Improved approaches for the collection and study of organisms that inhabit the mucus layer of the gastrointestinal tract as opposed to just those expelled in the feces.
- Precision medicine and the microbiome:
  - Use of advanced genomic techniques for detailed assessment of microbiota functions in health and specific diseases.
  - Application of host genome and microbiome data to predict disease susceptibility and responses to therapy.
  - Targeting of the microbiota to enhance vaccine efficacy.
  - Assembly of limited "designer microbiotas" to initiate therapeutic reconstruction of dysbiotic patient microbiota.

The peaceful interaction between humans and the microbes that colonize them is essential for a healthy life. As such, the coexistence of the two entities occurs in an organized fashion, and we now know that this commences in utero and is subject to perturbations that can have dire consequences. The involvement of the microbiota as an instigator and/or sustainer of chronic inflammatory diseases and cancer continues to be explored. In the future, these studies may identify novel avenues for targeting certain conditions and restoring the immune balance. Our growing understanding of the essential roles of the microbiota in diverse physiological processes presents new opportunities to harness these capabilities across the clinical spectrum.

REFERENCES

1. Perez-Chanona E, Trinchieri G. The role of microbiota in cancer therapy. *Curr Opin Immunol.* 2016;39:75–81.
2. Kuss SK, Best GT, Etheredge CA, et al. Intestinal microbiota promote enteric virus replication and systemic pathogenesis. *Science.* 2011;334:249–252.
3. Gomez de Aguero M, Ganal-Vonarburg SC, Fuhrer T, et al. The maternal microbiota drives early postnatal innate immune development. *Science.* 2016;351:1296–1302.
4. Spits H, Di Santo JP. The expanding family of innate lymphoid cells: regulators and effectors of immunity and tissue remodeling. *Nat Immunol.* 2011;12:21–27.
5. Ismail AS, Severson KM, Vaishnava S, et al. Gammadelta intraepithelial lymphocytes are essential mediators of host–microbial homeostasis at the intestinal mucosal surface. *Proc Natl Acad Sci USA.* 2011;108:8743–8748.
6. Al-Nasiry S, Ambrosino E, Schlaepfer M, et al. The interplay between reproductive tract microbiota and immunological system in human reproduction. *Front Immunol.* 2020;11:378. https://doi.org/10.3389/fimmu.2020.00378.
7. Rogier EW, Frantz AL, Bruno ME, et al. Secretory antibodies in breast milk promote long-term intestinal homeostasis by regulating the gut microbiota and host gene expression. *Proc Natl Acad Sci USA.* 2014;111:3074–3079.
8. Chen K, Magri G, Grasset EK, Cerutti A. Rethinking mucosal antibody responses: IgM, IgG and IgD join IgA. *Nat Rev Immunol.* 2020;20:427–441. https://doi.org/10.1038/s41577-019-0261-1.

9. Niewiesk S. Maternal antibodies: clinical significance, mechanism of interference with immune responses, and possible vaccination strategies. *Front Immunol.* 2014;5:446.
10. Garofalo R, Chheda S, Mei F, et al. Interleukin-10 in human milk. *Pediatr Res.* 1995;37:444–449.
11. Bouskra D, Brezillon C, Berard M, et al. Lymphoid tissue genesis induced by commensals through NOD1 regulates intestinal homeostasis. *Nature.* 2008;456:507–510.
12. Klose CS, Artis D. Innate lymphoid cells as regulators of immunity, inflammation and tissue homeostasis. *Nat Immunol.* 2016;17:765–774.
13. Hepworth MR, Fung TC, Masur SH, et al. Immune tolerance. Group 3 innate lymphoid cells mediate intestinal selection of commensal bacteria-specific CD4(+) T cells. *Science.* 2015;348:1031–1035.
14. Olszak T, An D, Zeissig S, et al. Microbial exposure during early life has persistent effects on natural killer T cell function. *Science.* 2012;336:489–493.
15. Sarrabayrouse G, Bossard C, Chauvin JM, et al. CD4CD8αα lymphocytes, a novel human regulatory T cell subset induced by colonic bacteria and deficient in patients with inflammatory bowel disease. *PLoS Biol.* 2014;12:e1001833.
16. Campbell C, et al. Extrathymically generated regulatory T cells establish a niche for intestinal border-dwelling bacteria and affect physiologic metabolite balance. *Immunity*. 2018;48(1245–1257):e1249. https://doi.org/10.1016/j.immuni.2018.04.013.
17. Cebula A, Seweryn M, Rempala GA, et al. Thymus-derived regulatory T cells contribute to tolerance to commensal microbiota. *Nature.* 2013;497:258–262.
18. Round JL, Lee SM, Li J, et al. The Toll-like receptor 2 pathway establishes colonization by a commensal of the human microbiota. *Science.* 2011;332:974–977.
19. Kim YG, Sakamoto K, Seo SU, et al. Neonatal acquisition of *Clostridia* species protects against colonization by bacterial pathogens. *Science.* 2017;356:315–319. https://doi.org/10.1126/science.aag2029.
20. Kim CH, Park J, Kim M. Gut microbiota-derived short-chain fatty acids, T cells, and inflammation. *Immune Netw.* 2014;14:277–288.
21. Cahenzli J, Koller Y, Wyss M, et al. Intestinal microbial diversity during early-life colonization shapes long-term IgE levels. *Cell Host Microbe.* 2013;14:559–570.
22. Blaser MJ. Antibiotic use and its consequences for the normal microbiome. *Science.* 2016;352:544–545.
23. Deehan EC, Walter J. The fiber gap and the disappearing gut microbiome: implications for human nutrition. *Trends Endocrinol Metab.* 2016;27:239–242.
24. Fonseca DM, Hand TW, Han SJ, et al. Microbiota-dependent sequelae of acute infection compromise tissue-specific immunity. *Cell.* 2015;163:354–366.
25. Hand TW, Dos Santos LM, Bouladoux N, et al. Acute gastrointestinal infection induces long-lived microbiota-specific T cell responses. *Science.* 2012;337:1553–1556.
26. Atarashi K, Tanoue T, Ando M, et al. Th17 cell induction by adhesion of microbes to intestinal epithelial cells. *Cell.* 2015;163:367–380.
27. Hueber W, Sands B, Vandemeulebroecke M, et al. Inhibition of IL-17A by secukinumab is ineffective for Crohn's disease. *J Crohn's Colitis.* 2011;5:S7.
28. Feller M, Huwiler K, Schoepfer A, et al. Long-term antibiotic treatment for Crohn's disease: systematic review and meta-analysis of placebo-controlled trials. *Clin Infect Dis.* 2010;50:473–480.
29. Joossens M, Huys G, Cnockaert M, et al. Dysbiosis of the faecal microbiota in patients with Crohn's disease and their unaffected relatives. *Gut.* 2011;60:631–637.
30. Jostins L, Ripke S, Weersma RK, et al. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature.* 2012;491:119–124.
31. Philpott DJ, Sorbara MT, Robertson SJ, et al. NOD proteins: regulators of inflammation in health and disease. *Nat Rev Immunol.* 2014;14:9–23.
32. Ramanan D, Tang MS, Bowcutt R, et al. Bacterial sensor Nod2 prevents inflammation of the small intestine by restricting the expansion of the commensal *Bacteroides vulgatus*. *Immunity.* 2014;41:311–324.
33. Wheeler ML, Limon JJ, Bar AS, et al. Immunological consequences of intestinal fungal dysbiosis. *Cell Host Microbe.* 2016;19:865–873.
34. Cadwell K, Patel KK, Maloney NS, et al. Virus-plus-susceptibility gene interaction determines Crohn's disease gene Atg16L1 phenotypes in intestine. *Cell.* 2010;141:1135–1145.
35. Yang JY, Kim MS, Kim E, et al. Enteric viruses ameliorate gut inflammation via Toll-like receptor 3 and Toll-like receptor 7-mediated interferon-β production. *Immunity.* 2016;44:889–900.
36. Ridaura VK, Faith JJ, Rey FE, et al. Gut microbiota from twins discordant for obesity modulate metabolism in mice. *Science.* 2013;341:1241214.
37. Turnbaugh PJ, Ley RE, Mahowald MA, et al. An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature.* 2006;444:1027–1031.
38. Boulange CL, Neves AL, Chilloux J, et al. Impact of the gut microbiota on inflammation, obesity, and metabolic disease. *Genome Med.* 2016;8:42.
39. Lee YK, Menezes JS, Umesaki Y, et al. Proinflammatory T-cell responses to gut microbiota promote experimental autoimmune encephalomyelitis. *Proc Natl Acad Sci USA*. 2011;108(suppl. 1):4615–4622.
40. Goldszmid RS, Dzutsev A, Trinchieri G. Host immune response to infection and cancer: unexpected commonalities. *Cell Host Microbe.* 2014;15:295–305.
41. Parkin DM. The global health burden of infection-associated cancers in the year 2002. *Int J Cancer.* 2006;118:3030–3044.
42. Dejea C, Wick E, Sears CL. Bacterial oncogenesis in the colon. *Future Microbiol.* 2013;8:445–460.
43. Gopalakrishnan V, Helmink BA, Spencer CN, et al. The influence of the gut microbiome on cancer, immunity, and cancer immunotherapy. *Cancer Cell.* 2018;33:570–580. https://doi.org/10.1016/j.ccell.2018.03.015.
44. Viaud S, Saccheri F, Mignot G, et al. The intestinal microbiota modulates the anticancer immune effects of cyclophosphamide. *Science.* 2013;342:971–976.
45. Findley K, Oh J, Yang J, et al. Topographic diversity of fungal and bacterial communities in human skin. *Nature.* 2013;498:367–370.
46. Oh J, Conlan S, Polley EC, et al. Shifts in human skin and nares microbiota of healthy children and adults. *Genome Med.* 2012;4:77.
47. Flowers L, Grice EA. The skin microbiota: balancing risk and reward. *Cell Host Microbe.* 2020;28:190–200. https://doi.org/10.1016/j.chom.2020.06.017.
48. Nakamizo S, Egawa G, Honda T, et al. Commensal bacteria and cutaneous immunity. *Semin Immunopathol.* 2015;37:73–80.
49. Morizane S, Gallo RL. Antimicrobial peptides in the pathogenesis of psoriasis. *J Dermatol.* 2012;39:225–230.
50. Cho SH, Strickland I, Boguniewicz M, et al. Fibronectin and fibrinogen contribute to the enhanced binding of *Staphylococcus aureus* to atopic skin. *J Allergy Clin Immunol.* 2001;108:269–274.
51. Bassis CM, Erb-Downward JR, Dickson RP, et al. Analysis of the upper respiratory tract microbiotas as the source of the lung and gastric microbiotas in healthy individuals. *MBio.* 2015;6:e00037.
52. O'Dwyer DN, Dickson RP, Moore BB. The lung microbiome, immunity, and the pathogenesis of chronic lung disease. *J Immunol.* 2016;196:4839–4847.

Immunology of the Skin

Hui Xu, Nabiha Yusuf, and Craig A. Elmets

Skin, which is the largest organ in the body, is the major interface between an individual and his/her environment. It comprises 12% to 16% of the body's weight and functions chiefly to protect against environmental threats. It also plays an important homeostatic role in maintaining body temperature, water and electrolyte balance, and vitamin D production. It acts as a barrier against invading pathogens, exogenous chemicals, and the destructive effects of physical agents, such as the sun, wind, and heat. To perform these various functions, the epidermis of the skin consists of three specialized compartments—the stratum corneum, the granular layer, and the basal layer—that are layered on top of each other (Fig. 23.1). In turn, the epidermis rests on the dermis, which then rests on the hypodermis. Adnexal tissues, which include hair, nails, sebaceous glands, and eccrine and apocrine sweat glands, are embedded within these tissues to provide specialized protective functions.

KEY CONCEPTS

Primary Functions of Skin

- Body's largest organ (12% to 16% of body weight with a surface area of ≈2 m^2).
- Principal organ for homeostatic thermoregulation and regulation of fluid retention and evaporation.
- Specialized sensory cells create a skin-wide tactile interface with the environment (e.g., fingertips with >2500 sensory cells/cm^2).
- Photosynthetic surface that uses ultraviolet radiation (UVR) to produce cholecalciferol (vitamin D_3).
- Physical barrier to physical stresses, including shear forces, extreme temperatures, wind, water, and solar radiation. Tight junctions between granular layer keratinocytes block entry of small molecules and microbes.
- By virtue of detoxification enzymes, pharmacological barrier to chemicals and other carcinogens (e.g., UVR and ionizing radiation).
- Regenerative organ that continuously replaces older, damaged skin cells, which are eliminated and shed from skin as anucleated squames.
- Immunological barrier using innate and adaptive immune mechanisms against exogenous antigens, microbial pathogens, and endogenous neoplastic cells.

The outermost layer of skin is the epidermis. Blood and lymphatic vessels are absent. However, oxygen and nutrients diffuse to epidermal cells from the microvasculature that is housed in the dermis that lies beneath the epidermis. Approximately 95% of cells within the epidermis are keratinocytes. These form a self-renewing stratified squamous epithelium, which differentiates from bottom to top from a regenerating basal layer containing cuboidal cells to the outer most superficial layer, called the *stratum corneum*, composed of flat, anucleate, compact scales. Keratin intermediate filaments are the major proteins produced by keratinocytes. The type of heterodimers that are made by keratins during differentiation affects the cytoskeletal structure and cell morphology.

Keratinocytes bind to one another through specialized adherence junctions called *desmosomes*. Keratin filaments attach to desmosomal proteins to provide structural, tensile strength. As differentiation proceeds, the composition of the keratin and desmosomal proteins changes. For example, basal layer keratinocytes synthesize keratins 5 and 14, and the desmosomal protein desmoglein 3 is more abundant than desmoglein 1. In the stratum corneum, keratins 1 and 10 and desmoglein 1 are all highly expressed, but keratins 5 and 14 and desmoglein 3 are absent. Finally, the compact tough outer scales (squames) of the stratum corneum are cross-linked with keratin filaments. The differences in protein localization between the epidermal layers have consequences for immunologically mediated skin diseases. Pemphigus vulgaris, a disease in which autoantibodies are formed against desmogleins 1 and 3 (Chapter 63), presents with blisters that originate in the suprabasal layer of the epidermis. Conversely, the closely related disease pemphigus foliaceus, in which autoantibodies recognize only desmoglein 1, results in blisters located exclusively in the upper epidermis.[1]

The remaining cells located in the epidermis are pigment-producing melanocytes, neuroendocrine Merkel cells (important for mechanoreception), and Langerhans cells (LCs) (specialized epidermal antigen-presenting cells [APCs]). Melanocytes are derived from the neural crest, and progenitors seed the epidermis early in development. They synthesize melanin pigment in organelles called *melanosomes*. Melanosomes mature into the melanin-filled granules, which are then transferred and internalized by keratinocytes. Melanin absorbs the damaging effects of ultraviolet radiation (UVR) and protects skin keratinocytes. When stimulated by cytokines, melanocytes express several immunologically relevant proteins, including intracellular adhesion molecule 1 (ICAM-1; CD54), costimulatory receptor CD40, and major histocompatibility complex (MHC) class I and II molecules (Chapter 5).[2]

Malignant melanomas arise from melanocytes. These tumors evade the host immune response, at least in part, through expression of programmed death ligand 1 (PD-L1) (Chapter 17). PD-L1 is an immune checkpoint molecule found on tumor cells and myeloid derived suppressor cells (Chapter 80). The interaction of PD-L1 with PD-1 receptors on T cells inhibits their antitumor activity. The monoclonal antibodies (mAbs),

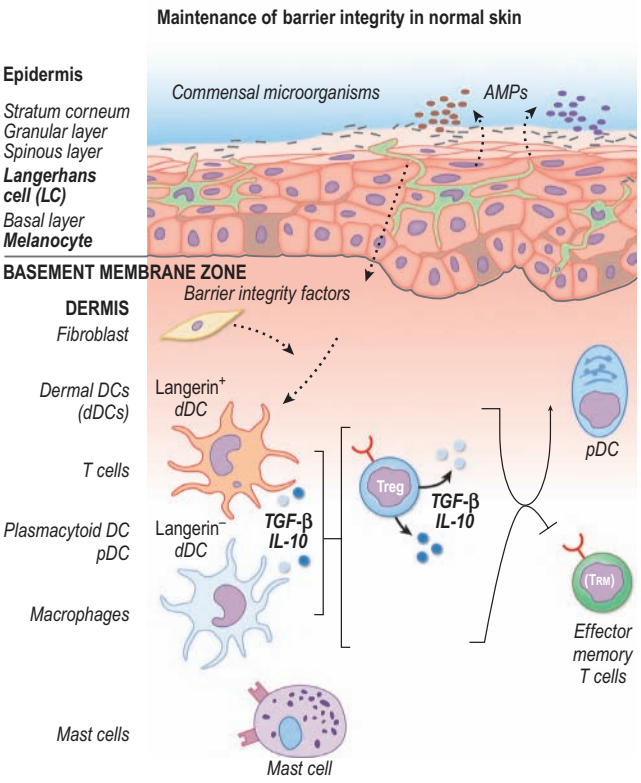


FIG. 23.1 Histological Features of Normal Skin. The epidermis is made up of keratinocytes organized into stratified layers. The most superficial layer is the stratum corneum *(light pink)*. Granular layer cells also secrete antimicrobial peptides *(AMPs)*. The basal layer *(tan)* contains the proliferative cells that give rise to all the differentiated suprabasal keratinocytes. Keratinocytes are connected to each other through tight junctions *(dark grey)*. Basal cells are bound to the basement membrane *(black line)* and share that scaffold with interspersed melanocytes *(brown)*. Langerhans cells *(LCs) (green)* are epidermal antigen-presenting cells (APCs) that reside in the suprabasal layer of the epidermis. The dermis contains a number of dendritic cell (DC) subsets, including langerin$^+$ *(orange)* and langerin$^-$ *(light blue)* dermal DCs (dDCs). Plasmacytoid DCs *(pDCs)* are found in the dermis and secrete type I interferons *(aqua)*. Resident memory T cells (T_{RM}) *(green)* represent previously primed T cells in many lineages (Th1, CTL, Th17, Tc17, Treg) poised to respond when their specific antigens are presented locally. They are quiescent in uninfected skin. Mast cells *(purple)* reside near endothelial venules (not shown). *TGF*, Transforming growth factor. (Drawing by Laura Timares, PhD.)

pembrolizumab and nivolumab, that block the PD-1 interaction with PD-L1 can restore antitumor responses. Tumor immunotherapy with these antibodies results in objective responses and prolonged survival in patients with metastatic melanoma. Ipilimumab, which blocks cytotoxic T lymphocyte antigen-4 (CTLA-4), another immune checkpoint inhibitor, is also effective against melanoma.

The dermis lies beneath the epidermis and serves as a connective tissue layer that provides elasticity and tensile strength to skin. It is filled with a matrix of collagen bundles, elastic fibers, glycoproteins, proteoglycans, and glycosaminoglycans, all of which are produced by dermal fibroblasts. The production of collagen is a dynamic process that involves continual remodeling. Fibroblasts must balance ongoing synthesis of matrix components with the production of matrix metalloproteinases (MMPs) that degrade the matrix. Scleroderma and morphea (also known as localized scleroderma) (Chapter 56) are inflammatory disorders that lead to the overproduction of dermal collagen, thus hardening the skin.

Other cell types normally found in the dermis are mast cells and diverse sets of tissue macrophages (Chapter 44) and dendritic cells (DCs) (Chapter 6).[3] The dermis also supports the microvasculature of skin. Endothelial cells of dermal arterioles, capillaries, and post capillary venules behave differently from those in larger vessels. They express E- and P-selectins, to which circulating skin-homing T cells bind through their homing receptors cutaneous lymphocyte-associated antigen (CLA) and P-selectin glycoprotein ligand (PSGL), respectively. Both homeostatic T-cell trafficking and recruitment of inflammatory leukocytes into skin from the circulation require expression of E- and P-selectins (Chapter 16). The egress of leukocytes from blood occurs primarily through post capillary venules. Egress is regulated, in part, by endothelial responses to cutaneous cytokines and chemokines. New vessels can also develop in skin. They can worsen immunologically mediated dermatological diseases and promote tumor development.

The dermis is separated from the epidermis by the basement membrane. This structure creates a scaffold to which basal keratinocytes and melanocytes are attached. It acts as a selective diffusion barrier that permits passage of necessary small-molecule nutrients while retarding entry of macromolecules and inflammatory cells. In patients with systemic lupus erythematosus (SLE) (Chapter 52), antigen–antibody complexes accumulate at the basement membrane. In addition, skin-specific proteins present in the basement membrane are targets of such autoimmune blistering diseases as bullous pemphigoid and epidermolysis bullosa acquisita (Chapter 63).

The hypodermis lies beneath the dermis and attaches skin to underlying muscle. It is composed mainly of adipose tissue, which cushions skin impacts. Adipose cells produce leptin, which is implicated in a variety of inflammatory skin diseases, including psoriasis (Chapter 64).

The skin protects individuals from exogenous threats through three major barriers: *(i)* physical, *(ii)* pharmacological or detoxifying, and *(iii)* immunological. *The physical barrier* includes the hair and stratum corneum. The stratum corneum provides an impermeable hydrophobic cover of protein-filled squames that hinders the entry of invading microorganisms or toxic chemicals into the body. It also reflects and scatters ultraviolet radiation (UVR), blocking it from reaching the deeper, regenerating layers of skin. *The pharmacological barrier* includes detoxifying and repair enzymes synthesized by epidermal cells. Some metabolize chemicals and xenobiotics, whereas others repair DNA damage resulting from UVR or environmental mutagens.

The immunological barrier includes cells and molecules unique to skin that are historically referred to as "skin-associated lymphoid tissue." They include elements of both innate and acquired immunity. Locally they enhance host defenses that protect affected skin from antigens and pathogens that are present in that site. Systemically, they alert the rest of the body to encourage and expand host defenses. Although these protective measures work exceedingly well, they do not work perfectly. Disturbances in these protective mechanisms can result in increased infections and malignancies when deficient or in immunologically mediated skin diseases when excessive (Table 23.1).

KEY CONCEPTS

Immunological Defenses of Skin

- Innate defenses provide the first line of protection against environmental antigens and pathogens.
- In response to pathogen-associated molecular patterns (PAMPs) and danger-associated molecular patterns (DAMPs), granular layer keratinocytes produce cutaneous microbicidal peptides and proinflammatory mediators.
- At initial skin breach, neighboring keratinocytes and dendritic cells (DCs) initiate and perpetuate a proinflammatory cascade of events.
- Langerhans cells internalize microbes, produce cytokines, migrate from skin to regional lymph nodes, and mature into potent antigen-presenting cells (APCs).
- In lymph nodes, DCs deliver antigens from affected skin to lymph node (LN) resident DCs to initiate and amplify skin-specific immunity.
- Keratinocyte-derived proinflammatory mediators and chemokines act on post endothelial venules to promote extravasation of neutrophils, monocytes/macrophages, natural killer (NK) cells, and memory T cells from the circulation into the dermis.
- In order to clear affected epidermis of invading pathogens and cellular debris, immigrant inflammatory leukocytes attach to activated keratinocytes through interactions with CD54 (intercellular adhesion molecule 1 [ICAM-1]).

MOLECULES PRIMARILY ASSOCIATED WITH INNATE IMMUNITY AND SKIN

Pattern Recognition Receptors

By rapidly sensing and responding to microbial pathogens and environmental toxins, epidermal keratinocytes act as innate first responders (Chapter 3). Keratinocyte pattern recognition receptors (PRRs) recognize highly conserved sequences in macromolecular components of microbial pathogens and host-derived danger indicators. Collectively, PRRs are promising therapeutic targets for cutaneous immune disorders and skin cancer.

Identification of pathogen-associated molecular patterns (PAMPs) and danger-associated molecular patterns (DAMPs) is principally undertaken by an array of PPRs, which include Toll-like receptors (TLRs), NOD-like receptors (NLRs), RIG-I-like receptors (RLRs), and C-type lectin receptors (CLRs).[4] Human keratinocytes express TLRs 1 to 6, and 9. Human melanocytes constitutively express TLRs 2 to 5, and 9.[5]

Human and murine DC subsets show differential TLR expression profiles. For example, conventional DCs (cDCs) express combinations of TLRs 1 to 5 and 8, whereas TLR expression is restricted to virus-detecting endosomal TLRs 7 and 9 in plasmacytoid DCs (pDCs). The latter is consistent with their primary function as major interferon-α (IFN-α) producers.

TLR ligands are derived from conserved microbial products (e.g., lipopolysaccharides, lipoteichoic acid, peptidoglycans, mannans, nucleic acids from viral and bacterial pathogens, and endogenous DNA from stressed host cells). DAMPs are recognized by NLRs. In the clinic, the TLR7 agonist imiquimod is used as an effective topical agent for eliminating human papillomavirus (HPV)–induced genital warts and nonmelanoma skin tumors.

The nucleotide-binding oligomerization domain-like receptors (NLRs) have over 20 members in humans, reflecting the broad spectrum of PAMPs and DAMPs they recognize. Mutations and gene single nucleotide polymorphisms are associated with inflammatory diseases in the skin. Also present are cytosolic NLRs. Although often with partial overlap, TLR and NLR

TABLE 23.1 Immunodermatological Disorders That Affect the Skin

- Papulosquamous disorders
  - Psoriasis
  - Lichen planus
  - Cutaneous graft-versus-host disease
  - Acute, subacute, and discoid lupus erythematosus
- Eczematous disorders
  - Atopic dermatitis
  - Allergic contact dermatitis
- Urticarial disorders
  - Urticaria and angioedema
  - Erythema multiforme
  - Stevens-Johnson syndrome
  - Toxic epidermal necrolysis
  - Autoinflammatory diseases (e.g., cryopyrin-associated periodic syndromes, Muckle-Wells syndrome, familial cold urticarial, neonatal-onset multisystem inflammatory disease [NOMID], deficiency of the interleukin-1–receptor antagonist [DIRA], tumor necrosis factor receptor–associated periodic syndrome [TRAPS])
- Purpuric disorders
  - Leukocytoclastic vasculitis
  - Medium vessel vasculitides (polyarteritis nodosa, granulomatosis with polyangiitis, eosinophilic granulomatosis with polyangiitis (EGPA))
- Vesiculobullous diseases
  - Pemphigus
  - Bullous pemphigoid
  - Paraneoplastic pemphigus
  - Epidermolysis bullosa acquisita
  - Dermatitis herpetiformis
  - Linear immunoglobulin A (IgA) bullous dermatoses
  - Pemphigus gestationis
- Pigmentary disorders
  - Vitiligo
- Hair disorders
  - Alopecia areata
- Autoimmune disorders
  - Acute, subacute, and discoid lupus erythematous
  - Dermatomyositis
  - Mixed connective tissue disease
  - Scleroderma and morphea
- Photodermatoses
  - Polymorphous light eruption
  - Solar urticaria
  - Chronic actinic dermatitis
  - Photoallergic contact dermatitis
- Allergic drug eruptions
- Disorders of subcutaneous tissues
  - Erythema nodosum
  - Erythema induratum
- Immunodeficiencies
  - Ataxia telangiectasia
  - Chronic mucocutaneous candidiasis
  - Chronic granulomatous disease
  - Hyper-immunoglobulin-E (IgE) syndrome (HIES)
  - Leukocyte adhesion molecule deficiency
  - Severe combined immunodeficiency
  - Warts–hypogammaglobulinemia–infections–myelokathexis syndrome (WHIM syndrome)
  - Wiskott-Aldrich syndrome
  - Organ transplant recipients on immunosuppressive medications

expression profiles differ among epidermal and leukocyte cell subsets.

NLRs can be divided into four subfamilies: (i) NLRA, (ii) NLRB, (iii) NLRC, and (iv) NLRC. NLRC is expressed in many tissues, including keratinocytes in the skin. Primary human keratinocytes also express mRNA transcripts for 11 of 14 NLRP family genes.[4] NLR units that recruit and activate the inflammatory protease caspase-1 are known as inflammasomes. NLRs act independently, but in synergy with TLR signals, to trigger fast-acting potent innate immune responses against microbes. They can either promote nuclear factor kappa B (NF-κB) activation or form inflammasomes that generate interleukin-1 (IL-1) and IL-18 family proinflammatory mediators. The set of TLRs and NLRs triggered determines the type and magnitude of the immunomodulatory or inflammatory response that will result.

CLRs are involved in recognition of types of viruses. Several CLRs, including DC-specific ICAM-3 grabbing non-integrin (DC-SIGN), Langerin, mannose receptor, C-type lectin domain family 4 member A (CLEC4A), C-type lectin domain family 5 member A (CLEC5A), and C-type lectin domain family 9 member A (CLEC9A), bind to viruses that invade the skin. Dectin-1 is a membrane-associated glycoprotein that is found on keratinocytes, DCs, and monocytes. It recognizes β-glucan and is involved in the host response to fungal pathogens. Its activation results in production of IL-1β, IL-6, and IL-23 (Chapter 14), which promotes development of Th17 cells, which are important antifungal effector cells (Chapter 11). Patients lacking dectin-1 exhibit Th17-cell deficiency, recurrent vulvovaginal *Candida albicans* infections, and onychomycosis (Chapter 29). Mutations in Card9, a mediator of dectin-1 signal transduction, also lead to Th17 deficiency and chronic mucocutaneous candidiasis.[6]

RLRs, (RIG-I, MDA5, and LGP2) sense viral dsRNA during cutaneous viral infections and DNA released from damaged host cells. RLRs have an important role in type I IFN production.[7] RIG-I signaling can drive IL-23 production and induce psoriasis-like skin disease in mice.[8]

Antimicrobial Peptides

Skin is a production site for antimicrobial peptides, which are important components of the innate immune response.[9] Over 20 antimicrobial peptides have been identified. They function at the earliest stages of infection by microorganisms that breach the stratum corneum. These antimicrobial peptides disrupt bacterial and fungal membranes, and viral envelopes. They have broad effects on innate and adaptive immune responses. The two best characterized skin antimicrobial peptides are the cathelicidins and β-defensins. They are synthesized by keratinocytes, cells of the epidermal sebaceous and eccrine glands, and dermal mast cells. Cathelicidins are secreted as precursor proteins (e.g., hCAP18), which are then processed to an active form (e.g., LL-37). They are detected at low levels in unperturbed skin and are strongly increased following infection or disruption of the epidermal barrier.

Cathelicidins interact with a variety of cell-surface receptors (e.g., TLR-like and endothelial growth factor [EGF] receptors). In skin, they enhance leukocyte migration, stimulate the secretion of cytokines (Chapter 14) and chemokines (Chapter 15), and promote angiogenesis. Vitamin D_3, whose synthesis is initiated in skin by UVR, plays an important role in regulating cathelicidin production through epigenetic mechanisms.[10]

Abnormalities in antimicrobial peptides have been implicated in a variety of immunological and non-immunological skin diseases. Deficiencies in antimicrobial peptides have been found in patients with atopic dermatitis (Chapter 48), which may explain why they are at risk for viral (herpes simplex, vaccinia, HPV-induced warts, poxvirus-induced molluscum contagiosum) and bacterial *(Staphylococcus aureus)* skin infections.

Rosacea manifests as an acneiform eruption with erythema, telangiectasias, and flushing of the face. Excessive amounts of cathelicidins are associated with the characteristic inflammation and angiogenesis in that disease.[11]

In psoriasis (Chapter 64), cathelicidins are postulated to combine with self-DNA from damaged keratinocytes. This stimulates type I IFN production by pDCs, which is a key to the development of pathogenic Th17 cells.[12]

Non-immune mechanisms that cause physical disruption of the skin barrier (e.g., shear forces, chemical, thermal, or UV damage) stimulate "sterile" inflammatory responses that produce similar proinflammatory mediators and antimicrobial peptides to protect vulnerable sites during wound healing.

Cytokines and Chemokines in the Skin

Skin is a rich source of cytokines (Chapter 14) and chemokines (Chapter 15). They are involved in both innate and adaptive immune responses. They influence the magnitude and type of cutaneous immune response generated (Table 23.2).[13] Although many chemokines and cytokines produced in skin are identical to those secreted by non-skin cells, in skin they can have unique effects.

Cutaneous cytokines and chemokines are also important contributors to systemic responses to injury, wound healing, carcinogenesis, and pigmentation. For example, IL-1 can remodel the dermis by inducing matrix metalloproteinases (MMP) production and augmenting collagen production. Production of these mediators in skin involves several different cell types and environmental stimuli (see Table 23.2). Polymorphisms in cytokine genes and their receptors have been associated with a variety of inflammatory diseases. For example, polymorphisms in the tumor necrosis factor (TNF) signaling pathway and in the IL-23 receptor are risk factors for psoriasis.

Two cytokines produced by keratinocytes and epidermal LCs are particularly important for activation of T cells. IL-12 facilitates the generation of Th1 cells; IL-23 has a stimulatory role for the generation of Th17 cells (Chapter 11). Because IL-23 promotes Th17-cell development, antibodies that block the p19 subunit of IL-23 (guselkumab, risankizumab, tildrakizumab, bimekizumab) or the p23 subunit common to both IL-12 and IL-23 (ustekinumab) are effective therapeutic agents for psoriasis.

TABLE 23.2 Cytokines Produced by the Skin

- Keratinocytes
  - Proinflammatory (IL-1, IL-6, IL-33, TNF-α)
  - Immunosuppressive and antiinflammatory (IL-10, TGF-α, TGF-β, IL-1 receptor antagonist [IL-1RA])
  - Colony-stimulating factors (GM-CSF, G-CSF, and M-CSF)
  - Immunomodulatory (IL-7, IL-12, IL-15, IL-18, IL-19, IL-20, IL-23)
  - Chemokines (CXCL8/IL-8, CCL2/MCP-1, CCL20/MIP3α, CCL5/RANTES).
- Langerhans cells
  - Proinflammatory (IL-1α, IL-1β)
  - T-cell maturation (IL-12, IL-23)
  - Chemokines (CCL3/MIP1α, CCL4/MIP1αβ, CCL5/RANTES)
- Melanocytes
  - Proinflammatory (IL-1β, IL-6, TNF-α)
  - Type I interferons (IFN-α, IFN-β)
  - Chemokines (CXCL8/IL-8, CCL2, CCL3, and CCL5)

Chemokines facilitate leukocyte migration and thus skin inflammation and antigen-specific responses (Table 23.3).[13] CXCL12 facilitates DC migration out of skin. CCL17 and CCL27, produced by keratinocytes and dermal cells, facilitate transmigration of circulating skin-homing T cells (expressing the chemokine receptors CCR4 and CCR10). Human skin–resident T cells are retained through the interaction of their chemokine receptor CCR8 and its ligand CCL1 bound to the dermal matrix.[13] Secondary lymphoid chemokine (SLC, CCL21) attracts $CCR7^+$ DCs and T cells from skin to lymph nodes. During inflammation, CXCL1 and CXCL8/IL-8 expedite infiltration of granulocytes and monocytes into skin. Cutaneous T-cell lymphoma (Chapter 78) is a disease in which a malignant population of skin-homing T cells accumulate in skin. The malignant T cells express CCR4 chemokine receptor. Humanized mAbs directed against the CCR4 receptor on T cells (mogamulizumab) are effective at treating this condition.[14]

CYTOKINE SIGNALING

The Janus kinase/signal transducer and activator of transcription (JAK/STAT) pathway is important for downstream signaling of several inflammatory cytokines and multiple growth factors. When cytokines bind to their receptors, it causes phosphorylation of JAK and STAT proteins. The dimerized STATs translocate to the nucleus, and regulate transcription of inflammatory cytokine genes.[15] Numerous inflammatory skin diseases are triggered by soluble inflammatory mediators, which depend on JAK/STAT signaling. Therapeutic intervention can thus simultaneously block the functions of several cytokines. Currently, JAK inhibitors (JAKi) are approved in oral formulations (tofacitinib and baricitinib). They are being tested as therapeutic agents for alopecia areata, atopic dermatitis, dermatomyositis, cutaneous graft-versus-host-disease, hidradenitis suppurativa, lichen planus, lupus erythematosus, psoriasis, and vitiligo.[15] Because of adverse effects associated with JAKi, nanoformulations with greater cutaneous absorption are in development for topical treatment.

Cytokine-Related Autoinflammatory Diseases

Mutations in cytokines are a common feature of autoinflammatory disorders.[16] Examples are familial Mediterranean fever, cryopyrin-associated periodic syndromes, deficiency of the IL-1 receptor antagonist (DIRA), and TNF receptor associated

TABLE 23.3 T-Cell Subsets and Trafficking in Skin

T-Cell Subset	Homing-R[a]	Chemokine-R	From	Homing to	Ligands for Homing and Chemokine Receptors
Naïve T cells	L-selectin (CD62L) Lymph node (LN) homing receptor	$CCR7^+$	Blood	LN	L-selectin ligand: GlyCAM-1 integrin on LN HEV[b]
Central memory T cells (T_{CM}) $CD45RA^+$	L-selectin	$CCR7^+$	LN	LN and the circulation	CCR7 ligand: LN-derived chemokine CCL21 (also called 6Ckine), secondary lymphoid-tissue chemokine (SLC)
Resident memory T cells (T_{RM}) $CD45RO^+$	Cutaneous lymphocyte-associated antigen (CLA^+) (P-selectin glycoprotein ligand-1 [PSGL-1]) Skin-homing receptor	$CCR4^+$	LN	Blood/dermis	CLA ligand: E-selectin on activated post capillary venules CCR4 ligand: skin-derived inflammatory chemokines: MCP-1, MIP-1, RANTES, TARC, CCL22
T_{RM} Subsets Resident in Dermis	**Homing-R**	**Chemokine-R**	**Function**	**Cytokines Produced**	**Notes**
T regulatory (Treg)	CLA^+	$CCR5^+$	Regulatory; resolution phase; inhibit autoreactivity	Interleukin (IL)-10, transforming growth factor (TGF)-β	Tregs are 5%–10% of CLA^+ dermal T cells Treg two-way trafficking–migration from dermis to LN, then recirculate back to dermis CCR5 binds RANTES, MIP-1
T-helper (Th)1, Tc1	CLA^+	$CCR4^+$	Type I antiviral	Interferon (IFN)-γ, tumor necrosis factor (TNF)-α, IL-2	Promotes cellular immunity Development of cytotoxic T cells (CTLs)
Th2, Tc2	CLA^+	$CCR4^+$	Type II parasites cleared	IL-4, IL-5, IL-10, and IL-13	Promotes humoral immunity Recruitment of eosinophils
Th17, Tc17	CLA^+	$CCR4^+$	Inflammatory; antifungal; antibacterial		Activation is dependent on IL-23 derived from keratinocytes/Langerhans cells (LCs)/dendritic cells (DCs)

T-cell distribution density in skin ≈1 million T cells/cm^2.
[a]Homing-R—an adhesion molecule on leukocytes that recognizes and binds site-specific adhesion molecules expressed on high endothelial venules in the lymph node (LN) and activated post capillary venules in the dermis and other organs.
[b]Selectin is the lymph node homing receptor, whereas CLA is the skin-homing receptor.
HEV, High endothelial venules; these are post capillary venous swellings in lymph nodes and most secondary lymphoid organs.

periodic syndrome (TRAPS). These diseases cause fever and arthralgias. Cutaneous manifestations are also particularly common and include atypical urticarial and vasculitic rashes, and other unusual expressions of cutaneous inflammation. For example, DIRA, associated with a mutation in the gene that encodes the IL-1–receptor antagonist, causes severe pustulosis, ichthyosiform lesions, psoriasis-like changes, and nail abnormalities.[16]

KEY CONCEPTS

- Skin-resident and recirculating immunocompetent cells and immune mediators defend against microorganisms, noxious exogenous chemicals, and somatically mutated cells.
- Excessive or dysregulated immune system activity can give rise to autoimmune skin diseases (e.g., production of desmogleins-1 and -3 binding antibodies in pemphigus).
- Underperforming immune system activity can lead to cutaneous immunodeficiency (e.g., skin cancers in patients receiving immunosuppressive medications).
- Skin-specific immune system components provide targets for new therapies (e.g., treatment with neutralizing IL-17 binding antibodies in psoriasis).

CELLULAR COMPONENTS OF THE CUTANEOUS ADAPTIVE IMMUNE RESPONSE

Innate immunity forms the first line of defense against microbial threats in an antigen-nonspecific manner. The host's adaptive immune response provides a second line of protection. In contrast to innate immunity, adaptive immunity is able to distinguish self from nonself to eliminate microbial-infected or mutated cancer cells in an antigen-specific manner without damaging normal cells. The adaptive immune response in skin is carried out primarily by T cells. T-cell mediated immunity is dependent on activation by antigen-presenting DCs.

DENDRITIC CELLS

DCs are a heterogeneous group of leukocytes that, as the name implies, are composed of a central nucleated body from which dendrites emanate. This ensures a large surface area for interaction with T cells and with which to sample the surrounding environment (Fig. 23.2) (Chapter 6). Under resting conditions, two different types of DCs are present in skin: myeloid-derived cDCs and pDCs. cDCs are potent APCs. pDCs are less dendritic and are ineffective APCs, but produce type I IFNs in the skin. pDCs reside in the dermis and are a cellular component of the innate immune response. Unlike conventional DCs, pDCs function primarily to release type I IFN-α and -β. pDCs have also been implicated in the pathogenesis of psoriasis (Chapter 64). In perturbed skin, inflammatory DCs (iDCs) accumulate to produce abundant levels of TNF-α and inducible nitric oxide synthase (iNOS). They are derived from circulating monocytes that differentiate into DCs upon transendothelial migration into inflamed dermis.

To perform antigen-presenting function, LCs and DCs express antigen capture receptors that internalize bound microbes and antigens into processing endosomes, where antigens are cleaved and trimmed into short peptides and then loaded into the peptide-binding groove of MHC (human leukocyte antigen [HLA]) molecules in human) (Chapters 5 and 6). Peptide–MHC molecular complexes are then "presented" at the surface of DCs for recognition by antigen-specific T-cell receptors (TCRs) (Chapter 4) on naïve and/or memory T cells. Through integration of environmental signals, LCs and DCs are able to program naïve T cells into appropriate, functionally distinct subsets of skin-homing effector cells that can respond to a specific pathogen.

Epidermal Langerhans Cells

LCs are specialized epidermal DCs originally discovered by Paul Langerhans in 1868. They are potent APCs for activation of T cells. In normal skin, resting LCs reside in the suprabasal layer of the epidermis. There they cover the entire skin surface in a net-like structure (see Fig. 23.2). LCs represent 1% to 3% of all cells in the epidermis. They are poised to sense and respond to microbial infections or physical breaches of epidermal integrity. LCs are identified by selective expression of the antigen capture receptor langerin (CD207), a C-type lectin. Langerin internalizes to form a definitive "tennis racket–shaped" organelle called the *Birbeck granule*, which is a specialized antigen-processing compartment (see Fig. 23.2). Human LCs are also identified by their expression of CD1a. This protein is structurally similar to MHC/HLA class I molecules but presents lipid and glycolipid microbial antigens, rather than peptides, to specialized populations of T cells. In vivo studies of mouse LCs reveal that they can extrude their dendritic processes across the skintight junction present in the granular layer, through a specialized pore, called *tricellulin*, to sample the microbial environment in the stratum corneum. Langerin receptors concentrate at the tips of the extruded dendrites to aid in the capture of external pathogens.

LCs have long been considered the paradigm-defining cDCs that transfer site-specific antigens to the lymph node, where they inform and activate CD4 and CD8 T cells (Chapter 9) to develop into differentiated skin-homing helper and cytotoxic effector T cells to eliminate the invading threat. Limited studies of primary human LCs indicate they are potent inducers of cytotoxic CD8 T cells and CD4 Th2 cells.[17]

In addition to their potent cross-presenting activity to CD8 T cells (Chapter 12), human LCs also favor development of CD4 regulatory T cells (Tregs) (Chapter 13), antimicrobial Th17 cells, and T follicular helper (Tfh) cells for antibody development (Chapter 11).[17,18] LCs that emigrate from the epidermis leave a hole in the skin surface network. This hole is subsequently replenished for the short term under inflammatory conditions by iDCs and for the long term by resident LC-committed stem cells that seed the epidermis during fetal development.

Dermal Dendritic Cells

Dermal DCs efficiently present to T cells antigens that have reached the dermis. In normal skin, multiple subpopulations of resident dermal DCs can be detected.[19] In mice, langerin-positive and -negative DCs are found in the dermis. Both are developmentally distinct from LCs. In humans, most dermal DCs are langerin negative; only a small number of langerin-low dermal DCs can be detected. Nevertheless, human langerin-negative dermal DCs share functional characteristics with mouse langerin-positive dermal DCs. They are able to cross-present captured intracellular bacterial or viral antigen from infected skin cell debris to activate CD8 T cells without being infected themselves. Thus, they are able to forgo the canonical MHC class I antigen processing pathway, which requires infection for endogenous expression of the pathogen antigens. Mouse

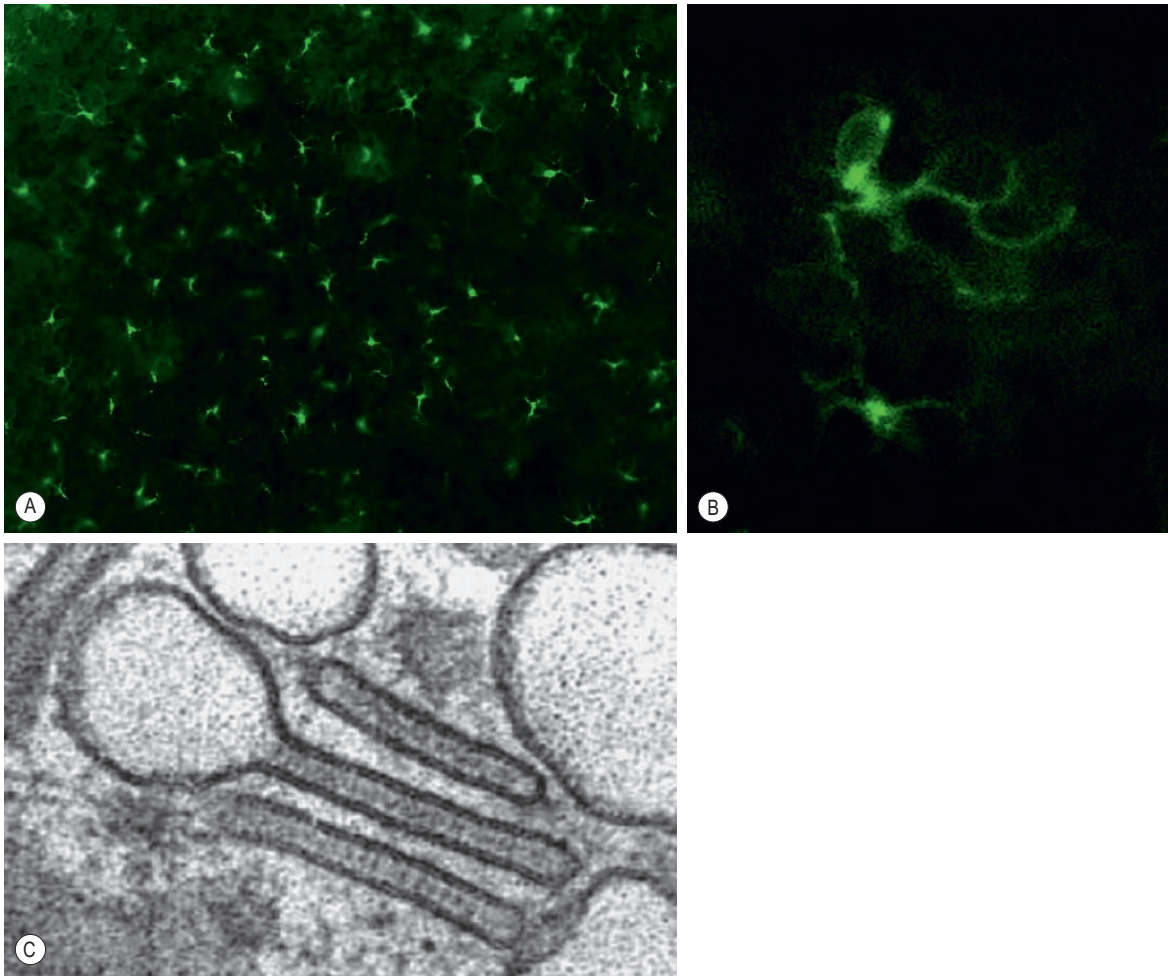


FIG. 23.2 Langerhans Cells (LCs). (A) A horizontal view of epidermis (murine) stained with an antibody to major histocompatibility complex (MHC) class II molecules. Only LCs are positive in unperturbed skin. Dendritic shaped LCs are spaced throughout the epidermis with dendrites stretching out to meet and communicate with neighboring LCs. (B) High-power view of an LC. Note contact between adjacent LCs. (C) Transmission electron microscopy of Birbeck granules in LCs from human epidermis. The Birbeck granule has a "tennis racket" morphology and is formed by internalized langerin molecules. (From Romani N, Clausen BE, Stoitzner P, et al. Langerhans cells and more: langerin-expressing dendritic cell subsets in the skin. *Immunol Rev*. 2010:234[1]:120–141.)

langerin-positive dermal DCs are also critical for activating Tfh cells for IgG2a/c and IgG2b isotype antibody production. In the resting state, all DCs are poor stimulators of T cells. Interaction with so-called immature DCs can generate unresponsive antigen-specific T cells, a key mechanism for maintaining immune tolerance, which is needed to prevent the immune system from attacking self-tissues.

T-CELL SUBPOPULATIONS AND THE SKIN

In response to instructional cues from cutaneous DCs, antigen-specific T cells differentiate into specialized subsets, each with functions that focus on eliminating specific types of pathogens (see Table 23.3). Pre-sensitized, functionally committed T cells that home to skin remain in the dermis as long-lived resident memory T cells (T_{RM}). The skin of a healthy adult contains nearly 20 billion T_{RM} cells, about twice the number present in the entire blood volume.[20] Among the best-characterized differentiated CD4 T-cell subsets are Th1, Th2, Th17, Th22, and Tregs, each of which has a distinct cytokine profile. CD8 T cells primarily function as cytotoxic effector cells, but also express cytokine profiles analogous to CD4 subsets, termed Tc1, Tc2, and Tc17.[21]

Th1 Responses

Th1 cells are the principal mediators of immunity that eradicate intracellular pathogens and tumors. Their major effector cytokine is IFN-γ. IFN-γ from Th1 cells stimulates macrophages to phagocytose and generate oxidative bursts that aid intracellular killing of microbes. IFN-γ also upregulates expression of class I and class II MHC molecules and ICAM-1 (CD54) on keratinocytes, dermal microvascular endothelial cells, and fibroblasts, and induces them to secrete proinflammatory cytokines and chemokines. Deficiencies in Th1 cells increase susceptibility to cutaneous microbial infections and tumors. Th1 cells are responsible for cutaneous delayed-type hypersensitivity responses (e.g., tuberculosis skin test reaction) and allergic contact dermatitis (Chapter 48).

Th2 Responses

Th2 cells are involved in immune responses important for eradication of extracellular parasites and bacterial infections. They produce IL-4, IL-5, IL-10, and IL-13. Th2 cell-mediated inflammation is characterized by the presence of eosinophils (Chapter 45) and basophils as well as extensive mast cell degranulation. In mice, Th2-cell deficiency profoundly increases susceptibility to *Leishmania* infection in skin. In humans, Th2 cells play a critical role in the pathogenesis of atopic dermatitis (Chapter 48). Dupilumab is a monoclonal antibody that inhibits IL-4/IL-13 signaling. It effectively blocks the decrease of skin barrier function, class switch to IgE, and differentiation of Th2 cells. It is approved for the treatment of moderate to severe atopic dermatitis.[22]

Th17 Responses

Th17 cells produce IL-17A, IL-17F, IL-21, and IL-22. Th17 cells protect against extracellular bacterial and fungal infections in skin. Both IL-17A and IL-17F enhance protective immune responses by inducing keratinocytes production of granulocyte colony-stimulating factor (G-CSF), and antimicrobial peptides. Mice deficient in IL-17A are highly susceptible to *S. aureus* infections in skin. IL-17 and IL-22 are important mediators of psoriasis and are implicated in other autoimmune skin disorders. Anti-human IL-17 antibodies (secukinumab and ixekizumab) and an IL-17 receptor antagonist (brodalumab) are effective in over 75% of individuals with moderate to severe psoriasis.[23] Because IL-23 promotes Th17-cell development, antibodies that block p19 subunit of IL-23 (guselkumab, risankizumab, tildrakizumab, bimekizumab) or the p40 subunit common to both IL-12 and IL-23 (ustekinumab) are effective therapeutic agents for psoriasis.[23]

Regulatory T Cells

Tregs diminish immune responses in skin, an important function for dampening inflammatory responses and limiting tissue damage (Chapter 13). Conversely, an increase in Treg suppression can impair host immune surveillance against cutaneous tumors. Natural Tregs (nTregs) develop in the thymus in response to self-antigens, whereas induced Tregs (iTregs) develop in the periphery in response to foreign antigens. Both are commonly characterized as $CD4^+/CD25^+/Foxp3^+$. Both produce immunosuppressive cytokines (IL-10 and/or transforming growth factor [TGF]-β). They can also express immunoregulatory surface molecules, such as CTLA-4, PD-1, glucocorticoid-induced TNFR family–related gene (GITR), and lymphocyte activation gene 3 (LAG-3). They can directly inhibit T-cell activation and function, and suppress the antigen-presenting activity of cDCs. A defect in Treg function will result in autoimmune and inflammatory diseases. About 5% to 10% of the T cells resident in normal human skin are Tregs.[24]

CD8 T-Cell Immunity

CD8 T cells differentiate into cytotoxic T lymphocytes (CTLs), which are essential for the elimination of viral infections and cancerous cells (Chapter 12). CD8 T cells are important effectors of immune surveillance. Their presence in tumor tissues is a good prognostic factor for certain types of cancer.[25] Diminished CD8 T-cell function increases the susceptibility to squamous cell and basal cell carcinomas. Cytotoxic CD8 T cells and CD8 T cells producing IFN-γ or IL-17 (Tc1 and Tc17, respectively) are key mediators in allergic contact dermatitis (Chapter 48).[26]

γδ T Cells

γδ T cells, also known as dendritic epidermal T cells (DETCs), bridge innate and adaptive immunity. In place of conventional αβ chain TCRs, they express a more restricted repertoire of γ and δ chain T-cell receptors (TCRs) (Chapter 4). They reside in epithelial tissues of skin, gut, lungs, and reproductive tract.[27] In humans, γδ T cells account for at least 10% of the T cells in the epithelium. Epithelium-resident γδ T cells differ from circulating γδ T cells with respect to their development, selection, TCR diversity, and effector functions. Epithelial γδ T cells are not restricted by MHC class I or class II molecules (Chapters 5 and 6). Instead, they recognize PAMP and DAMP molecules presented by non-classic MHC molecules, such as CD1. They use alternate sets of costimulatory molecules, such as the junctional adhesion molecule–like protein (JAML) and its ligand, the coxsackievirus and adenovirus receptor (CAR). They are thought to be partially activated under static conditions. This enables them to mount responses to pathogenic stimuli rapidly. γδ T cells are also involved in epidermal wound repair, can influence DC activation, and can directly present antigens.

MAST CELLS

Skin is a rich source of mast cells (Chapter 44).[28] These cells are derived from $CD34^+$ progenitors in bone marrow and migrate into tissues at environmental interfaces. In skin, they are present in the dermis at a density of up to 20,000/mm^2. They are typically concentrated around the dermal microvasculature, appendages, and nerves. Their cytoplasmic granules contain a plethora of preformed mediators that include histamine and heparin; the proteases tryptase, chymase, carboxypeptidase, arylsulfatase A, β-hexosaminidase, and β-glucuronidase; and the cytokines TNF-α, GM-CSF, IL-3, IL-4, IL-5, IL-6, IL-8, and IL-13.

Mast-cell activation plays a prominent role in urticaria and angioedema. Mast-cell activation is also implicated in the pathogenesis of bullous pemphigoid, leukocytoclastic vasculitis, atopic dermatitis, allergic contact dermatitis, and mastocytosis (urticaria pigmentosa).

EVENTS INVOLVED IN THE GENERATION OF A PROTOTYPIC INNATE AND ADAPTIVE IMMUNE RESPONSE IN THE SKIN

Keratinocytes constitutively produce a pool of inactive proinflammatory precursors: pro–IL-1α, pro–IL-1β, pro–IL-18, and others. In response to immunological and inflammatory stimuli, these molecules can be immediately processed into active mediators by inflammasomes, which are rapidly formed in the presence of DAMPs. IL-1β, a potent stimulator of endothelial cell activation, permits rapid infiltration of inflammatory myeloid cells and T cells from blood into the dermis. TNF-α contributes to dermal microvascular endothelial cell activation and augments expression of ICAM-1 (CD54) on keratinocytes and endothelial cells. Infiltrating T cells express the β_2-integrin molecule leukocyte function–associated antigen-1 (LFA-1), which binds to ICAM-1 expressed on affected cells in the dermis and epidermis. IL-6 helps drive the development of IL-17–producing Th17 cells. IL-6 can also perpetuate chronic inflammation.

Activated keratinocytes respond to these proinflammatory mediators by generating a second set of cytokines and chemokines (including GM-CSF, TNF-α, and CXCL8/IL-8) with

autocrine and paracrine effects. They activate unaffected neighboring keratinocytes to amplify proinflammatory signals. These cytokines are a key to triggering adaptive immune responses by activating cutaneous LCs and DCs to become mature potent APCs. APC interaction with antigen-specific, skin-resident memory T cells stimulates proliferation and expansion of differentiated effector T cells at the affected skin site. Keratinocyte-derived IL-8 is a potent chemoattractant for neutrophils. The influx of leukocytes further amplifies local immune responses, inflammation, and associated tissue damage. Stimulation of TLRs 7 and 9 within small numbers of resident dermal pDCs leads to abundant production of type I IFN-α, inducible nitric oxide synthase (iNOS), and arginase, which aid in eradication of viral infection and other invading organisms.

This first-wave innate immune response is followed by a slower, but more specific, second-wave adaptive immune response to control and eliminate antigenic stimuli (FIG. 23.3). Antigens are taken up by epidermal LCs and dermal DCs. They carry antigens from the affected skin site to draining lymph nodes, where they present those antigens to activate and expand T cells specific for the antigens to which the skin was exposed (Chapter 6). To emigrate from the epidermis to regional lymph nodes, LCs downregulate E-cadherin, allowing them to detach from surrounding keratinocytes. During migration to the draining lymph node, LCs and DCs undergo changes that halt antigen capture activity (by downregulation of antigen capture receptors) and increase expression of T-cell interactive molecules: MHC class I and II molecules, costimulatory molecules CD80 and CD86, and ICAM-1 (CD54).

Naïve T cells in lymph nodes that express TCRs for the antigens in skin form APC–T cell conjugates during which T cells are programmed to differentiate into functionally distinct T-cell subsets. These newly activated T cells leave the lymph node by downregulating their lymph node–homing receptors, CCR7, and L-selectin (CD62L) and upregulating the skin-specific homing receptor CLA and chemokine receptors CCR4 and CCR10 to enter the affected skin site.[13] Unless antigen is present in skin for an extended period, the sensitization phase

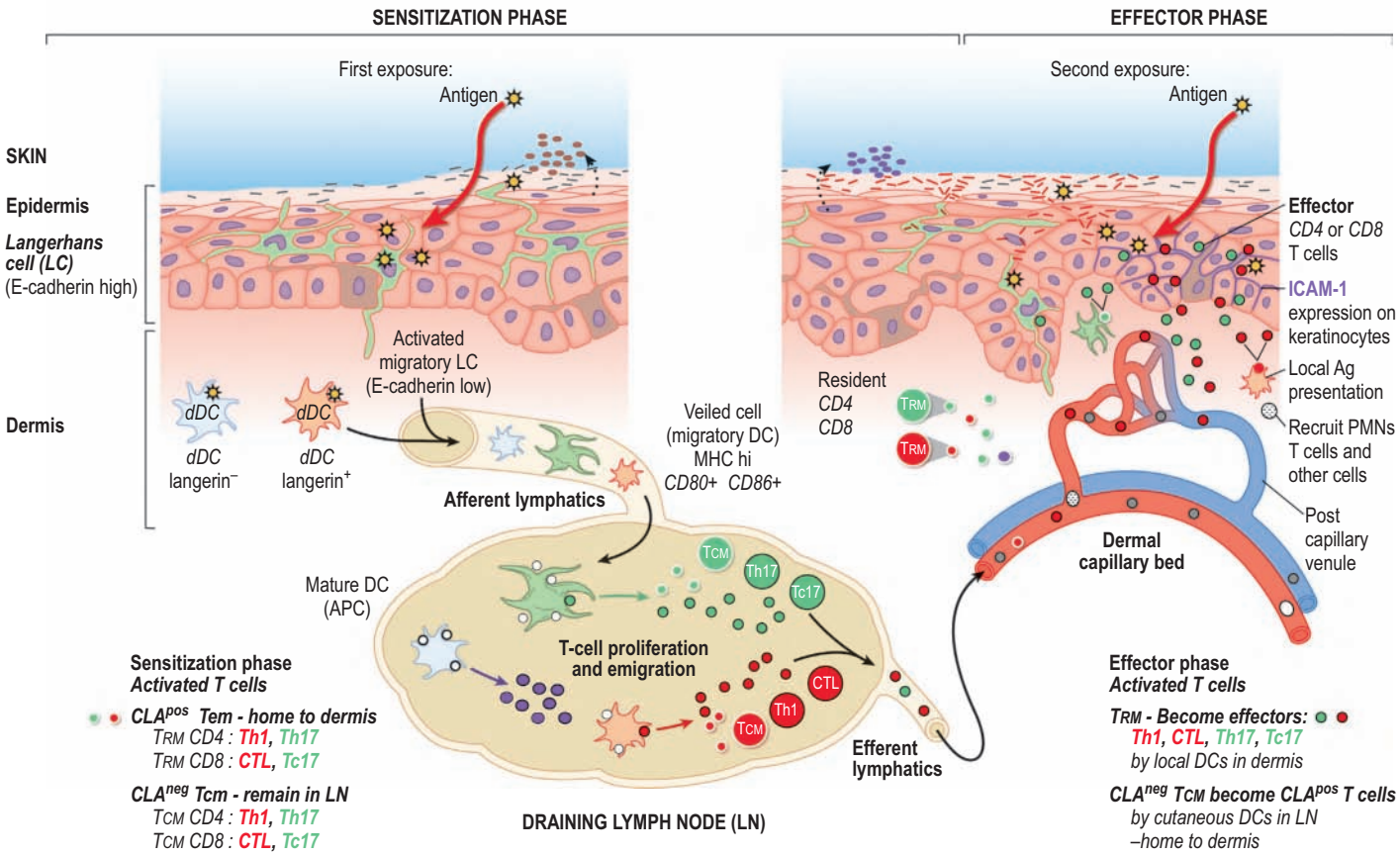


FIG. 23.3 Phases for the Development of Immune Responses in Skin. During the sensitization phase, dendritic cells *(DCs)* capture antigens in skin and migrate through afferent lymphatics to the skin-draining lymph nodes *(LNs)* where they present the antigens to naïve T cells. T-cell differentiation is induced by activated cutaneous DCs, and they develop into resident memory T *(T_{RM})* cells and central memory T *(T_{CM})* cells representing different lineages (Th1, CTL, Th17, Tc17), depending on DC programming. Activated T_{CM} that leave LNs and migrate to skin become skin resident T_{RM}, whereas non-activated T_{CM} remain in lymph nodes. The effector phase occurs in response to subsequent exposures to antigen. DCs become activated, which, in turn, stimulates local T_{RM} to become effector cells in situ. Activated cutaneous DCs that reach the draining LN activate T_{CM} to become skin-homing T effector cells, which are recruited to the inflamed dermal site. These effector cells, in concert with recruited cells, will enter into the epidermal layer by binding intercellular adhesion molecule 1 *(ICAM-1)* molecules present on activated keratinocytes. They act to clear infection, eliminate pathogen-infected cells, remove debris, and repair the skin barrier. (Drawing by Laura Timares, PhD.)

on first exposure to antigen does not cause an inflammatory reaction and often goes unnoticed.

In the dermis of resting skin, resident memory T cells (T_{RM}), previously activated during the sensitization phase, persist in the absence of antigens and are readily available for further encounters with the same antigen. Central memory T cells (T_{CM}) reside in lymph nodes. Unlike T_{RM}, T_{CM} do not express high levels of skin-homing receptors CLA and CCR4 or exhibit immediate robust responses to antigen restimulation. T_{CM} in lymph nodes or spleen can be activated by presentation of antigen by skin-draining LCs or DCs. Activated T_{CM} enter the circulation, express skin-homing receptors, and extravasate into skin sites where antigen is present, thereby further contributing to the immune response. CD4 and CD8 T_{RM} and recruited T_{CM} that interact with antigen-activated LCs and DCs in the dermis are further stimulated to proliferate in situ to *(i)* amplify effector cell numbers, *(ii)* augment cytokine and chemokine production, and *(iii)* recruit additional inflammatory leukocytes, all of which act in concert to eradicate the antigen present in skin. Events are terminated by regulatory T cells and immunosuppressive cytokines such as IL-10, TGF-β, and IL-1 receptor antagonist (IL-1RA). This allows the skin to return to its non-inflamed status. IL-10 is a potent inhibitor of adaptive and inflammatory cellular responses, whereas IL-1RA effectively blocks the activity of IL-1β. In addition, macrophages and certain DC subsets can inhibit ongoing adaptive immune responses through their expression of indoleamine 2,3-dioxygenase (IDO), an enzyme that alters tryptophan metabolism in T cells.

ANTIBODIES AND SKIN

Although the role of IgG, IgA, and IgE antibodies (Chapters 4 and 8) in normal skin homeostasis is not well understood, Igs often play key roles in the pathogenesis of several different dermatological diseases. Pathogenic IgG causes blistering disorders, including pemphigus, bullous pemphigoid, epidermolysis bullosa acquisita, paraneoplastic pemphigus, and leukocytoclastic vasculitis. IgA is key to the pathogenesis of dermatitis herpetiformis, linear IgA bullous dermatosis, and IgA-mediated cutaneous vasculitis. IgE-dependent skin diseases include urticaria and angioedema and bullous pemphigoid.

Urticaria

Urticaria (Chapter 46) is a form of IgE antibody–mediated adaptive immunity in the skin. Mast cells express high-affinity surface receptors for the Fc portion of IgE (FcεRI), which binds and retains IgE for long periods. Antigen-induced cross-linking of IgE bound to FcεRI initiates calcium- and energy-dependent events that culminate in fusion of granules to the plasma membrane and release of granule contents. Degranulation releases potent prostanoids, histamines, and inflammatory cytokines that rapidly recruit inflammatory cells. Non-antigenic stimuli, such as opiates, C5a, anaphylatoxin, stem cell factor, and substance P, can activate degranulation through FcR-independent opioid, adenosine, and β-adrenergic receptors. The products and contents of mast-cell granules are causative agents of acute vasodilatation, edema, pruritus, and rapid influx of leukocytes and eosinophils into skin.

Hyper-Immunoglobulin E Syndrome

Hyper-immunoglobulin E (IgE) syndrome (HIES) is an autosomal dominant disease that develops cutaneous manifestations of severe dermatitis, recurrent infections of staphylococcal abscesses, and, sometimes, cutaneous candidiasis. Dominant negative mutations in the signal transducer and activator of transcription 3 (STAT3) gene are associated with this disease. Deficient STAT3 signaling leads to impaired β-defensin expression, which may explain some of the clinical manifestations.[29]

ENVIRONMENTAL CHALLENGES AND THE SKIN-ASSOCIATED LYMPHOID TISSUE

Chemicals

Environmental chemicals that are able to breach the cutaneous physical barrier are often identified by the host, correctly or incorrectly, as unwelcoming molecules that should be neutralized or eliminated by the immune system. When this occurs to a seemingly innocuous molecule, it causes allergic contact dermatitis. Examples include urushiol, the active moiety in poison ivy, oak, and sumac, and nickel sulfate, responsible for the dermatitis caused by nickel-containing jewelry. Once beyond the stratum corneum, these chemicals bind to epidermal proteins, generating "modified self" proteins as immunogenic complete antigens. Animal models of allergic contact dermatitis have been studied extensively to understand the mechanism of action of T-cell-mediated adaptive immune responses in the skin.

Solar Ultraviolet Radiation

Sunlight is the major environmental agent to which skin is exposed. Injudicious exposure to UV spectrum can lead to sunburn, skin aging, skin cancers, and photosensitivity diseases, many of which have an immunological pathogenesis. Although much of the investigation into the immunological effects of UVR was conducted in animal models, many of the observations have been corroborated in humans.

In mice, as in humans, chronic exposure to UVR results in the development of highly antigenic skin tumors, capable of stimulating a vigorous antitumor response in untreated mice. Despite their antigenic nature, these tumors grow progressively in their original host. This apparent paradox was resolved in studies showing that, in addition to producing mutant neoplastic cells, UVR also impairs cell-mediated immune surveillance, which, under normal circumstances, eliminates mutant cells before they develop into clinically apparent tumors. Thus, mutant cells progress to become tumors only in an environment of immune suppression. For example, organ transplant recipients who are treated with immunosuppressive medications have a greatly increased risk of developing aggressive skin cancers.[30]

UVR mediates its effects on the immune system by perturbing the function of skin APCs. UV-irradiated APCs are poor stimulators of Th1 cells but are able to activate Tregs, which promote antigen-specific immunological tolerance. Enhanced production of the suppressive cytokines IL-10 and TGF-β, as well as by a reduction in the Th1 activation cytokine IL-12 also play a role. It may seem surprising that an environmental carcinogen, such as UVR, suppresses immunological function in skin. One proposed hypothesis is that altered epithelial proteins are constantly generated by UVR exposure, necessitating chronic induction of immune tolerance to UVR-damaged proteins to preserve the integrity of the skin barrier.

The immunosuppressive effect of UVR has been exploited for therapeutic purposes. UVR phototherapy is used to suppress pathological immune responses in such inflammatory skin diseases as psoriasis and atopic dermatitis.[30]

ON THE HORIZON

- Development of small-molecule inhibitors of JAK/STAT signaling that can be given orally or topically for the treatment of psoriasis.
- Elucidation of the mechanisms underlying comorbidities, such as metabolic syndrome and major adverse cardiovascular events in patients with psoriasis.
- Development of new cytokine inhibitors to treat a spectrum of skin diseases (such as IL-17 and IL-23 inhibitors for the management of hidradenitis suppurativa).
- Targeting of dendritic cell subsets (DCs) in order to optimize vaccination.
- Use of cytokines or cytokine inhibitors to manipulate cutaneous immune responses in immunologically mediated skin diseases (such as vitiligo, alopecia areata, cutaneous lupus erythematosus, and dermatomyositis).

REFERENCES

1. Hammers CM, Stanley JR. Recent advances in understanding pemphigus and bullous pemphigoid. *J Invest Dermatol*. 2020;140(4):733–741.
2. Yohn JJ, Critelli M, Lyons MB, et al. Modulation of melanocyte intercellular adhesion molecule-1 by immune cytokines. *J Invest Dermatol*. 1990;95(2):233–237.
3. Sumpter TL, Balmert SC, Kaplan DH. Cutaneous immune responses mediated by dendritic cells and mast cells. *JCI Insight*. 2019;4(1):e123947.
4. Takeuchi O, Akira S. Pattern recognition receptors and inflammation. *Cell*. 2010;140(6):805–820.
5. Burns EM, Yusuf N. Toll-like receptors and skin cancer. *Front Immunol*. 2014;5:135.
6. Conti HR, Gaffen SL. IL-17-mediated immunity to the opportunistic fungal pathogen *Candida albicans*. *J Immunol*. 2015;195(3):780–788.
7. Kawamura T, Ogawa Y, Aoki R, et al. Innate and intrinsic antiviral immunity in skin. *J Dermatol Sci*. 2014;75(3):159–166.
8. Zhu H, Lou F, Yin Q, et al. RIG-I antiviral signaling drives interleukin-23 production and psoriasis-like skin disease. *EMBO Mol Med*. 2017;9(5):589–604.
9. Takahashi T, Gallo RL. The critical and multifunctional roles of antimicrobial peptides in dermatology. *Dermatol Clin*. 2017;35(1):39–50.
10. Schauber J, Dorschner RA, Coda AB, et al. Injury enhances TLR2 function and antimicrobial peptide expression through a vitamin D-dependent mechanism. *J Clin Investig*. 2007;117(3):803–811.
11. Dorschner RA, Williams MR, Gallo RL. Rosacea, the face of innate immunity. *Br J Dermatol*. 2014;171(6):1282–1284.
12. Lande R, Botti E, Jandus C, et al. The antimicrobial peptide LL37 is a T-cell autoantigen in psoriasis. *Nat Commun*. 2014;5:5621.
13. Tan SY, Roediger B, Weninger W. The role of chemokines in cutaneous immunosurveillance. *Immunol Cell Biol*. 2015;93(4):337–346.
14. Kim YH, Bagot M, Pinter-Brown L, et al. Mogamulizumab versus vorinostat in previously treated cutaneous T-cell lymphoma (MAVORIC): an international, open-label, randomised, controlled phase 3 trial. *Lancet Oncol*. 2018;19(9):1192–1204.
15. Solimani F, Meier K, Ghoreschi K. Emerging topical and systemic JAK inhibitors in dermatology. *Front Immunol*. 2019;10:2847.
16. Georgin-Lavialle S, Ducharme-Benard S, Sarrabay G, et al. Systemic autoinflammatory diseases: clinical state of the art. *Best Pract Res Clin Rheumatol*. 2020:101529.
17. Clausen BE, Kel JM. Langerhans cells: critical regulators of skin immunity? *Immunol Cell Biol*. 2010;88(4):351–360.
18. Yao C, Zurawski SM, Jarrett ES, et al. Skin dendritic cells induce follicular helper T cells and protective humoral immune responses. *J Allergy Clin Immunol*. 2015;136(5):1387–1397. e1–7.
19. Heath WR, Carbone FR. The skin-resident and migratory immune system in steady state and memory: innate lymphocytes, dendritic cells and T cells. *Nat Immunol*. 2013;14(10):978–985.
20. Clark RA. Resident memory T cells in human health and disease. *Sci Transl Med*. 2015;7(269):269rv1.
21. He D, Wu L, Kim HK, et al. $CD8^+$ IL-17-producing T cells are important in effector functions for the elicitation of contact hypersensitivity responses. *J Immunol*. 2006;177(10):6852–6858.
22. Seegraber M, Srour J, Walter A, et al. Dupilumab for treatment of atopic dermatitis. *Expert Rev Clin Pharmacol*. 2018;11(5):467–474.
23. Menter A, Strober BE, Kaplan DH, et al. Joint AAD-NPF guidelines of care for the management and treatment of psoriasis with biologics. *J Am Acad Dermatol*. 2019;80(4):1029–1072.
24. Uttarkar S, Brembilla NC, Boehncke WH. Regulatory cells in the skin: pathophysiologic role and potential targets for anti-inflammatory therapies. *J Allergy Clin Immunol*. 2019;143(4):1302–1310.
25. Yusuf N, Nasti TH, Katiyar SK, et al. Antagonistic roles of $CD4^+$ and $CD8^+$ T-cells in 7,12-dimethylbenz(a)anthracene cutaneous carcinogenesis. *Cancer Res*. 2008;68(10):3924–3930.
26. He D, Wu L, Kim HK, et al. IL-17 and IFN-gamma mediate the elicitation of contact hypersensitivity responses by different mechanisms and both are required for optimal responses. *J Immunol*. 2009;183(2):1463–1470.
27. Sutoh Y, Mohamed RH, Kasahara M. Origin and evolution of dendritic epidermal T cells. *Front Immunol*. 2018;9:1059.
28. Galli SJ, Grimbaldeston M, Tsai M. Immunomodulatory mast cells: negative, as well as positive, regulators of immunity. *Nat Rev Immunol*. 2008;8(6):478–486.
29. Minegishi Y, Saito M, Nagasawa M, et al. Molecular explanation for the contradiction between systemic Th17 defect and localized bacterial infection in hyper-IgE syndrome. *J Exp Med*. 2009;206(6):1291–1301.
30. Elmets CA, Cala CM, Xu H. Photoimmunology. *Dermatol Clin*. 2014;32(3):277–290, vii.

24

Immunology of Mucosal Surfaces

Prosper N. Boyaka and Kohtaro Fujihashi

Mammals have evolved into a sophisticated network of cells and molecules that serves to maintain homeostasis on exposed mucosal surfaces. This system is anatomically and functionally distinct from its bloodborne counterpart and is strategically located at the portals through which most pathogenic microorganisms enter the body. This specific branch of the immune system may have developed in response to the size of the mucosal surfaces, which cover an area of ≈70 m^2 in the airways and ≈400 m^2 in the gut of an adult human, and the large numbers of commensal bacteria and exogenous antigens to which the mucosae are exposed.[1]

THE INNATE MUCOSAL DEFENSE SYSTEM

Cells and molecules that contribute to innate defense of the mucosa include the physical barrier provided by epithelial cells, the movement of the epithelial cilia, the production of mucus by goblet cells, the secretion of molecules with innate antimicrobial activity, and the cytolytic activity of natural killer (NK) cells (Fig. 24.1, A). Innate lymphoid cells (ILCs; Chapter 3) were identified recently as key players in innate mucosal immunity (See Fig. 24.1, B). In concert with the commensal microbiota (Chapter 22), these innate mechanisms provide a first line of defense against exogenous antigens and invading pathogens. Epithelial Cells and Other Effectors of the Mucosal Physical Barrier

KEY CONCEPTS

Innate Defenses of the Mucosal Immune System

The innate defenses of the mucosal immune system provide a first line of protection against the entry of exogenous antigens and microbes, as well as invading pathogens. These defenses include:

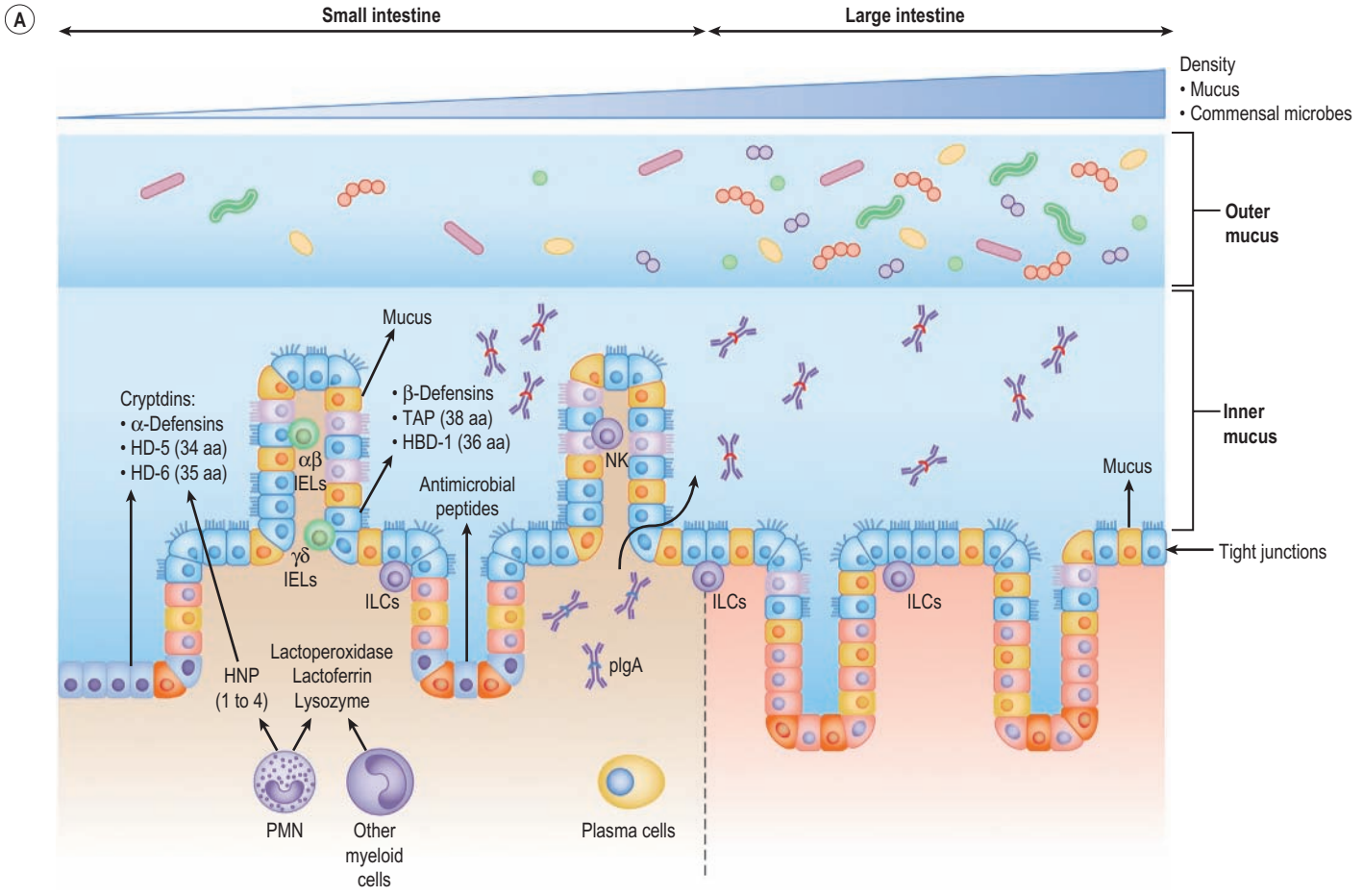
- *Physical and chemical barriers:* the epithelium, epithelial cell cilia, goblet cell mucus production, acid production by the stomach, lipid production by skin.
- *Mucosal antimicrobial molecules:* Paneth cell production of α-defensins in the small intestine; epithelial cell production of β defensins in the oral mucosa, trachea, bronchi, mammary glands, and salivary glands: lactoferrin, lysozyme, lactoperoxidase, and secretory leukocyte protease inhibitor (SLPI).
- *Cellular innate immunity:* mucosal natural killer (NK) cells, innate lymphoid cells (ILCs), dendritic cells (DCs), polymorphonuclear neutrophils (PMNs), mast cells and eosinophils.

All mucosal surfaces are covered by epithelial cells, which contribute to their selective barrier function. In the gastrointestinal (GI) tract, tightly joined enterocytes constitute the cellular component of the physical barrier and are covered by a blanket of mucus. Mucus consists of glycoproteins secreted into the lumen by goblet cells. This layer of glycoproteins interferes with the attachment of microorganisms to the mucosal surface. The replacement of damaged or infected enterocytes by crypt epithelial cells, which differentiate into enterocytes as they migrate toward the desquamation zone at the villus tip, ensures the integrity of this barrier. Multilayered squamous epithelial cells cover other mucosal surfaces, including the oral cavity, pharynx, tonsils, urethra, and vagina. These epithelia lack tight junctions. Instead, mucus coats the intercellular space between the lower stratified epithelial cell layers. Polymeric immunoglobulin A (pIgA) and commensal microbes support the physical barrier function of mucosal tissues. Perturbation of the commensal microbiome, which is present in the mucus ecosystem, facilitates opportunist infections by pathogens such as *Clostridium difficile*.

Defensins and Other Mucosal Antimicrobial Peptides

Selected epithelial cell subsets contribute to innate responses through the production of antimicrobial peptides, iron transporters, and enzymes. Defensins are 30 to 40 amino acid β-sheet peptides with antiviral activity and antimicrobial effects similar to those of antibiotics. Defensins are structurally segregated into α and β categories. α-Defensins are secreted by tracheal epithelial cells and by Paneth cells in intestinal crypts. α-Defensins are homologous to peptide mediators of nonoxidative microbial cell killing in neutrophils (termed *human neutrophil peptides* [HNPs]).[2] Human β-defensin 1 (HBD-1) is expressed in the epithelial cells of the oral mucosa, trachea, bronchi, mammary glands, and salivary glands, whereas HBD-5 is expressed in the gut. Inflammatory cytokines (Chapter 14), including interleukin-1 (IL-1), IL-17, tumor necrosis factor (TNF), and bacterial lipopolysaccharide (LPS), regulate defensin production.

Other antimicrobial products of the epithelium include lactoferrin, lysozyme, peroxidases, secretory phospholipase A2 (S-PLA2), and cathelin-associated peptides. Lactoferrin, a member of the transferrin family, is found in exocrine secretions. S-PLA2 and lysozymes are released by Paneth cells and high concentrations of lysozyme (1209 to 1325 μg/mL) are found in tears, saliva, colostrum, serum, and urine. Human milk contains lysozyme in concentrations ranging from 20 to 245 μg/mL, depending on the lactation period. Milk leukocytes produce myeloperoxidase (MPO), and mammary gland cells produce human lactoperoxidase (hLPO). Both peroxidases display properties similar to those of human salivary peroxidases (hSPO). Secretory leukocyte protease inhibitor (SLPI) is found in human saliva, nasal secretions, tears, cervical mucus, and seminal fluid. It is believed to be responsible for the anti–human immunodeficiency virus (HIV) properties of external secretions.



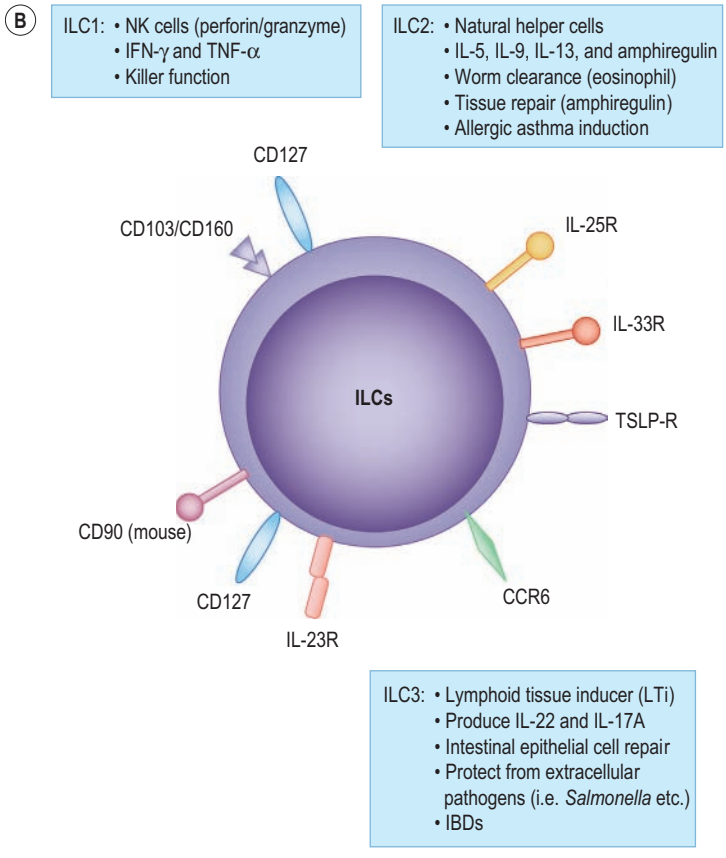


FIG. 24.1 Innate Mucosal Host Defense Factors. (A) A thick coat of mucus prevents penetration of exogenous macromolecules, commensal microbes, and potential pathogens. The epithelial cell barrier is connected via tight junctions and contains both αβ and γδ intraepithelial T lymphocytes (IELs). The crypt regions contain Paneth cells, which produce cryptins (α-defensins). β-Defensins are products of epithelial cells and form a defensin network. Other innate factors such as lysozyme, lactoperoxidase, lactoferrin, and phospholipases, also serve in antimicrobial defense. (B) Three types of innate lymphoid cells are identified based on their cytokine production.

Mucosal Innate Lymphoid Cells

ILCs, recently described, constitute groups of lymphoid cells that are predominantly found in mucosal tissues. ILCs lack T- and B-cell receptors. They contribute to innate regulation of homeostasis through their ability to rapidly produce cytokines. Similar to CD4 T-helper (Th) cells (Chapter 11), ILCs can be divided into three major groups according to the cytokines they produce.[3,4] Group 1 ILCs (ILC1) produce interferon-γ (IFN-γ) and are considered similar to Th1 cells. Cells with ILC1 phenotypes express CD103/CD160 and CD127 and include NK cells, which are large granular lymphocytes found in the lamina propria and the intraepithelial compartment. Group 2 ILCs are similar to Th2 cells and produce IL-5, IL-9, IL-13, and amphiregulin. ILC2 are responsible for innate responses in allergies and asthma. As with Th17 cells, group 3 ILCs (ILC3) secrete IL-17 and/or IL-22. Although lymphoid tissue inducer (LTi) cells are ILC3, functional properties of these cells are distinct from other mucosal ILC3.[3,4]

Two types of ILC1 occur in the tonsils and the mucosa of the GI tract. $CD103^+$, $CD160^+$ ILCs are found in the intraepithelium compartment of the intestines and produce "*perforin*" a granzyme like NK cells. This ILC1 subset requires Nfil3 and T-bet transcription factors for its development. An ILC1 population expressing low levels of RORγt and aryl hydrocarbon receptor (Ahr) similar to ILC3 was identified in humans. These ILCs express IL-7Rα (CD127), but not conventional lymphocyte lineage or NK cell markers. Both subsets of ILC1 may be involved in the induction of inflammatory bowel disease (IBD).[3] ILC2 are seen in the lungs, upper respiratory mucosa, gut, and skin. By producing IL-5, IL-13, and amphiregulin, they play key roles in the clearance of parasites, including *Nippostrongylus brasiliensis* and *Trichuris muris*. ILC2 are also important in the development of asthma and allergies, including atopic dermatitis. Although ILC2 were reported to promote airway hypersensitivity during acute influenza virus infection, the amphiregulin they produce could also help maintain lung epithelial cell homeostasis.[3] A recent study showed that ILC2 also play a key role in the induction of IgA Ab responses. Thus, absence of ILC2 prevented the induction of IgA Abs that protect the stomach by eliminating IgA-coated bacteria, including pathogenic *Helicobacter pylori*.[5]

ILC3 display heterogenic cytokine profiles in both humans and mice. Thus, human ILC3 bearing NKp44 and CCR6 produce IL-22 alone. However, human ILC3 isolated from patients with Crohn disease produce IL-17 and IFN-γ. Mouse LTi and LTi-like ILC3 produce both IL-17 and IL-22. $NKp46^+$ ILC3 secrete IL-22, as well as IFN-γ under certain conditions, but not IL-17. Another subset of mouse ILC3 that produces IFN-γ, IL-17, and IL-22 can be found in the large intestine. ILC3-derived IL-22 induces antimicrobial peptide responses by intestinal epithelial cells (IECs). Furthermore, IL-22 from ILC3 was reported to enhance innate immunity against *Salmonella* and bacterial infection secondary to influenza virus infection.[3,4] Because ILC3 are also involved in the induction of colonic inflammation and colorectal cancer, they may act as a double-edged sword in the mucosal immune system.[3,4]

THE COMMON MUCOSAL IMMUNE SYSTEM

Higher-order mammals have developed an organized secondary lymphoid tissue system in the GI and upper respiratory tracts. Gut-associated lymphoid tissues (GALTs) include Peyer patches (PPs), the appendix, and solitary lymphoid nodules in the GI tract (Chapter 2). Tonsils and adenoids comprise nasopharyngeal-associated lymphoid tissues (NALTs). Experimental animals, such as rabbits, rats, and guinea pigs, exhibit organized bronchus-associated lymphoid tissues (BALTs) that also occur in human airway branches with inflammation.[1] Together, GALTs and NALTs in humans and GALTs, BALTs, and NALTs in experimental species are termed *mucosa-associated lymphoid tissue* (MALT).

The vast areas of the mucosal immune system characterized by diffuse collections of lymphoid cells are termed "*effector tissues*." These include the interstitial tissues of the mammary, lacrimal, salivary, sweat, and all other exocrine glands, as well as the lamina propria and the epithelium of the GI tract. The lamina propria areas of the upper respiratory and genitourinary tracts are also lymphoid effector sites. MALT is connected with effector sites through the migratory patterns of effector cells.

KEY CONCEPTS

The Common Mucosal Immune System

The term *mucosa-associated lymphoid tissue* (MALT) comprises discrete and diffuse collections of lymphoid tissues that share distinctive features, including a unique type of epithelium, a distinct architecture, a unique set of antigen-presenting cells (APC) and B cells, where switching to immunoglobulin A (IgA) predominates. The involved tissues include:

- *Gut-associated lymphoid tissues (GALTs):* Peyer patches (PPs), the appendix, and solitary lymphoid nodules in the gastrointestinal (GI) tract
- *Nasopharyngeal-associated lymphoid tissues (NALTs):* tonsils and adenoids
- *Effector tissues:* the interstitial tissues of the mammary, lacrimal, salivary, sweat, and all other exocrine glands; the lamina propria and the epithelium of the GI tract; and the lamina propria areas of the upper respiratory and genitourinary tracts.

Mucosa-Associated Lymphoid Tissue as an Inductive Site

MALT has a unique type of epithelium for antigen uptake. Its features include a characteristic architecture, antigen-presenting cell (APC), and B-cell areas with germinal centers where switching to IgA predominates. The columnar epithelium that covers MALT is infiltrated with lymphocytes and APCs, leading to the term follicle-associated epithelium (FAE). Lacking goblet cells, the FAE is covered with far less mucus than normal enterocytes. Soluble and particulate luminal antigens are taken up by microfold (M) cells and are delivered to adjacent APCs. M cells have been described in PPs, the appendix, and tonsils, and represent 10% to 15% of cells within the FAE.[5] M cells are also found in isolated lymphoid follicles (ILFs) and at the tips of the villus, where they are termed "*villous M cells*."[5] The microvilli of these cells, which are less dense than those of adjacent enterocytes, offer a portal of entry into MALT (Fig. 24.2). The M cell is often identified by an invagination of the basolateral membrane into a "pocket" normally occupied by lymphocytes and APCs (Fig. 24.3).

M cells appear ideal for antigen uptake owing to a well-developed microvesicle system that includes endosomes. However, it remains unclear whether M cells act as classic APCs. M cells also provide a portal of entry for some pathogens, such as invasive strains of *Salmonella typhimurium*, but not for noninvasive strains of *S. typhimurium* and reoviruses.

FIG. 24.2 The Microfold (M) Cell. A scanning electron micrograph of an M cell with adjacent enterocytes. The M cell has selectively bound *Escherichia coli* 0157. Note that a thick brush border is lacking, facilitating the binding and uptake of microparticles. (Courtesy of Dr. Tatsuo Yamamoto, Niigata University.)

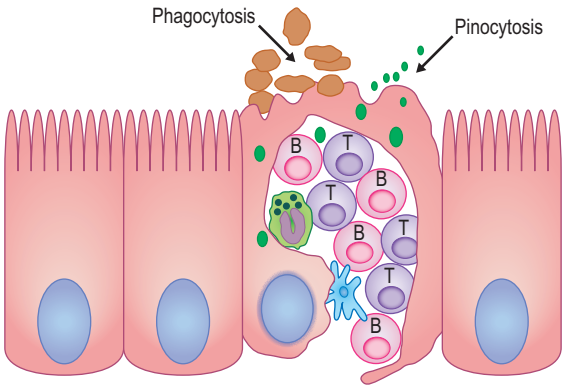


FIG. 24.3 Microanatomic Features of Microfold (M) Cells. The M cell forms a "pocket" containing memory lymphocytes. It actively pinocytoses soluble antigens and phagocytoses particulates such as viruses, bacteria, and microspheres. (Courtesy of Dr. Svein Steinsvoll, University of Oslo.)

Gut-Associated Lymphoid Tissues

Each PP contains a dome region that is positioned under the FAE. This dome region is populated by T cells, B cells, macrophages (MØs), and dendritic cells (DCs). It includes follicles that contain germinal centers. The presence of all three major APC types in the dome (i.e., memory B cells, MØs, and DCs) makes it likely that antigen uptake occurs immediately following release from M cells (Fig. 24.4, A). M cell pockets in PPs contain approximately equal numbers of T and B cells, but fewer MØs. Approximately 75% of the T cells are Th cells.

GALT B-cell follicles are enriched in IgA-bearing B cells, suggesting that they are major sites for B-cell μ to α switching (Chapters 4 and 7). The interfollicular regions of PPs contain high endothelial venules (HEVs) (Chapters 2 and 16). Both CD4 and CD8 TCRαβ T cells are found in these interfollicular regions, with CD4 T cells representing the predominant phenotype. Both naïve and memory T cells are present in PPs, with one-third in cell cycle (see Fig. 24.4; Table 24.1). Lymphotoxin-α (LT-α), lymphotoxin-β (LT-β), and TNF (Chapter 14) are critical for lymphoid tissue organogenesis (Chapter 2). LT-$\alpha^{-/-}$ mice are deficient in secondary lymph nodes, whereas LT-$\beta^{-/-}$ mice have mesenteric and cervical lymph nodes but lack peripheral lymph nodes and PPs. Tumor necrosis factor–receptor I (TNF-RI)$^{-/-}$ mice lack or display abnormal PP structures, whereas TNF$^{-/-}$ mice exhibit normal PPs.

Nasopharyngeal-Associated Lymphoid Tissues (NALTs)

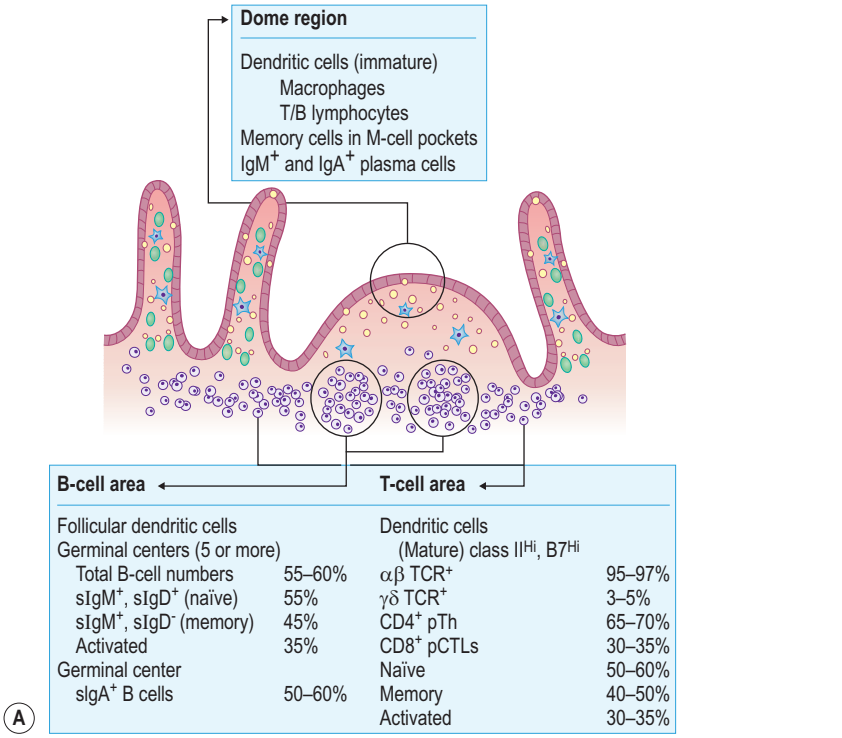
Strategically positioned at the entry of the respiratory and the digestive tracts are the accumulations of lymphoid tissues that comprise the palatine, lingual, and nasopharyngeal tonsils, which collectively form the *Waldeyer ring*. These tissues resemble both lymph nodes and PPs, including a FAE with M cells in the tonsillar crypts that is essential for selective antigen uptake (see Fig. 24.3). Germinal centers containing B and T cells, plasma cells, and APCs are also present. Tonsillar tissues can serve as a source of precursors of IgA plasma cells found in the upper aerodigestive tracts, as well as inductive sites for systemic and mucosal immune responses.[6] The LTα1β2 – LT-β receptor signaling pathway is essential for the maintenance, but not the initiation of NALT organogenesis.[7] Signaling via the IL-7/IL-7R and the L-selectin/peripheral lymph node addressin (PNAd) adhesion molecules both play important roles in the organization of NALTs.[8]

Other Sites for Mucosal Induction of an Immune Response

The follicular structures analogous to PPs in the large intestine are known as rectal-associated lymphoid tissues (RALTs). Unlike most other mucosal tissues in humans, the large intestine lamina propria is home to more IgA2- than IgA1-producing cells.[8] Eye drop administration of antigen elicits secretory immunoglobulin A (SIgA) antibody responses in ocular and nasal mucosae. Thus, both tear-duct associated lymphoid tissue (TALT) and conjunctiva-associated lymphoid tissue (CALT) take up antigens for the initiation of mucosal immune responses as a component of MALT.[9] Immunization via the epicutaneous and sublingual routes is emerging as a potential method for induction of mucosal immunity, and structures facilitating these responses are being investigated.

LYMPHOCYTE HOMING INTO MUCOSAL COMPARTMENTS

Mesenteric lymph node cells of orally immunized animals can repopulate the lamina propria of the gut, mammary glands, lacrimal glands, and salivary glands with antigen-specific IgA plasma cells (see Fig. 24.4, B),[1] pointing to the existence of a "common" mucosal immune system. This concept has undergone further refinement, with evidence that migration of cells into and from NALTs follows rules different from those for GALTs and the GI tract.



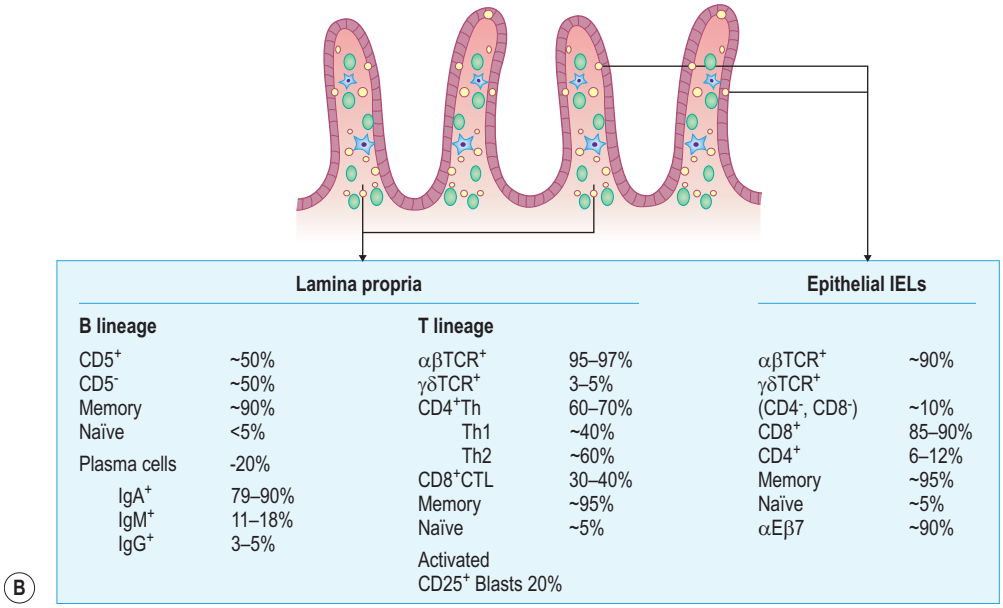


FIG. 24.4 Structural Features and Cellular Components of Gut-Associated Lymphoreticular Tissues (GALTs). (A) The dome region is covered by the follicle-associated epithelium (FAE) with its characteristic microfold (*M*) cells. Major features of the dome include M cells with lymphocyte pockets, scattered plasma cells, and immature dendritic cells (DCs). The B-cell area contains five or more germinal centers with high frequencies of surface immunoglobulin (Ig)A+ B cells. The adjacent T-cell area contains mature interdigitating DCs and precursors of CD4 Th and CD8 CTL. (B) Structural features and cellular characteristics of mucosal effector sites. The lamina propria is equally populated by B1 and B2 cells, both of which differentiate into IgA+ plasma cells. Note that memory B and T lymphocytes are also both present in this compartment. Although intraepithelial lymphocytes (*IELs*) in humans are mainly T-cell receptor (*TCR*) αβ+, significant numbers of TCRγδ+ T cells are also found in this compartment.

TABLE 24.1 Major T-Cell Subpopulations Associated With Murine Peyer Patches

T-Cell Phenotype	Percentage of Total T Cells
$CD3^+\alpha\beta$ T-cell receptor $(TCR)^+$	95–97
$CD3^+\gamma\delta$ TCR^+	3–5
$CD3^+$, $CD4^+$ (precursors of T-helper [Th] cells)	65–70
$CD3^+$, $CD8^+$ (precursors of cytotoxic T lymphocytes [CTLs])	30–35
Naïve ($CD45RB^{Hi}$)	50–60
Memory ($CD45RB^{Lo}$, $CD45RO^{Hi}$)	40–50
Blasts (in cell cycle)	30–35

Lymphocyte Homing in the Gastrointestinal Tract

Naïve lymphocytes enter mucosal or systemic lymphoid tissues from blood through specialized HEVs (Fig. 24.5, A) (Chapter 16). In GALTs, HEVs are present in the interfollicular T-cell zones.[10] In effector sites such as the lamina propria of the gut, the HEVs tend to occur near villi crypts (see Fig. 24.5, B). Mucosal addressin cell adhesion molecule-1 (MAdCAM-1) is the most important addressin expressed by PP HEVs or lamina propria venules (LPVs). PNAd and vascular cell adhesion molecule 1 (VCAM-1) are the principal addressins expressed by peripheral lymph node and skin HEVs, respectively (Chapter 16).

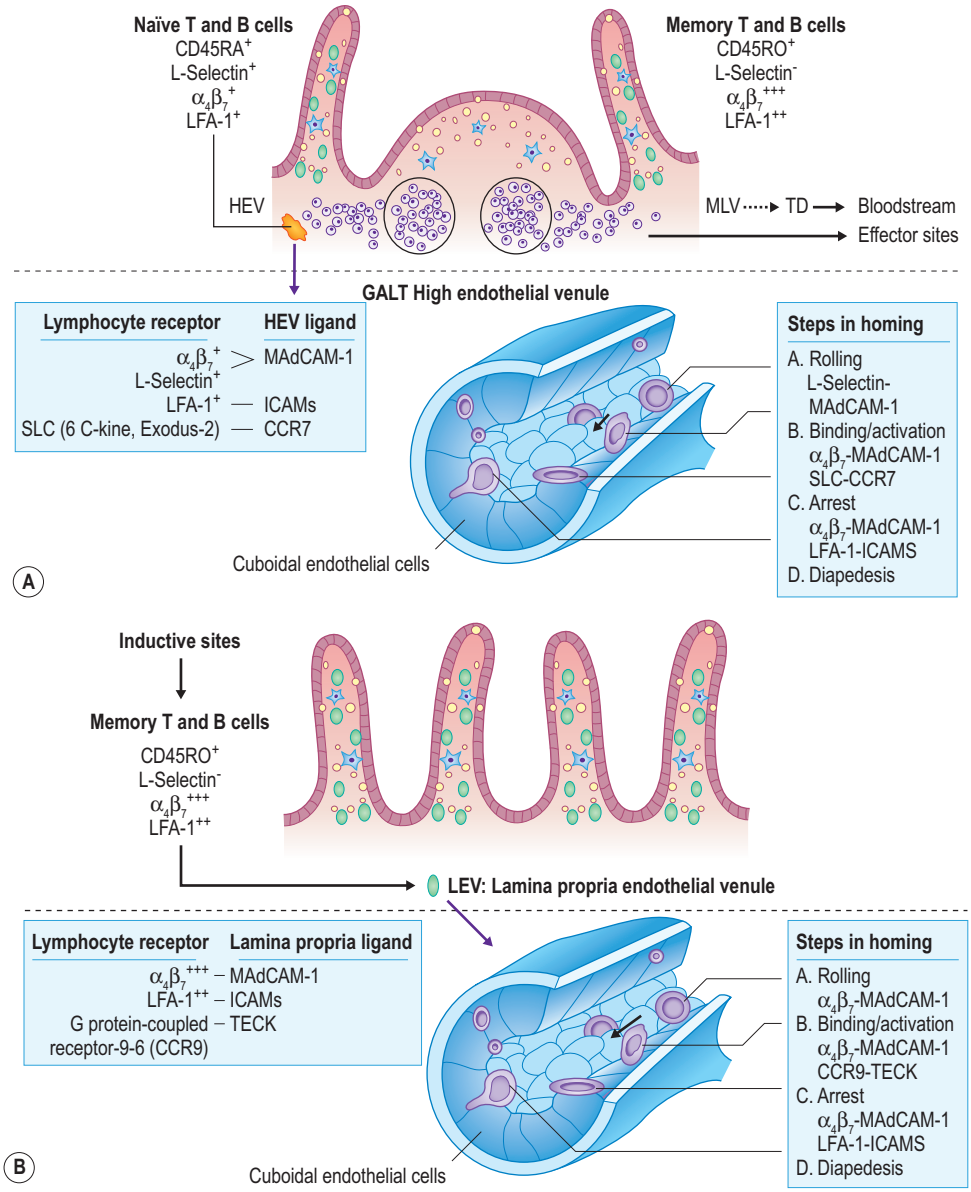


FIG. 24.5 Structural Features and Lymphocyte Homing to Gut-Associated Lymphoreticular Tissues (*GALTs*). (A) High endothelial venules (*HEVs*) occur in T-cell areas and express the ligands mucosal addressin cell adhesion molecule-1 (*MAdCAM-1*), intercellular adhesion molecule 1 (*ICAM-1*), and *CCR7*. Naïve T and B cells, which are L-selectin$^+$, $\alpha_4\beta_{7^+}$, and leukocyte function–associated antigen-1 (*LFA-1*)$^+$, all participate in rolling, binding activation, arrest, and diapedesis in the HEV. Memory B and T cells express $\alpha_4\beta_7$ and LFA-1 at a higher level. (B) Lymphocyte receptors and addressin ligands involved in homing to mucosal effector sites of the gastrointestinal (GI) tract. The majority of B and T cells exhibit a memory phenotype with co-expression of high levels of $\alpha_4\beta_7$ and LFA-1. The expression of the G-protein–coupled receptor *CCR9* allows the homing steps that occur on lamina propria venules.

Integrins are a large class of homing receptors characterized by heterodimeric α and β chains (Chapter 16). In general, the β1 integrin characterizes the homing receptor for skin, whereas the β_7 integrin characterizes the receptor for the gut. The pairing of α_4 with β_7 is thus responsible for lymphocyte binding to MAdCAM-1, which is expressed on HEVs in PPs and gut LPVs (see Fig. 24.5).[10] The C-type lectins—L-, E-, and P-selectins (Chapter 16)—also serve as homing receptors. For example, L-selectin can bind to carbohydrate-decorated MAdCAM-1 and is an important initial receptor for homing into GALT HEVs.

Chemokines (Chapter 15) are also involved in immune-cell homing in mucosal tissues. For example, loss of secondary lymphoid tissue chemokine (SLC) results in lack of naïve T-cell or DC migration into the spleen or PPs. Conversely, memory $\alpha_4\beta_{7^{hi}}$ T cells that express the receptor for thymus-expressed chemokine (TECK), CCR9, migrate into the lamina propria of the GI tract. Both human $\alpha_E\beta_{7^+}$ and $\alpha_4\beta_{7^{hi}}$ CD8 T cells express CCR9, suggesting that TECK-CCR9 is also involved in lymphocyte homing and the arrest of intraepithelial lymphocytes (IELs) in the GI tract epithelium (see Fig. 24.4) (Chapters 15 and 16).

PPs and GALTs contain both naïve and memory T- and B-cell subsets, whereas the lamina propria consists of memory T and B cells and terminally differentiated plasma cells (see Table 24.1 and see Fig. 24.4). Naïve lymphocytes destined for GALTs express L-selectin, moderate levels of $\alpha_4\beta_7$ ($\alpha_4\beta_{7^+}$) and lymphocyte function–associated antigen-1 (LFA-1). Memory lymphocytes destined for lamina propria express higher levels of $\alpha_4\beta_7$ ($\alpha_4\beta_{7^{hi}}$) and lack L-selectin. Initial rolling is dependent on $\alpha_4\beta_7$ interactions with LPV MAdCAM-1. Activation-dependent binding and extravasation require LFA-1–ICAM binding. $\alpha_4\beta_7$ also mediates binding to E-cadherin, and CCR9 expression can result in activation-dependent entry into the epithelial cell compartment.

Cryosections of human tissues have revealed naïve lymphocytes in HEVs that express both L-selectin and $\alpha_4\beta_7$, whereas memory lymphocytes in efferent lymphatics express $\alpha_4\beta_7$, but not L-selectin. The majority of cells in mesenteric lymph nodes, including B-cell blasts, tend to be of the memory phenotype and are $\alpha_4\beta_{7^{hi}}$, L-selectin^{lo}. Ig-containing B-cell blasts also express high levels of $\alpha_4\beta_7$. This separation of naïve and memory T and B cells for entry into GALT HEVs or LPVs has important implications in vaccine development (Chapter 87).

An oral cholera vaccine was reported to elicit transient IgA antibody-forming cells (AFCs) in blood and subsequent IgA anticholera toxin AFCs in duodenal tissues.[11] In a separate study, peripheral blood AFCs induced after parenteral immunization were L-selectin$^+$, whereas those induced after oral and rectal immunization were predominantly $\alpha_4\beta_{7^+}$ AFCs.[11] In the latter study, most of the AFCs produced IgA, but some also expressed IgG. After nasal immunization, AFCs expressed both L-selectin and $\alpha_4\beta_7$ homing receptors. The APCs in GALTs were shown to produce high levels of retinoic acid, which promotes expression of $\alpha_4\beta_7$. Thus, enteric immunization of GALTs more effectively triggers $\alpha_4\beta_7$ memory IgA and IgG B cells, which can then migrate into the bloodstream.

Lymphocyte Homing in Nasopharyngeal-Associated Lymphoid Tissues and Lung-Associated Tissues

Unlike PP HEVs, which are found in T-cell zones, murine NALT HEVs are found in B-cell zones and express PNAd, either alone or associated with MAdCAM-1. Moreover, anti–L-selectin antibodies—but not anti–MAdCAM-1 antibodies—block the binding of naïve lymphocytes to NALT HEVs, suggesting a role for L-selectin and PNAd in the binding of naïve lymphocytes to these HEVs.[12]

During pulmonary immune responses, induction of VCAM-1, E-selectin, and P-selectin in the pulmonary vasculature is matched by increased expression of P-selectin ligand on peripheral blood CD4 and CD8 T cells.[13] As the cells accumulate in the bronchoalveolar lavage (BAL) fluid, the number of cells that express P-selectin ligand in blood declines. Very late antigen-4 (VLA-4) appears to be involved, as migration of VLA-4$^+$ T cells into BAL fluid is impaired following treatment with anti-α_4 antibody. Following systemic immunization, most NALT effector B cells express L-selectin, with only a few cells expressing $\alpha_4\beta_7$.[14] Effector B cells induced by nasal immunization display a more promiscuous pattern of adhesion molecules, with a large majority expressing both L-selectin and $\alpha_4\beta_7$.

The Common Mucosal Immune System Revisited

The homing pattern that has been elucidated in the GI tract after immunization of GALTs has been the model for all mucosal immune sites. As summarized above, the specific set of homing receptors and ligand addressins expressed in the GI tract are absent in NALTs and associated lymph nodes. It remains possible, and even likely, that memory lymphocytes from the gut may enter NALTs for additional priming and reprogramming of homing receptors. Likewise, memory lymphocytes induced in NALTs may traffic to lung and genitourinary tract tissues as well as to the GI tract. Thus, the rules for the homing of naïve lymphocyte precursors to NALTs need to be more clearly defined.

INDUCTION OF MUCOSAL IMMUNITY

Mucosal immune responses are typified by SIgA antibodies, the predominant Ig isotype in external secretions. The resistance of SIgA to endogenous proteases makes them uniquely suited to protect mucosal surfaces. The development of mucosal adaptive immunity requires cytokine signals from CD4, as well as from CD8, T cells, DCs, MØs, B cells, and nonclassic APCs (e.g., epithelial cells). B-cell commitment (Cμ to Cα switching) and the interactions of B and T cells are of central importance for induction of pIgA-producing cells.

Mucosal Antigen-Presenting Cells

Large macromolecules are taken up by M cells in the GI tract. Just beneath the follicle epithelium, N418$^+$, 2A1$^+$, NLDC-145$^-$, and M342$^-$ DCs, form a dense layer of cells in the subepithelial dome (SED) where CD4 T cells can be found.[15] Another subset of DCs, N418$^+$, 2A1$^+$, NLDC-145$^+$, and M342$^+$ DCs populate the interfollicular T-cell regions, where both CD4 and CD8 T cells reside. DCs in the dome region are immature, highly endocytic, and express low levels of major histocompatibility complex (MHC) (Chapter 5) and B7 molecules. DCs in the T-cell area are mature, with low endocytic activity, high levels of MHC class I and II molecules, and B7 molecule expression (see Fig. 24.4, A). DCs are also found in NALTs, where they play essentially the same role as in GALTs.

IECs express MHC class II and class I molecules and present peptides to primed CD4 and CD8 T cells. Human and murine IECs also express CD1d, a nonclassic MHC class I molecule involved in the presentation of lipid and glycolipid antigens (Chapter 5).

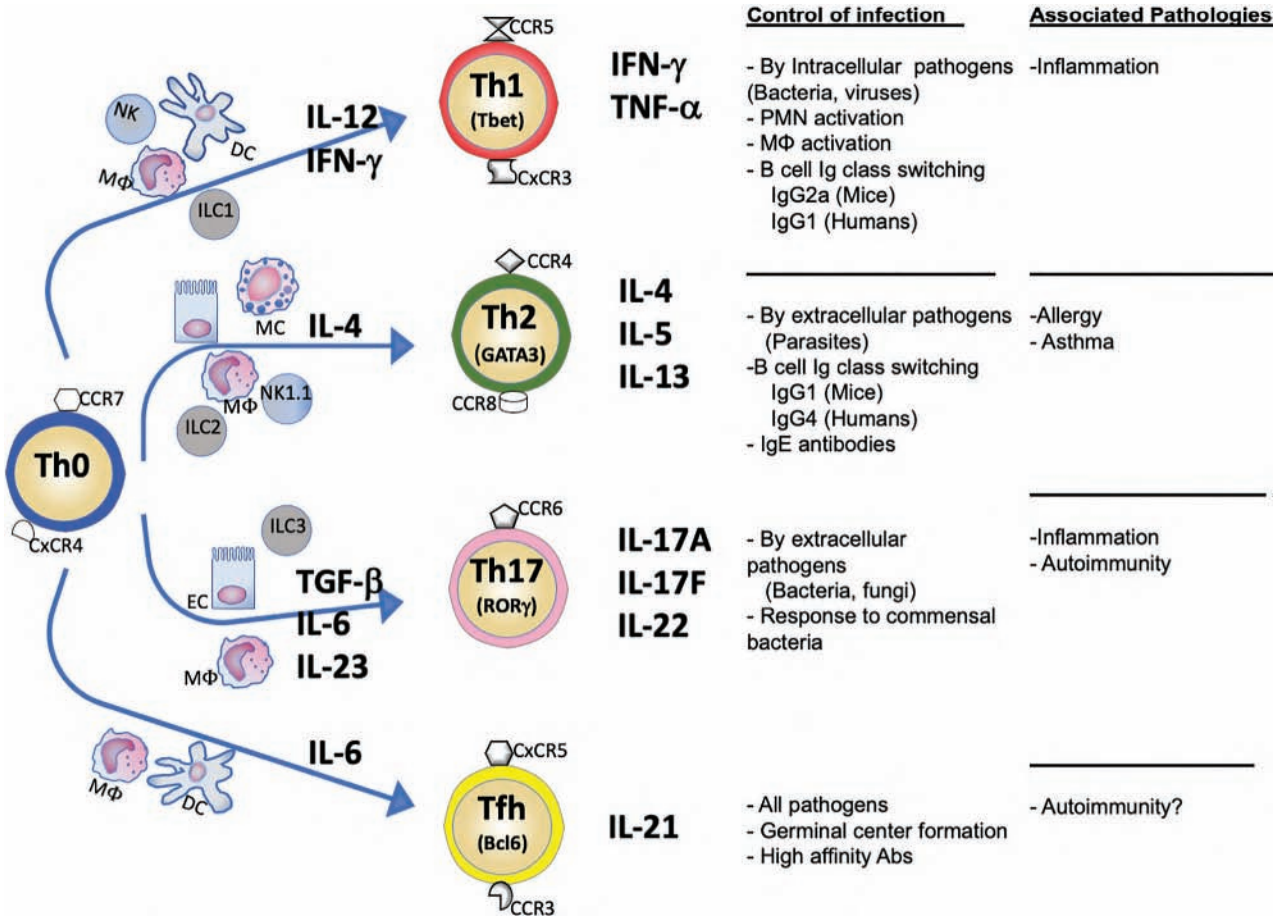


FIG. 24.6 T-Helper (*Th*) Cell Subset Development in Mucosal Tissues. The cellular and cytokine environment induces Th0 cells to develop into Th1, Th2, or Th17 subsets that can be discriminated, based upon their cytokine production. Antigen-presenting cells (APCs) produce IL-12 in response to microbial assault and, together with interferon (*IFN*)-γ produced by natural killer (*NK*) cells and group 1 innate lymphoid cells (*ILC1*), induce mature Th1 cells. Th1 cells express select chemokine receptors and through IFN-γ synthesis, activate macrophages (*MØs*), and induce B cells to produce opsonizing antibodies. Other cells, such as NK1.1, mast cells, and ILC2, respond to parasite/antigen/allergen with interleukin (*IL*)-4 production. IL-4 induces Th0 to Th2 differentiation. Epithelial cells (*ECs*) also produce cytokines that facilitate Th2 cell differentiation. Th2 cells produce IL-4, -5, -6, -9, -10, and -13, which help regulate mucosal secretory immunoglobulin A (*SIgA*) antibody responses. Transforming growth factor (*TGF*)-β, IL-6, and IL-23 produced by epithelial cells, MØs and other cells, help promote the differentiation of Th17 cells. The cytokines produced by Th17 cells contribute to several functions for the host response to commensal bacteria and protection against fungal infections. Follicular T-helper cells (*Tfh*) are a subset of Th cells that help the germinal center formation and the development of high-affinity antibodies (*Abs*).

CD4 T-Helper Cell Subsets in Mucosal Immunity

Th cells are classified as Th1, Th2, Th17, regulatory T cells (Treg), or follicular Th (Tfh) cells, according to the cytokines they produce (Chapter 11). Th1 cells produce IFN-γ, LT-α, LT-β, and TNF, whereas Th2 cells produce IL-4, IL-5, IL-6, IL-9, IL-10, and IL-13 (Fig. 24.6). In mice, mucosal Th1-type responses are associated with cell-mediated immunity and B-cell responses with characteristic IgG2a antibodies. Th2 cells support the production of IgA, as well as IgG1, IgG2b, and IgE. In humans and mice, Th1 and Th2 cells regulate the development of the opposite subset reciprocally through IFN-γ and IL-4 secretion, respectively (see Fig. 24.6). Human Th1 cells and IFN-γ responses are associated with IgG1, IgG3c-fixing antibodies with low IgG2, and undetectable IgG4 antibody levels,[16] suggesting that in humans IL-4 promotes IgG4 and IFN-γ promotes IgG1 (see Fig. 24.6).

Treg and Th17 cells play a role in mucosal homeostasis and inflammatory responses (Chapter 13). Human tonsil CD4 T cells expressing the B-cell follicle homing receptor CXCR5 are identified as Tfh cells that help B-cell differentiation.[17] Foxp3$^+$ Treg in PPs can apparently differentiate into Tfh cells, which express the chemokine CXCR5, the transcription factor Bcl-6, and the cytokine IL-21, to promote germinal center formation and IgA synthesis in the gut.[18]

B-Cell Isotype Switching and Immunoglobulin A Plasma Cell Differentiation

Isotype switching is preceded by transcriptional activation of the isotype in question (Chapter 4). IL-4 and TGF-β induce surface IgM-positive (sIgM$^+$) B cells to switch to IgE and IgA. TGF-β1 can induce sIgM$^+$ to sIgA$^+$ B-cell switches, and addition

of TGF-β1 to LPS-triggered mouse B-cell cultures increases IgA synthesis. In humans, anti-CD40 stimulation of tonsillar B cells, together with TGF-β1 in the presence of IL-10, stimulates IgA synthesis[1]Cα1 transcripts can also be induced by B-cell mitogen plus TGF-β, and Cα2 transcripts can be induced by TGF-β together with IL-10.

DCs can also induce surface IgA$^+$ B cells via direct stimulation of B cells with B-cell activation factor of the TNF family (BAFF) and a proliferation-inducing ligand (APRIL).[19] APRIL–transmembrane activator and CAML interactor (TACI) signaling play a key role in CD40-independent IgA class switching in mice.[19] In humans, functional mutations in TACI can result in IgA deficiency (IgAD; Chapter 33). Differentiation of sIgA$^+$ B cells into IgA-producing plasma cells is dependent on IL-5 and IL-6.[20]

VACCINE DEVELOPMENT AND MUCOSAL IMMUNE RESPONSES

Mucosal SIgA antibodies, as well as Th cell and cytotoxic T lymphocyte (CTL) responses, can be induced by pathogens triggering organized mucosal inductive sites. Effective protection against virulent mucosal pathogens requires prophylactic immune responses that can be achieved through mucosal vaccines. In contrast to conventional injected vaccines, those administered via mucosal routes can stimulate organized mucosal inductive sites and trigger both mucosal immune responses as a first line of defense at the portal of pathogen entry, and systemic immune responses that neutralize pathogens that have penetrated that barrier. Thus, safe adjuvants and delivery systems boosting SIgA antibodies and mucosal immunity are being developed for immunization by needle-free vaccines. These efforts are in large part as a result of knowledge gained from studies of bacterial enterotoxins and nontoxic derivatives (Fig. 24.7).

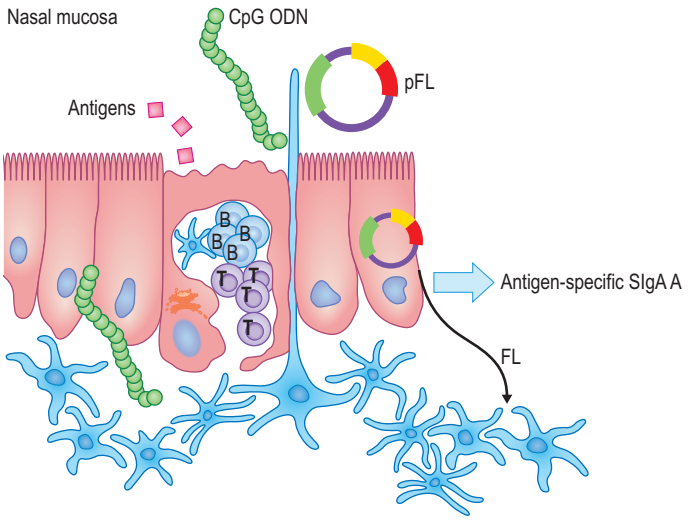


FIG. 24.7 Dendritic cell (DC) targeting with a nasal adjuvant. Nasal administration of CpG oligodeoxynucleotide (*ODN*) and plasmid expressing FLT3 ligand cDNA (*pFL*) specifically target DCs in nasopharyngeal-associated lymphoid tissues (NALTs). These nasal DC-targeting vaccines successfully elicit protective antigen-specific secretory immunoglobulin A antibody (*SIgA Ab*) responses in older adults.

Lessons From Studies of Bacterial Enterotoxins

Earlier studies of cholera toxin (CT) and heat-labile toxin I (LT-I) from *Escherichia coli* helped establish that mucosal (i.e., oral or nasal) administration of vaccines was an effective approach for the induction of both mucosal and systemic immunity, to co-administered vaccine antigens (Chapter 87). These closely related molecules are AB-type toxins consisting of two structurally and functionally separate enzymatic A subunits and binding B subunits (see Fig. 24.7). The B subunit of cholera toxin (CT-B) binds to GM1 gangliosides, whereas the B subunit of heat-labile toxin I (LT-B) binds to GM1 as well as GM2 asialo-GM1 gangliosides. The A subunits of these toxins are adenosine diphosphate (ADP)–ribosyl transferases. Binding of the B subunits to ganglioside receptors on target cells allows the A subunits to reach the cytosol where they elevate cyclic adenosine monophosphate (cAMP) levels.

Cellular Targets of Vaccine Adjuvants Can Shape the Immune Response

Studies with CT and LT-I revealed the importance of the cellular targets for shaping the profile of immune responses induced by mucosal adjuvants. In fact, CT promotes CD4 Th2 and Th17 responses, whereas LT-I also induces a CD4 Th1 (i.e., IFN-γ) response.[21] As discussed below, studies with mutants of these enterotoxins and other toxins (e.g., *Bacillus anthracis* edema toxin[22]) have shown that their enzymatic activities are dispensable for vaccination.

Potential Central Nervous System Targeting Is a Safety Concern with Nasal Vaccines

Cholera induces diarrhea as a result of its ability to elevate cAMP in epithelial cells, thereby promoting secretion of water and chloride ions into the intestinal lumen. Diarrhea is thus the primary limiting factor for the use of oral enterotoxin as an adjuvant in humans. The olfactory neuroepithelium in the nasopharynx constitutes approximately 50% of the nasal surface and has direct neuronal connection to the olfactory bulbs (OBs) in the central nervous system (CNS). Nasally delivered enterotoxins can enter and/or target olfactory neurons and therefore gain access to OBs and deeper structures in the brain parenchyma. These adverse effects are, in large part, mediated by the ADP–ribosyl transferase activity and the nature of the cellular receptors targeted. Both CT and LT-I bind to GM1 on epithelial cells and require endocytosis followed by transport across the epithelial cell to reach the basolateral membrane. GM1 gangliosides are also abundantly expressed by the neuronal and microglial cells of the CNS.[23] CT or CT-B, when administered nasally to mice, enters the olfactory nerves and epithelium (ON/E) and OBs by mechanisms that are selectively dependent on GM1.[24] The targeting of CNS tissues by nasally administered bacterial enterotoxins is clearly related to a higher incidence of Bell palsy (facial paresis) among volunteers of a nasal vaccination trial given LT-I as a mucosal adjuvant. Bell palsy occurring among study subjects who received a non-living nasal influenza vaccine (Nasalflu) led to its withdrawal from the market in the year 2000.

CLINICAL RELEVANCE

Examples of Mucosal Adjuvants and Delivery Systems for the Induction of Targeted Immunity

Genetically Engineered Bacterial Toxins
- Examples are derivatives of the enterotoxin cholera toxin (CT) and heat-labile toxin (LT-I) from *Escherichia coli*.

Nucleic Acid Toll-Like Receptor Ligands
- These sequences typically contain a transcription unit designed to express the antigen in question that is coupled to an adjuvant/mitogen unit, such as CpG motifs.

Mucosal Cytokines and Innate Factors as Adjuvants
- Mucosal delivery of specific cytokines or innate factors can reduce the risk of adverse systemic effects while targeting the immune response to the mucosa.

Transgenic Plants
- Plants, such as potatoes, bananas, lettuce, and rice, can be engineered to express both B- and T-cell antigen epitopes, providing a simple delivery system for oral vaccination or oral tolerance induction.

New Mucosal Adjuvants and Delivery Systems

Nontoxic Derivatives of Bacterial Enterotoxins

To circumvent the toxicity of enterotoxins, mutants of CT (mCT) and LT (mLT) molecules were generated by site-directed mutagenesis in the active site of the A subunit of CT or LT, or in the protease-sensitive loop of LT. These mutants induced comparable levels of antigen-specific serum IgG and IgA antibodies as wild-type CT and significantly higher levels than those induced by recombinant CT-B.[25] One of the mutants also induces Th2-type responses through a preferential inhibition of Th1-type CD4 T cells. mLT molecules, whether possessing a residual ADP–ribosyltransferase activity (e.g., LT-72R) or totally devoid of it (e.g., LT-7 K and LT-6 K3), can also function as mucosal adjuvants for nasal vaccine antigens in mice.[26] As LT induces a mixed CD4 Th1- and Th2-type response,[21] one might envisage the use of mLTs when both Th1- and Th2-type responses are desired.

The use of GM1-receptor binding holotoxins as nasal mucosal adjuvants is currently not recommended because of the risk for their accumulation in the CNS. However, nontoxic mCT could overcome these potential problems. To this end, a model adjuvant has been developed by combining the ADP-ribosylating ability of native CT (nCT) with a dimer of an Ig-binding fragment, D, of *Staphylococcus aureus* protein A.[27] This CTA1-DD molecule directly binds to B cells of all isotypes, but not to MØs or DCs. Despite the lack of a mucosal binding element, the B-cell–targeted CTA1-DD molecule is as strong an adjuvant as nCT. Notably, CTA1-DD promoted a balanced Th1/Th2 response with little effect on IgE antibody production. CTA1-DD did not induce inflammatory changes in the nasal mucosa and, most importantly, did not bind to or accumulate in the OBs or the CNS.[27] CTA1-DD is an example of the use of nonganglioside targeting adjuvants and delivery systems as new tools for the development of safe and effective nasal vaccines.

Nucleic Acid Toll-Like Receptor Ligands

Toll-like receptor 3 (TLR3) and TLR9 recognize the pathogen-associated microbial pattern double-stranded RNA (dsRNA) and unmethylated DNA, respectively (Chapter 3). The latter contains immunostimulatory sequences consisting of short palindromic nucleotides located around a CpG dinucleotide core (e.g., CpG motifs). CpG motifs bind to intracellular TLR9 and induce cytokine secretion (i.e., IL-6, IFN-α, IFN-β, IFN-γ, IL-12, and IL-18) by a variety of immune cells. CpG motifs can enhance both systemic and mucosal immune responses when given nasally to mice.[28] An injection of bacterial DNA or CpG motifs with a DNA vaccine or with a protein antigen promotes Th1-type responses even in mice with preexisting Th2-type immunity. Stimulation of TLR3 by dsRNA results in the production of type I IFNs (i.e., IFN-α/β), which stimulate antibody responses to injected vaccines.[29] The synthetic TLR3 ligand polyinosinic–polycytidylic acid (poly I:C) has been shown to enhance responses of CD8 T cells to an experimental nasal influenza vaccine in mice and promote heterosubtypic protection via stimulation of TLR3 signaling by nonhematopoietic radioresistant cells.[30]

Cyclic dinucleotides that bind stimulator of interferon gamma genes (STING) were recently shown to be potential alternatives to cAMP-inducing bacterial toxins, and derivatives as vaccine adjuvants for induction of mucosal immunity. For example, STING ligands of bacterial origin, including 3′3′-cGAMP, c-di-AMP, and c-di-GMP, have been shown to effectively elicit mucosal and systemic immune responses following nasal or sublingual immunization. Like the enterotoxins CT and LT, the STING ligands stimulate Th17 responses, an important observation since Th17 cells are believed to be crucial for production of high-affinity T-dependent IgA.[31]

Mucosal Cytokines and Innate Factors as Adjuvants

Mucosal delivery of cytokines offers a means to prevent the adverse effects associated with the large and repeated parenteral doses often required for the effective targeting of tissues and organs. For example, nasal delivery permits acquisition of significant serum levels of IL-12 at one-tenth the dose required for inhibition of serum IFN-γ by parenteral administration.[1] Earlier studies have shown that nasal administration of tetanus toxoid with IL-12 as adjuvant induced high titers of SIgA antibody responses in the GI tract, vaginal washes, and saliva.[32] Similar results were reported when mice were nasally immunized with soluble influenza H1 and N1 proteins and IL-12. Related studies showed that mucosally administered IL-12 can redirect antigen-specific Th2-type responses toward the Th1 type or promote mixed Th1- and Th2-type responses, depending on the mucosal route and timing of delivery.[1]

FMS-like tyrosine kinase 3 ligand (FL) binds to the FMS-like tyrosine kinase receptor Flt3/Flk2. FL mobilizes and stimulates myeloid and lymphoid progenitor cells, DCs, and NK cells. Although FL dramatically augments numbers of DCs in vivo, it fails to induce their activation. Treatment of mice by systemic FL injection can induce marked increases in the numbers of DCs in both systemic (i.e., spleen) and mucosal lymphoid tissues (i.e., iLP, PPs, and mesenteric lymph nodes). Although this increase in mucosal DCs can, in some cases, initially enhance induction of oral tolerance, it favors the induction of immune responses by mucosal or systemic vaccines. Nasal administration of plasmid or adenovirus encoding FL cDNA (pFL or Ad-FL) with protein antigens was shown to induce antigen-specific secretory IgA (SIgA) and protective immunity.[28] Thus FL cDNA may be an alternative to costly treatments with FL protein.

Transgenic Plants

Edible plants have been engineered to synthesize and assemble one or more antigens that retain both T- and B-cell epitopes, thereby inducing systemic and mucosal immune responses in both mice and humans.[33,34] To circumvent potential denaturation of the plant antigen during cooking, recombinant bananas that can accumulate up to 1 mg of vaccine antigen per 10 g of banana were developed. Most recently, the CT-B subunit has been expressed under the control of the rice seed storage protein glutelin promoter (MucoRice-CT-B). Oral feeding of powdered MucoRice-CT-B to mice and nonhuman primates resulted in the induction of both systemic and mucosal antibody responses for protection against CT.[35] Unlike plant vaccines expressed in banana, tomato, or lettuce, MucoRice vaccines were shown to be stable for years at room temperature.

ON THE HORIZON

Development of Transgenic Plants as Vehicle for Vaccine Administration

- The MucoRice system is a novel strategy for vaccine development.
- The MucoRice system may also be used as a passive neutralizing antibody delivery system.

SYNTHESIS AND FUNCTIONS OF SECRETORY ANTIBODIES

Mucosal SIgA differs from serum IgA in both molecular composition and specific antibody activity. Humans possess two Cα gene segments, Cα1 and Cα2 (Chapter 4), the use of which defines the two IgA subclasses, IgA1 and IgA2.[8] These IgA subtypes differ primarily in their hinge regions (Chapter 8). IgA1 antibodies contain an additional 13 amino acids in the hinge region, and this renders them more flexible and susceptible to IgA1-specific proteases produced by certain bacteria. IgA1-secreting cells are prevalent in most human mucosal tissues, especially the small intestine and the respiratory tract, whereas the human colon and genital tract are enriched by IgA2-secreting cells. SIgA is mostly viewed as a barrier at mucosal surfaces to regulate interactions with commensals and prevent adhesion and colonization of pathogens, as well as an effective means to neutralize viruses and toxins. Nonetheless, these antibodies confer the additional advantage of having anti-inflammatory properties.[8]

In external secretions, adult levels of SIgA are reached considerably earlier (1 month to 2 years) than in the serum (adolescence). Approximately 98% of SIgA antibodies are produced locally in mucosal tissues, with only a minor fraction deriving from the circulation.Polymeric Immunoglobulin Receptor and plgA Transport

The polymeric Ig receptor (pIgR) is synthesized as a transmembrane protein by epithelial cells and is found on the basolateral surface of epithelial cells. It acts as a receptor for the endocytosis of pIgA(dimeric) and pentameric IgM, both of which contain a J-chain. The pIgR is produced by bronchial epithelial cells, renal tubules, glands, and the epithelia of the small and large intestines.[1] The pIgR is not expressed by the FAE (including M cells) of PPs, but only by the adjacent columnar epithelial cells. Furthermore, pIgR is expressed in the upper respiratory tract, which includes the nasal cavity, tonsils, trachea, bronchi, and tracheobronchial glands. Expression in the lower lungs is restricted to the pulmonary alveolar cells.

In female reproductive tissues, the expression of pIgR is influenced by the sex hormones. It is low in the vagina, absent in the ovary and myometrium, and very high in the fallopian tubes and uterus. Normal kidneys do not express pIgR, whereas epithelial cells in the lower urinary tract may normally express pIgR and transport pIgA into urine. The expression of pIgR can be upregulated by several cytokines, such as IFN-γ, TNF, IL-1α, IL-1β, and TGF-β.

KEY CONCEPTS

Secretory Immunoglobulin A

- Unlike serum immunoglobulin A (IgA), mucosal secretion of IgA reaches adult levels early in life (1 month to 2 years after birth).
- The polymeric Ig receptor (pIgR) is expressed on the basolateral surface of epithelial cells and facilitates the active transport of polymeric (mainly dimeric) IgA, as well as pentameric IgM into mucosal secretions.
- SIgA protects the host by inhibiting microbial adherence, neutralizing viruses, enzymes, and toxins, and engaging in anti-inflammatory activities by means of inhibiting IgM and IgG complement activation.
- Although selective IgA deficiency could occur, the most common is primary immune deficiency; normally, it is clinically inconsequential. Some affected subjects develop recurrent mucosal infections, including sinusitis, otitis media, bronchitis, and pneumonias of viral or bacterial origin, as well as acute diarrhea caused by viruses, bacteria, or parasites such as *Giardia lamblia*.

IgA-Mediated Inhibition of Microbial Adherence

The inhibition of microbial adherence plays a critical initial role in the protection of the host. This inhibition is mediated by both specific and nonspecific mechanisms. The surface of microorganisms interacting with SIgA becomes less hydrophobic and thus more likely to be entrapped in mucus. A series of recent studies have revealed that binding of SIgA to intestinal microbes is more sophisticated than originally envisioned. Thus, individual SIgA can specifically bind to a broad, but well-defined subset of bacteria.[36] Furthermore, while most commensals are dually targeted by SIgA1 and SIgA2 in the small intestine of humans, the situation appears to be different in the large intestine where some bacteria are either preferentially or exclusively recognized by IgA2. Finally, SIgA antibodies are more effective at agglutinating microorganisms than membrane-bound IgA, and the agglutinating ability of SIgA specific for capsular polysaccharides of *Haemophilus influenzae* appears to be crucial to preventing colonization by *H. influenzae*.

Neutralization by SIgA of Viruses, Enzymes, and Toxins

SIgA antibodies have been shown to be effective at neutralizing viruses in several experimental systems (e.g., influenza virus, Epstein-Barr virus [EBV], HIV) and at different steps in the infectious process. SIgA specific for influenza hemagglutinin can interfere with the initial binding of influenza virus to target cells or with the internalization and the intracellular replication of the virus. In vitro experiments employing polarized murine epithelial cells have demonstrated that antibodies specific to rotavirus and hepatitis virus can neutralize the viruses inside epithelial cells. Finally, SIgA can neutralize the catalytic activity of many enzymes of microbial origin.

Anti-Inflammatory Actions Mediated by SIgA Antibodies

IgA antibodies are unable to activate complement by either the classical or the alternative pathway (Chapter 40). Nevertheless, they can interfere with IgM- and IgG-mediated complement activation.[8] SIgA can inhibit phagocytosis, bactericidal activity, and chemotaxis by polymorphonuclear neutrophils (PMNs), monocytes, and MØs. IgA can downregulate the synthesis of TNF and IL-6, as well as enhancing the production of IL-1R antagonists by LPS-activated human monocytes. Thus, the anti-inflammatory properties of IgA are of significant importance for the integrity of the mucosa in that IgA can limit bystander tissue damage that may result from the continuous interactions of the mucosa with myriad dietary and environmental antigens. Systemically, circulating IgA also appears to help limit inflammatory reactions that result from complement fixation and phagocyte activation, and it contributes to the inhibition of IgE-dependent anaphylactic responses.

IgA Deficiency

Selective IgAD is the most common primary immune deficiency (PID) in individuals of European descent (Chapter 33). The clinical diagnosis of IgAD depends on the relative absence of IgA in the serum. However, the most important manifestations of the disorder primarily reflect the absence of both SIgA1 and SIgA2 in the external secretions. Thus IgAD affects both the mucosal and systemic immune compartments, with only rare individuals exhibiting a superselective loss of either IgA1 or IgA2 alone.

MUCOSAL CYTOTOXIC T LYMPHOCYTES

M cells have specific receptors for mucosal virus that allow certain viruses, such as reoviruses, to enter the cells in both NALTs and GALTs. It is likely that enteric viruses, such as rotavirus, and respiratory pathogens, such as influenza virus and respiratory syncytial virus (RSV), also enter the mucosal inductive pathway via M cells.[37] After enteric infection or immunization, antigen-stimulated CTLs are disseminated from PPs into mesenteric lymph nodes via the lymphatic drainage. Oral immunization with live virus can thus induce antigen-specific CTLs in both mucosal and systemic lymphoid tissues.

KEY CONCEPTS

Mucosal Cytotoxic T Lymphocytes

- After enteric infection or immunization, antigen-stimulated cytotoxic T lymphocytes (CTLs) are disseminated from Peyer patches into mesenteric lymph nodes via the lymphatic drainage.
- Oral immunization with a live virus can induce antigen-specific CTLs in both mucosal inductive and effector tissues for mucosal immune responses, as well as in systemic lymphoid tissues for serum immune responses.

Enteric Viruses and Mucosal Cytotoxic T Lymphocytes

CD8 CTLs (Chapter 12) play a central role in rotavirus and reovirus immunity. Reovirus-induced CTL precursors (pCTLs) in GALTs migrate to the systemic compartment. Reovirus-specific CD8 CTLs associated with the αβ T-cell population are also observed in intraepithelial T lymphocytes. Oral delivery of rotavirus increases pCTLs in GALTs and results in their dissemination throughout the murine lymphoid system within 3 weeks. Moreover, adoptively transferred CD8 T cells mediate the clearance of rotavirus infection in severe combined immunodeficiency mice.

Respiratory Viruses and Mucosal Cytotoxic T Lymphocytes

Studies of immune responses after nasal infection with influenza virus in CD4-coreceptor knockouts, or other mice in which this subset had been depleted, have shown that CD4 T cells do not affect the induction of pCTLs or significantly alter clearance of infection. Clearance of influenza is unaltered by the use of β2-microglobulin knockout mice, which lack CD8 T cells, or of mice that have been treated with monoclonal anti-CD8. γδ T cells with several Vδ chain specificities increase in the infected site as clearance occurs, which suggests a regulatory role for γδ T cells in antiviral immunity.

Mucosal AIDS Models for Cytotoxic T-Lymphocyte Responses

Approximately 80% of new HIV-1 infections result from sexual transmission (Chapter 41). Studies using the rhesus macaque and the simian immunodeficiency virus (SIV) vaginal infection model have provided evidence that pCTLs occur in female macaque reproductive tissues and that infection with SIV induces CTL responses. This important finding was extended to vaginal infection with an SIV/HIV-1 chimeric virus (SHIV) containing the HIV-1 89.6 *env* gene. Other studies have shown that nasal immunization with SIV/HIV components induces antibody responses in vaginal secretions.

Other Mucosal Cytotoxic T-Lymphocyte Systems

Salmonella can elicit CD8 T-cell responses, including CTLs, to expressed proteins, and CD8 T cells induced by the parasite *Toxoplasma gondii* have been shown to be protective. Thus, mucosal CD8 CTLs can also be induced in nonviral situations. Significant questions remain as to the mechanism by which naïve CD8 T cells can be triggered to expand into pCTLs and to the rules for expression of effector CTLs and memory in the actual mucosal compartment that manifests the infection. pCTLs accumulate in immunologically privileged sites, but they do not develop a cytotoxic function until they encounter infected class I MHC-presenting target cells. It is possible that this mechanism protects the common mucosal immune system network from inadvertent cytotoxic inflammatory events.

MUCOSAL IMMUNE RESPONSES IN EARLY LIFE AND AGING

Although most of its structures are present at birth, the mucosal immune system requires further postnatal development and maturation before becoming fully functional (Chapter 21). GALTs, NALTs, and tonsils are present in humans at birth. The bacterial colonization after birth increases the number of immune cells and germinal centers in these sites and the number of secondary lymph nodes (e.g., mesenteric lymph nodes and cervical lymph nodes; Chapter 2) and generates innate lymphoid follicles (ILFs). Significant changes that occur after bacterial colonization include an increase in SIgA levels and in the numbers of IgA-secreting cells, Treg, and Th17 cells. BALT

only develops after birth. It is clear that the immature mucosal immune system in early life cannot protect against infectious pathogens entering mucosal surfaces. This gap is filled by maternal antibodies, which are acquired either before birth through the placenta or after birth via ingestion of milk.

KEY CONCEPTS

Mucosal Immunosenescence

- Early mucosal aging is evident in the gastrointestinal (GI) tract immune system.
- Nasal immunization is an effective route for the induction of mucosal and systemic immune responses in aging mice.
- Dendritic cell (DC)–targeting mucosal adjuvants are able to elicit protective pathogen-specific secretory immunoglobulin A (SIgA) antibody responses in aged mice.

Immune functions are known to deteriorate as a result of aging in several species (Chapter 21). The risk and severity of infections are higher, and the susceptibility to certain types of autoimmune diseases and cancer are greater in older adults, and responses to vaccination are diminished. Aging-associated alterations of the systemic immune compartments have been studied extensively. Dysfunctions occur in both B and T cells, although the latter are considered more susceptible to immunosenescence.[38]

In humans, older subjects were reported to have significantly higher concentrations of salivary SIgA antibodies compared with younger subjects, whereas whole gut lavages of aged and young subjects contain similar amounts of antibodies.[28] Analogous results have also been obtained for total IgA antibody responses in the serum of aged animals and humans. These results indicate an absence of aging-associated impairment in total IgA antibody levels in external secretions.

The GI tract in older adults is particularly susceptible to infectious diseases. Antigen-specific mucosal IgA antibody responses are diminished in aged animals, especially those in GALTs.[28] In older humans, pathogens that invade through mucosal surfaces of the airways, such as influenza virus, SARS-Cov-2, and the bacterial pathogen *Streptococcus pneumoniae*, cause more severe and more frequently lethal infections. The development of effective vaccines for older adults remains a largely unmet goal. To provide effective protection against influenza, *S. pneumoniae* and COVID-19 for this population, one should strongly consider developing a new generation of vaccines that could induce pathogen-specific immunity in the respiratory tract.[28] Although it has been shown that effective protection can be provided by pathogen-specific systemic IgG without mucosal IgA responses, pathogen-specific SIgA responses are a necessary component for providing a first line of effective immunity against these respiratory pathogens at their entry site.

REFERENCES

1. Kiyono H, Kunisawa J, McGhee JR, Mestecky J. The mucosal immune system. In: Paul WE, ed. *Fundamental Immunology.* 5th ed. Philadelphia, PA: Lippincott Williams & Wilkins; 2008:983–1030.
2. Clevers HC, Bevins CL. Paneth cells: maestros of the small intestinal crypts. *Annu Rev Physiol*. 2013;75:289–311.
3. Collonna M, Fuchs A, Cella M. Innate lymphoid cells in mucosal homeostasis, infections, autoimmune disorders, and tumors. In: Mestecky J, Strober W, Russell MW, eds. *Mucosal Immunology*. 4th ed. Waltham, MA: Elsevier; 2015:1003–1012.
4. Satoh-Takayama N, Kato T, Motomura Y, et al. Bacteria-induced group 2 innate lymphoid cells in the stomach provide immune protection through induction of IgA. *Immunity.* 2020;52(4):635–649. e4.
5. Williams IR, Owen RL. M cells: Specialized antigen sampling cells in the follicle-associated epithelium. In: Mestecky J, Strober W, Russell MW, eds. *Mucosal Immunology*. 4th ed. Waltham, MA: Elsevier; 2015:211–230.
6. Brandtzaeg P. Potential of nasopharynx-associated lymphoid tissue for vaccine responses in the airways. *Am J Respir Crit Care Med*. 2011;183(12):1595–1604.
7. Kiyono H, Fukuyama S. NALT- versus Peyer's-patch-mediated mucosal immunity. *Nat Rev Immunol*. 2004;4(9):699–710.
8. Woof JM, Mestecky J. Mucosal immunogloblins. In: Mestecky J, Strober W, Russell MW, eds. *Mucosal Immunology*. 4th ed. Waltham, MA: Elsevier; 2015:287–324.
9. Nagatake T, Fukuyama S, Kim DY, et al. Id2-, RORgammat-, and LTbetaR-independent initiation of lymphoid organogenesis in ocular immunity. *J Exp Med*. 2009;206(11):2351–2364.
10. Habtezion A, Nguyen LP, Hadeiba H, Butcher EC. Leukocyte trafficking to the small intestine and colon. *Gastroenterology.* 2016;150(2):340–354.
11. Brandtzaeg P. The mucosal B cell system. In: Mestecky J, Strober W, Russell MW, eds. *Mucosal Immunology*. Waltham, MA: Elsevier; 2015:623–681.
12. Csencsits KL, Jutila MA, Pascual DW. Nasal-associated lymphoid tissue: phenotypic and functional evidence for the primary role of peripheral node addressin in naïve lymphocyte adhesion to high endothelial venules in a mucosal site. *J Immunol.* 1999;163(3):1382–1389.
13. Wolber FM, Curtis JL, Milik AM, et al. Lymphocyte recruitment and the kinetics of adhesion receptor expression during the pulmonary immune response to particulate antigen. *Am J Pathol.* 1997;151(6):1715–1727.
14. Quiding-Jarbrink M, Nordstrom I, Granstrom G, et al. Differential expression of tissue-specific adhesion molecules on human circulating antibody-forming cells after systemic, enteric, and nasal immunizations. A molecular basis for the compartmentalization of effector B cell responses. *J Clin Invest*. 1997;99(6):1281–1286.
15. Lambrecht BN, Iwasaki A, Kelsall BL. Mucosal dendritic cells: origins, subsets, and biology. In: Mestecky J, Strober W, Russell MW, eds. *Mucosal Immunology*. 4th ed. Waltham, MA: Elsevier; 2015:489–541.
16. Widhe M, Ekerfelt C, Forsberg P, et al. IgG subclasses in Lyme borreliosis: a study of specific IgG subclass distribution in an interferon-gamma-predominated disease. *Scand J Immunol*. 1998;47(6):575–581.
17. Moser B. CXCR5, the defining marker for follicular B helper T (TFH) cells. *Front Immunol*. 2015;6:296.
18. Tsuji M, Komatsu N, Kawamoto S, et al. Preferential generation of follicular B helper T cells from Foxp3+ T cells in gut Peyer's patches. *Science.* 2009;323(5920):1488–1492.
19. Cerutti A, Chen K, Chorny A. Immunoglobulin responses at the mucosal interface. *Annu Rev Immunol*. 2011;29:273–293.
20. Fujihashi K, McGhee JR, Lue C, et al. Human appendix B cells naturally express receptors for and respond to interleukin 6 with selective IgA1 and IgA2 synthesis. *J Clin Invest*. 1991;88(1):248–252.
21. Fujihashi K, McGhee JR. Th1/Th2/Th3 cells for regulation of mucosal immunity, tolerance and inflammation. In: Mestecky J, Lamm ME, Strober W, eds. *Mucosal Immunology*. 3rd ed. San Diego, CA: Elsevier Inc.; 2005:539–558.
22. Duverger A, Carre JM, Jee J, et al. Contributions of edema factor and protective antigen to the induction of protective immunity by *Bacillus anthracis* edema toxin as an intranasal adjuvant. *J Immunol*. 2010;185(10):5943–5952.
23. Mancini P, Santi PA. Localization of the GM1 ganglioside in the vestibular system using cholera toxin. *Hear Res*. 1993;64(2):151–165.
24. van Ginkel FW, Jackson RJ, Yuki Y, McGhee JR. Cutting edge: the mucosal adjuvant cholera toxin redirects vaccine proteins into olfactory tissues. *J Immunol.* 2000;165(9):4778–4782.
25. Yamamoto S, Kiyono H, Yamamoto M, et al. A nontoxic mutant of cholera toxin elicits Th2-type responses for enhanced mucosal immunity. *Proc Natl Acad Sci (USA)*. 1997;94(10):5267–5272.

26. Rappuoli R, Pizza M, Douce G, Dougan G. Structure and mucosal adjuvanticity of cholera and *Escherichia coli* heat-labile enterotoxins. *Immunol Today*. 1999;20(11):493–500.
27. Lycke N, Bemark M. Mucosal adjuvants and long-term memory development with special focus on CTA1-DD and other ADP-ribosylating toxins. *Mucosal Immunol*. 2010;3(6):556–566.
28. Fujihashi K. Mucosal vaccines for aged: challenges and straggles in immunosenescence. In: Kiyono H, Pascual DW, eds. *Mucosal Vaccines: Innovation for Preventing Infectious Diseases*. 2nd ed London, UK: Elsevier; 2020:789–808.
29. Le Bon A, Schiavoni G, D'Agostino G, et al. Type I interferons potently enhance humoral immunity and can promote isotype switching by stimulating dendritic cells *in vivo*. *Immunity.* 2001;14(4):461–470.
30. Perez-Giron JV, Belicha-Villanueva A, Hassan E, et al. Mucosal polyinosinic-polycytidylic acid improves protection elicited by replicating influenza vaccines via enhanced dendritic cell function and T cell immunity. *J Immunol*. 2014;193(3):1324–1332.
31. Hirota K, Turner JE, Villa M, et al. Plasticity of Th17 cells in Peyer's patches is responsible for the induction of T cell-dependent IgA responses. *Nat Immunol*. 2013;14(4):372–379.
32. Boyaka PN, Marinaro M, Jackson RJ, et al. IL-12 is an effective adjuvant for induction of mucosal immunity. *J Immunol*. 1999;162(1):122–128.
33. Haq TA, Mason HS, Clements JD, Arntzen CJ. Oral immunization with a recombinant bacterial antigen produced in transgenic plants. *Science.* 1995;268(5211):714–716.
34. Tacket CO, Mason HS, Losonsky G, et al. Immunogenicity in humans of a recombinant bacterial antigen delivered in a transgenic potato. *Nat Med*. 1998;4(5):607–609.
35. Nochi T, Yuki Y, Katakai Y, et al. A rice-based oral cholera vaccine induces macaque-specific systemic neutralizing antibodies but does not influence pre-existing intestinal immunity. *J Immunol*. 2009;183(10):6538–6544.
36. Pabst O, Slack E. IgA and the intestinal microbiota: the importance of being specific. *Mucosal Immunol*. 2020;13(1):12–21.
37. Williams IR, Owen RL. M cells: Specialized antigen sampling cells in the follicle-associated epithelium. In: Mestecky J, Strober W, Russell MW, eds. *Mucosal Immunology*. 4th ed. Waltham, MA: Elsevier; 2015:211–230.
38. Goronzy JJ, Weyand CM. Understanding immunosenescence to improve responses to vaccines. *Nat Immunol*. 2013;14(5):428–436.

25

Host Defenses to Viruses

Scott N. Mueller and Barry T. Rouse

Viruses as obligate intracellular parasites require their host to replicate them and to facilitate their spread to others. In humans, most clinically relevant infections were derived from other animals, and this process continues. Recent examples include human immunodeficiency virus (HIV), Ebola virus, Zika virus, and severe acute respiratory syndrome (SARS) virus, including the novel coronavirus (CoV), SARS-CoV-2, that is responsible for the ongoing COVID-19 pandemic. In contrast to these examples, viral infections are rarely lethal, even if they are highly cytolytic to individual cells. Mortality commonly occurs when viruses jump species (SARS-related coronaviruses are believed to have originated in horseshoe bats in China), when the virus undergoes a major antigenic change (i.e., influenza viruses), or when host immunity is compromised. HIV (see Chapter 41) represents one of the more dramatic human examples of an exotic virus that kills its host. However, HIV kills slowly, providing ample time to spread to new hosts and an effective strategy for persistence in the species. Death or dire consequences following virus infection in mammals with inadequate immunity are well illustrated by observations that fetuses or neonates, especially if deprived of passive immunity, succumb to many agents well tolerated by healthy adults. The science of viral immunology seeks to understand mechanisms of virus-host interactions with a view to applying this knowledge to the design of effective vaccines and immunomodulators that control virus infections. These objectives are facilitated by an increasing wealth of immunological techniques, an expanding array of genetically manipulated animal models, and an abundance of high-throughput technologies, which generate data that can be subjected to complex computational analysis. Such analyses can yield signatures indicative of optimal immunogenicity and vaccine efficacy or failure and can explain the variable outcome of infections in individual hosts. In most situations, defense against viruses involves multiple immune components, and the impact of a single mechanism varies greatly according to the method by which individual viruses enter, replicate, and spread within the host. In this chapter, we highlight the principal means by which the host achieves immunity after infection by viruses. Table 25.1 presents an overview.

VIRAL ENTRY AND INFECTION

Access to target tissues presents numerous obstacles for entry and infection by most human viruses. Most effective of these are the mechanical barriers provided by skin and the mucosal surfaces, as well as the chemically hostile environment of the gut (Fig. 25.1). A number of common human viral pathogens enter through the gastrointestinal tract, including rotavirus, enteric adenoviruses, and hepatitis A virus (HAV). These are usually spread via person-to-person contact or contaminated food and water. Respiratory infections caused by influenza viruses, rhinoviruses, coronaviruses (including the SARS coronaviruses), measles virus, varicella-zoster virus (VZV), and respiratory syncytial virus (RSV) are often spread by aerosol transmission, as well as person-to-person contact. Many of the herpes viruses target the skin or the mucosae, such as herpes simplex virus (HSV) and VZV. HSV, in particular, can infect the oral and genital mucosae, the eye, and skin through small cuts and abrasions. Other herpes viruses, such as Epstein-Barr virus (EBV) and cytomegalovirus (CMV), target mucosae. CMV can also spread vertically from mother to baby or rarely via blood transfusions. Human papillomavirus (HPV) targets skin and mucosae and causes warts and may transform cells, inducing cancers, such as cervical cancer. Some viruses, such as West Nile virus, sengue virus, Semliki Forest virus, and Zika virus, can enter through the skin via insect vectors. HIV and hepatitis B virus (HBV) are commonly spread via sexual contact. HIV, HBV, and hepatitis C virus (HCV) can also infect humans by direct entry into the bloodstream via transfusions or contaminated needles.

Most human viruses replicate only in certain target tissues, this being mainly the consequence of viral receptor distribution. Many viruses use two receptors, such as the use of the CD4 coreceptor and the chemokine receptor CCR5 on T cells by HIV. After attachment to a cellular receptor, viruses may fuse with the cell membrane or be endocytosed and then gain entry into the cytoplasm or nucleus by fusing with the vesicular membrane (enveloped viruses, such as HSV and HIV), or translocate across the cell membrane or induce lysis of the endocytic vesicle once in the cytoplasm (nonenveloped viruses, such as Norwalk virus and poliovirus).[1] Viruses then utilize host cell machinery and specialized virally encoded proteins to replicate rapidly within the cell. Once they have multiplied within the cell, many viruses induce cytolysis to facilitate the release of new infectious virions (e.g., poxviruses, poliovirus, and herpes viruses). Other viruses are released from infected cells by budding through the cell membrane in the absence of cell death (e.g., HIV and influenza virus). Having entered the body, however, viruses encounter numerous innate defenses and activate the components of adaptive immunity. The latter usually assures that clinical disease, if not infection, will not become evident. Successful exploitation of these defenses through the use of vaccines (see Chapter 87) remains a central challenge for many human viruses, particularly those that cause chronic infections, such as HIV and HCV.[2]

TABLE 25.1 Viral Infections and Immunity

Viral Event	Obstacles	Time Course
Transmission	Mechanical and chemical barriers	0
Infection and replication	Innate immunity	0 →
Infection stopped or spreads	Viral antigens transported to lymphoid tissues	Within 24 h
Infection controlled	Specific antibodies and cell-mediated immunity	4–10 days
Sterile immunity	Immune memory	14 days to years
Viral persistence if infection not controlled	Immune disruption or evasion	Weeks to years

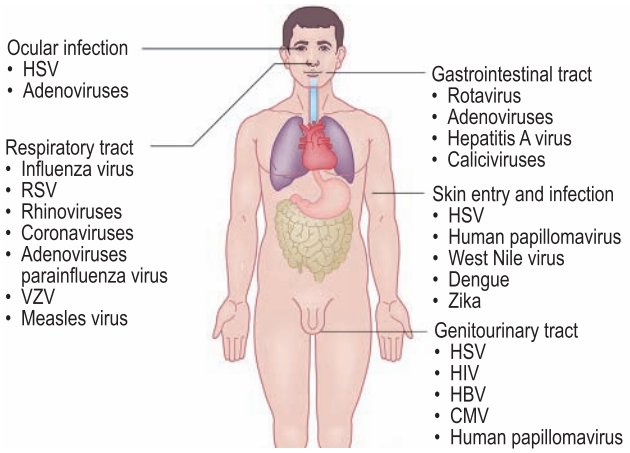


FIG. 25.1 Common Routes of Entry and Infection for Human Viral Pathogens. *CMV*, Cytomegalovirus; *HBV*, hepatitis B virus; *HIV*, human immunodeficiency virus; *HSV*, herpes simplex virus; *RSV*, respiratory syncytial virus; *VZV*, varicella-zoster virus.

TABLE 25.2 Sensors of Viral Infection

Toll-Like Receptors (TLRs)	
TLR3	dsRNA, MCMV, VSV, LCMV, HSV, EBV
TLR7 and TLR8	ssRNA, influenza virus, HIV, VSV
TLR9	dsDNA, HSV, MCMV
TLR2	MV hemagglutinin protein, HSV, HCMV
TLR4	MMTV envelope protein, RSV
RIG-I–Like Helicases (RLHs)	
RIG-I	Influenza virus, VSV, HCV, JEV, MV, RSV, Sendai virus, EBV
MDA-5	Poly(I:C), MV, Sendai virus, VSV, MCMV, picornaviruses
NOD-Like Receptors (NLRs)	
NLRP3	Influenza virus, Sendai virus, adenovirus, vaccinia virus
NOD2	Influenza virus, VSV, RSV
Other Sensors	
AIM2	Vaccinia virus, MCMV
ZBP1 (DAI)	Cytosolic dsDNA, HSV
IFI16	Cytosolic dsDNA, HSV
cGAS	Cytosolic dsDNA, HSV

AIM2, Absent in melanoma-2; *cGAS*, cyclic GMP-AMP synthase; *DAI*, DNA-dependent activator of IFN; *dsRNA*, double-stranded RNA; *EBV*, Epstein-Barr virus; *HCMV*, human cytomegalovirus; *HCV*, hepatitis C virus; *HIV*, human immunodeficiency virus; *HSV*, herpes simplex virus 1/2; *IFI16*, Gamma-interferon-inducible protein Ifi-16; *JEV*, Japanese encephalitis virus; *LCMV*, lymphocytic choriomeningitis virus; *MCMV*, murine cytomegalovirus; *MDA-5*, melanoma differentiation-associated gene; *MMTV*, mouse mammary tumor virus; *MV*, measles virus; *RSV*, respiratory syncytial virus; *ssRNA*, single-stranded RNA; *VSV*, vesicular stomatitis virus; *ZBP1*, Z-DNA–binding protein 1.

INNATE IMMUNITY TO VIRUSES

Viral infection induces an extensive array of defense mechanisms in the host. Innate defenses come into play to block or inhibit initial infection, protect cells from infection, or eliminate virus-infected cells. Innate mechanisms occur well before the effectors of adaptive immunity become active, but they are critical for the initiation of adaptive immunity via the elicitation of inflammation that promotes immune cell activation. The innate immune defenses are initiated via pattern recognition receptors (PRRs), which recognize pathogen-associated molecular patterns (PAMPs)[3] (see Chapter 3). These include transmembrane receptors of the Toll-like receptor (TLR) family, two families of intracellular receptors including the NOD-like receptors (NLRs) and the RIG-I–like helicases (RLHs), as well as the sensor molecule absent in melanoma-2 (AIM2). Additionally, the molecules cyclic guanosine monophosphate–adenosine monophosphate (GMP-AMP) synthase (cGAS), DDX41, IFI16, and Z-DNA–binding protein 1 (ZBP1) can sense cytosolic DNA (Table 25.2). These cellular sensors promote the expression of interleukin-1 (IL-1) and IL-18, type I (α/β) interferon (IFN-I), and a variety of IFN-stimulated genes and inflammatory cytokines, and chemokines. TLRs are cell surface or endosomal membrane-bound proteins expressed by numerous cells, including dendritic cells (DCs), macrophages, lymphocytes, and parenchymal cells. Expression of TLRs is largely inducible in most cell types, although some (TLR7/8/9) are constitutively expressed at high levels by specialized plasmacytoid DCs for rapid IFN production. Different TLR molecules recognize specific viral products, such as single- and double-stranded RNA (TLR 3 and TLR7/8, respectively) or double-stranded DNA (TLR9).

The RLHs retinoic acid-inducible gene I (RIG-I) and melanoma differentiation-associated gene (MDA-5) mediate cytoplasmic recognition of viral nucleic acids. These activate mitochondrial antiviral signaling (MAVS) proteins to stimulate IFN-I production and activate inflammasomes, which are molecular complexes that facilitate the activation of caspases and induce the production of proinflammatory IL-1β and IL-18. NLRs are a second class of cytosolic sensors of PAMPs that activate inflammasomes via the adapter protein ASC. These include the NLRP (or NALP), NOD, and IPAF/NAIP receptors. Three major inflammasomes have been shown to be involved in antiviral immunity: the NLRP3 inflammasome, the RIG-I inflammasome, and the AIM2 inflammasome.[3]

The innate defense system consists of multiple cellular components and many specialized proteins. The longest known and best-studied antiviral proteins are the α/β IFNs, which act by binding to the type I IFN receptor and result in the transcription of more than 100 IFN-stimulated genes. One consequence of this "antiviral state" is the inhibition of cell protein synthesis and the prevention of viral replication. Multiple leukocyte subsets are involved in innate defense, including macrophages,

KEY CONCEPTS

Major Antiviral Innate Defense Mechanisms

Acting to block infection:
- Natural antibodies
- Complement components
- Some cytokines and chemokines

Acting to protect cells from infection:
- Interferon-α/β (IFN-I)
- Interferon-γ (IFN-γ)
- Interleukin-1 (IL-1),18

Acting to destroy or inhibit virus-infected cells:
- Natural killer (NK) cells
- Natural killer T cells (NKT cells)
- Macrophages
- Neutrophils
- γδ T cells
- Nitric oxide

Involved in regulating antiviral inflammatory response:
- ILs-1, 6, 10, 12, 18, 23, 33
- Transforming growth factor (TGF)-β
- Chemokines (CCL2, 3, 4, 5)

DCs, neutrophils, natural killer (NK) cells, natural killer T cells (NKT cells), and γδ T cells. Furthermore, tissue cells, including fibroblasts, epithelial cells, and endothelial cells, express PRRs and respond to viral infection via the production of innate cytokines, including IFN-I and IL-1. IFN-I is a critical link between the innate and adaptive immune systems via activation of DCs and T cells, as well as protecting T cells from NK cell–mediated attack.[4] IFN-Is can also activate NK cells and induce other cytokines that promote NK responses, such as IFN-γ and IL-12. NK cells produce proinflammatory cytokines; they can kill infected cells and interact with DCs and are an important component of innate defense against viruses. NK cells can protect against some herpes viruses, which downregulate major histocompatibility complex (MHC) expression in the cells they infect. NK cells are also important in resistance to mouse and human CMV and possibly to HIV, influenza virus, and Ebola virus.[5] NK cells have also recently been shown to possess traits of adaptive immunity and, like T and B cells, can form populations of memory cells.[6] NK cells are regulated by an array of activating and inhibitory receptors, whose expression and function are just beginning to be understood. Uninfected cells are usually protected from NK cell cytolysis as they deliver negative signals, such as high expression of MHC molecules. In contrast, virus-infected cells are killed either because they deliver positive signals or because they lack adequate MHC-negative signals. NK cells may also control excessive immune responses to viruses by killing $CD4^+$ T cells and indirectly regulating cytotoxic T lymphocyte (CTL) responses. NKT cells may provide some antigen-specific innate immune protection against certain viruses, such as influenza virus.

Several classes of innate host proteins function in antiviral defense. These include natural antibodies, which may play a role in defense against some viral infections, as well as pentraxins and complement proteins. Some viruses may be directly inactivated by complement activation or be destroyed by phagocytic cells that bind and ingest complement-bound virions. Several proinflammatory cytokines and chemokines induced by virus infection also play key roles in defense. Foremost among these are IL-1 and other members of the IL-1 family, including IL-18 and IL-33. These cytokines influence both innate and adaptive immune cells and play critical roles in antiviral defense. Other antiviral cytokines are produced early, following infection, such as tumor necrosis factor (TNF), IFN-γ, IL-12, IL-6, and chemokines, such as MIP-1α. In particular, IL-12 is a potent inducer of IFN-γ from NK cells. Inflammatory chemokines also play an important role in innate antiviral defense by orchestrating macrophage, neutrophil, DC, and NK cell responses at the site of infection. Not only are these components of innate immunity involved in mediating initial protection against viruses but also several components (e.g., the PRRs; the cytokines IFN-I, IL-I, IL-33, and IL-12; and phagocytes, including macrophages, monocytes, and DCs) serve to shape the nature and effectiveness of the subsequent adaptive response to viral pathogens. For instance, DCs require innate signals, such as IFN-I and IL-12, for maturation and optimal T-cell activation. Furthermore, $CD8^+$ T cells responding to viruses need IFN-I and IL-33 signals for expansion and memory formation. Thus, both the magnitude and the type of innate response induced by virus infection have a marked influence on the generation of adaptive immune responses.

ADAPTIVE IMMUNITY TO VIRUSES

Innate immunity generally only slows, rather than stops, viral infection, allowing time for the adaptive immune response to begin. The two major divisions of adaptive immunity, antibody-mediated and T-cell–mediated, are mainly directed at different targets. Antibodies usually function by binding to free viral particles and, in so doing, block infection of the host cell (see Chapter 8). In contrast, T cells act principally by recognizing and destroying virus-infected cells or by orchestrating an inflammatory response that includes several antiviral components (see Chapter 12). As all viruses replicate within cells and many can spread directly between cells without reentering the extracellular environment, resolution of infection is reliant more on T-cell function than on antibody function. However, broadly neutralizing antiviral antibodies have the potential to be effective therapies against many different human infections, including HIV, influenza viruses, and Ebola virus. Recent advances have allowed researchers to isolate and identify human monoclonal antibodies (mAbs) against these and other pathogens,[7] offering promise of new therapies as well as significant insight for vaccine design. Antiviral antibodies are also very important as an immunoprotective barrier against reinfection. It is the presence of antibodies at portals of entry—most often mucosal surfaces—that is of particular relevance to influenza, HSV, and HIV infections. Yet, how to generate vaccines that induce optimal antibody responses, including broadly neutralizing antibodies, remains an important unsolved problem.

In some individuals, highly potent or highly cross-reactive antibodies are generated naturally in response to viral infection. Such broadly neutralizing antibodies, sometimes called super-antibodies, offer many potential advantages for therapy, especially if engineered to have a long half-life in vivo.[8] Current single-cell approaches to identify and isolate B cells for human monoclonal antibody generation will be decisive for the use of broadly cross-reactive human monoclonal antibodies for therapy, including against emerging and pandemic viruses, such as SARS-CoV-2, the causative virus of the recent COVID-19 pandemic.

Initiation of adaptive immunity is closely dependent on early innate mechanisms that activate antigen-presenting cells (APCs), principally subsets of DCs. APCs and lymphocytes are drawn into lymphoid tissues by chemokine and cytokine signals and are retained there for a few days to facilitate effective

intercellular interactions. The architecture of the secondary lymphoid tissues supports the coordinated interactions among the cells of the adaptive immune system through a network of supportive stromal cells and local chemokine gradients (see Chapter 2). The induction events occur in lymph nodes draining an infection site or in the spleen if the virus enters the bloodstream. The passage of viral antigens to lymph nodes usually occurs in DCs. Some viruses are able to compromise the function of APCs, such as HSV and measles virus, which can inhibit DC maturation.

B-cell activation occurs following antigen encounter in the B-cell follicles, and possibly the T-cell zones, in the spleen or lymph nodes. Some activated B cells become short-lived plasma cells, whereas others move to the edges of the B-cell follicles and interact with antigen-specific helper CD4 T cells via the presentation of antigenic peptides on B-cell MHC class II molecules. These Bcl6-dependent CD4 T follicular helper (Tfh) cells are specialized for providing help for B-cell responses and are needed to promote and regulate B-cell responses.[9] Activated B cells initiate germinal center (GC) reactions with the help of CD4 Tfh cells, ensuring somatic hypermutation and affinity maturation for the selection of high-affinity, antibody-producing, long-lived plasma cells, as well as memory B cells.[10] At the molecular level, upregulation of the transcription factors Blimp-1, XBP-1, and IRF-4 dictates plasma cell formation, whereas Pax-5 expression delineates B cells destined for GC reactions and the memory B-cell lineage.

Antibody binding to epitopes expressed by native proteins at the surface of free virions usually blocks viral attachment or penetration of target cells. Sometimes the consequence is viral lysis (with complement proteins also involved), opsonization, or sensitization for destruction by Fc receptor-bearing cells that mediate antibody-dependent cellular cytotoxicity (ADCC). Occasionally, however, Fc receptor binding of antibody-bound virus may facilitate infection and result in more severe tissue damage. This occurs in dengue fever and may happen in some instances in HIV infection. This antibody-dependent enhancement (ADE) effect occurs in dengue fever, some vaccines against dengue, and possibly in some instances in HIV infection.[11] Although ADE was a concern with the development of coronavirus vaccines because animal models challenged with SARS-CoV were found to develop ADE, this has not proven to be a problem following SARS-CoV-2 vaccination with current approved vaccines.

The antibody involved in the protection of mucosal surfaces in humans is predominantly secretory immunoglobulin A (IgA), but serum-derived IgG may also be protective, particularly in such sites as the vaginal mucosa. Both antibody isotypes act mainly to block infection of epithelial cells, although, in some instances, the antibody may transport antigen from within the body across epithelial cells to the outside. Mucosal antibody persists for a much shorter period compared with serum antibody, which explains, in part, why immunity to mucosal pathogens is usually of much shorter duration compared with immunity to systemic viral infections.

Like B-cell responses, T-cell responses to viral infections also begin within lymphoid tissues. Specific CD8 CTL precursors recognize antigen in the context of major histocompatibility complex (MHC) class I–peptide antigen complexes on DCs. The CD8 T cells become activated, proliferate, and differentiate into effectors. Expansion of these naïve antigen-specific precursors is considerable, often exceeding 10,000-fold, and results in an effector population that can account for 40% or more of a host's total CD8 T-cell population (Fig. 25.2). Various factors, including antigen and APCs, costimulatory molecules (e.g., CD28 and 4–1BB), and inflammatory cytokines (e.g., IFN-I and IL-12), are required to program the development of functional effector lymphocytes.[12] In some infections, CD4 T-cell help is also important to prime robust CTL responses via signals, including CD40 that are delivered to DCs. Activated CTL effectors then exit lymphoid organs and access almost all body locations via the bloodstream. However, effectors do not stay activated for long once the virus is cleared, and approximately 95% die by a process termed *activation-induced cell death*. Following this contraction phase, the remaining cells differentiate into memory cells, which remain as a more or less stable population in the host for many years. They represent an expanded pool of CTL precursors that can be activated upon secondary encounter with antigen and provide enhanced protection upon reinfection with the same virus (see next section). Although much of our knowledge of T-cell responses to viruses has been obtained from murine studies, it is increasingly clear that the fundamental principles are the same or similar in humans.

T-cell immunity against a particular virus involves both CD4 and CD8 T-cell subsets that recognize peptides derived from viral antigens bound to surface MHC proteins (class II and class I, respectively) (see Chapters 5 and 6). Complexes of viral peptides bound to MHC class II proteins are generated by APCs from scavenged and processed virus-infected cells or viral particles. Antigen–MHC class I complexes are expressed on the surface of infected cells, and antigen can also be transferred to

KEY CONCEPTS

Antiviral T- and B-Cell Immunity

Effector Systems	Recognized Molecules	Control Mechanisms
Antibody	Surface proteins or virions	Neutralization of virus, opsonization, or destruction of infected cells by ADCC
Antibody + complement	Surface proteins expressed on infected cells	Infected cell destruction by ADCC or complement-mediated lysis
Mucosal antibody (IgA)	Surface proteins or virions	Viral neutralization, opsonization, and transcytosis
CD4 T cells	Viral peptides (10–20 mers) presented on MHC class II surface, internal or nonstructural proteins presented by APCs	Antiviral cytokine and chemokine production; help for CD8 T-cell and B-cell responses; killing infected cells; regulatory functions to reduce immunopathology
CD8 T cells	Viral peptides (8–10 mers) presented on MHC class I surface, internal, or nonstructural proteins presented on infected cells or by cross-presentation	Killing infected cells or purging virus without cell death; antiviral cytokine and chemokine production

ADCC, Antibody-dependent cellular cytotoxicity; *APC*, antigen-presenting cell; *IgA*, immunoglobulin A; *MHC*, major histocompatibility complex.

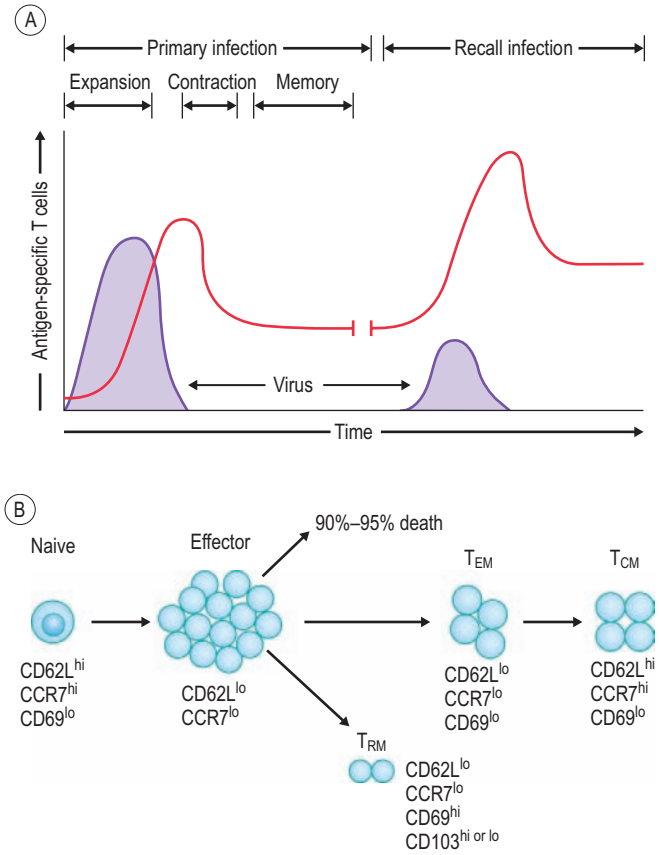


FIG. 25.2 Expansion/Contraction/Memory Phases of Adaptive Immunity and Memory Cell Subsets. (A) Dynamics of primary and secondary (recall) T-cell responses to viral infection *(red line)*. Both primary and recall T-cell responses undergo expansion and contraction phases, followed by stable immune memory. Recall responses induce a larger effector pool and reduced contraction, further boosting the memory pool. (B) Effector and memory T-cell differentiation. Antigen stimulation expands effector cells, most of which die during the contraction phase. Effector memory T cells (T_{EM}) that are formed gradually convert to central memory T cells (T_{CM}) over time, with corresponding changes in surface marker expression. Some effector T cells develop into resident memory T cells (T_{RM}) that persist in the tissues and do not reenter the circulation.

APCs from infected cells by a process known as *cross-presentation*. Recent experiments in mice have also demonstrated a role for the transfer of antigen between DCs as they migrate from infected tissues to the lymphoid tissues. Multiple subsets of DCs exist and specialize somewhat in antigen presentation on MHC-I or MHC-II.[13] During the process of activation, T cells can receive signals from multiple DC types in a temporally controlled sequence that coordinates CD4 and CD8 T-cell interactions.[12] Use of MHC class I and class II tetramers to directly visualize antigen-specific CD8 and CD4 T-cell responses, respectively, has demonstrated the significant size of T-cell responses to viruses, such that the majority of the activated T cells seen at the peak of the response are virus-specific.

CTLs function by recognizing virus-infected cells and killing them; this often involves perforins and cytotoxic granules containing granzymes. Effector CTLs can also induce death in target cells following engagement of the Fas ligand on the CTL with Fas on target cells. Both pathways lead to apoptosis of the target cell, involving the degradation of nucleic acids, including those of the virus. Alternatively, CD8 T cells also mediate defense through the release of various cytokines after antigen recognition. Some of the cytokines and chemokines most highly produced by CTLs include IFN-γ, TNF, lymphotoxin-α, and RANTES (CCL5) (see Chapters 14 and 15). These cytokines can have multiple antiviral effects on infected cells and on the cells around them, including the purging of virus from infected cells without killing the cells. This is particularly important for such viruses as HSV, which infects non-rejuvenating cells, such as nerve cells.

CD4 T cells are involved in antiviral defense as well as being modulators of inflammatory reactions to viruses. Multiple functional subsets of CD4 T cells are recognized based largely on the types of cytokines produced when they recognize antigen. CD4 T cells are more broadly reactive than CD8 T cells; they recognize larger peptides processed from viral proteins and are restricted by MHC class II. These CD4 T cells participate in antiviral immunity in several ways. They can act as helper cells for the development of high-affinity antibody responses and for more functional CD8 T-cell responses.[9,14] Additionally, CD4 T cells act as effectors and orchestrate inflammatory reactions, which either serve a protective function or, in some cases, become prolonged, causing chronic tissue damage (see Chapter 11). The latter can happen in HCV-mediated hepatitis and HSV-mediated stromal keratitis. Occasionally, CD4 T cells can mediate direct cytotoxicity, but they are less effective than CD8 T cells. The principal subsets of CD4 T cells involved in inflammatory reactions are T helper-1 (Th1) cells (producing mainly IFN-γ, TNF, IL-2) and Th17-producing cells (IL-17 and IL-22). A third effector subset, Th2 cells producing (IL-4, IL-5, and IL-13), also participates in inflammatory reactions, although in the case of viruses, these are usually more tissue-damaging than protective. This situation can occur in response to RSV infection. Regulatory T cells (Tregs) are a further subset of CD4 T cells of particular importance since these cells largely act to regulate the function of effector subsets and, in so doing, influence the severity and duration of inflammatory reactions (see Chapter 13).[15] Tregs produce antiinflammatory cytokines, such as IL-10 and transforming growth factor-β (TGF-β), and can be distinguished from other CD4 subsets by their expression of a unique transcription factor, FoxP3. The balance of CD4 T-cell subset representation in response to a viral infection is critical. In situations where responses become overtly tissue-damaging and chronic, the balance favors effector subsets. In such situations, changing the balance to favor Tregs can result in diminished lesions. How to achieve this objective by acceptable therapeutic approaches represents an active area of research.

IMMUNOLOGICAL MEMORY

Immunological memory is a cardinal feature of adaptive immunity. The goal of vaccinology is to induce long-lived immunological memory to protect against reinfection (see Chapter 87). Following infection with certain viruses, memory can be exceptionally long-lived, potentially for the life of the host (e.g., yellow fever and smallpox viruses).[16,17] Memory is defined by the persistence of specific lymphocytes and antibody-producing plasma cells rather than that of antigen to induce continuous lymphocyte activation. Humoral memory to viruses involves

long-lived plasma cells in bone marrow, which provide a continuous, low-level source of serum antibodies. This maintenance of humoral immunity also involves a population of homeostatically maintained memory B cells, which may be required to maintain stable numbers of long-lived plasma cells over time. The pool of memory T cells is regulated by low-level homeostatic division controlled by the cytokines IL-7 and IL-15. For memory CD8 T cells, IL-7 is primarily important for survival, whereas IL-15 is crucial for low-level proliferation to maintain the size of the memory T-cell pool.

KEY CONCEPTS

Principles of Antiviral Immunity

- Many human viral infections are successfully controlled by the immune system.
- Certain emerging viruses may overwhelm the immune system and cause severe morbidity and mortality.
- Other viruses have developed mechanisms to overwhelm or evade the immune system and persist.
- Individuals with defects in innate or adaptive immunity demonstrate more severe viral infections.
- T-cell immunity is more important for control than are antibodies in many viral infections.
- Antibodies are important to minimize reinfection, particularly at mucosal sites.
- Immune memory is often sufficient to prevent secondary disease, although not in all viral infections.
- Tissue-specific immune memory may be important to rapidly protect against reinfection at peripheral sites (e.g., skin and mucosae).

Immunological memory is defined by a pool of antigen-specific cells whose increased frequency enables rapid control of viral reinfection (see Fig. 25.2). IL-7Rα-expressing effector T cells are the precursors of this memory pool. This population of cells, which constitutes about 5% to 10% of the effector pool, preferentially survives the contraction phase and gradually differentiates into a stable memory population.[18] Upon reinfection, these memory cells can be rapidly activated and, by virtue of their increased frequency, mediate more rapid clearance of the viral pathogen. Moreover, repeated stimulation of memory cells via multiple infections with the same virus, or prime-boost vaccine regimes, further increases the size of the antigen-specific memory T-cell pool.[19] Restimulation also affects the activation status and tissue distribution of memory T cells, which may enhance protection from viral infection in mucosal and other tissues.

Experiments in humans and mice have demonstrated that memory T cells are heterogeneous. Memory T cells were divided into effector memory (T_{EM}) and central memory (T_{CM}) subsets, defined by the expression of two surface molecules involved in T-cell migration: CD62L and CCR7.[18] The $CD62L^{lo}CCR7^{lo}$ T_{EM} subset is found primarily in nonlymphoid tissues and the spleen, whereas the $CD62L^{hi}CCR7^{hi}$ T_{CM} subset is largely present in lymph nodes and the spleen. The current model predicts that effector T cells form the T_{EM} subset and that these cells gradually convert to a T_{CM} phenotype over time (Fig. 25.2, B). Although the conditions that control the rate of this conversion are unknown, it is likely that the amounts of antigen and inflammatory signals received during the effector phase greatly influence this by programming the epigenetic regulation of gene expression[20]. It has also been shown that CD4 T-cell help is required for the generation of long-lived memory CD8 T cells via interactions with DCs.

Studies suggest that T_{CM} are capable of mounting stronger proliferative responses following reinfection. Tissue-specific homing of T_{EM} cells permits them to enter sites of potential viral infection, such as skin and mucosae. However, we now know that many memory T cells found at sites of previous viral infections take up long-term residence in tissues.[21] This includes skin, intestines, lungs, the liver, and the brain (Fig. 25.3, A). These tissue-resident memory T cells (T_{RM} cells) are sequestered from the circulation and provide rapid protection against viruses, such as HSV in the skin, where they localize with a unique dendritic morphology and undergo slow surveillance of the tissue (Fig. 25.3, B). T_{RM} cells can produce various cytokines, including IFN-γ and IL-17. Notably, activation of T_{RM} cells can trigger enhanced early inflammation to drive local immunity. This is in contrast to T_{CM}, which migrate largely through lymphoid organs (spleen and lymph nodes), and T_{EM} cells, which can migrate through nonlymphoid tissues. These differences may define the physiological *raison d'être* for these memory T-cell subsets, highlighting that measurement of memory T cells in human peripheral blood is a poor representation of the total-body memory T-cell pool.

T_{RM} cells can be detected in tissues by using markers, such as CD69 and CD103, although these are imperfect identifiers and may not identify T_{RM} in all human tissues. T_{RM} cells in different anatomic locations share a common genetic signature and require common transcription factors for their formation, including T-bet, Eomes, Blimp1, and Hobit, as well as cytokines including TGF-β and IL-15. Yet, these cells also adopt a unique gene expression that is imprinted by the tissue environment and presumably imparts specialized functions on T_{RM} cells in each location. However, memory in certain peripheral tissues, such as lungs, appears to wane over time, suggesting that memory T cells may not persist in sufficient numbers in this site. This rationalizes a need for vaccines that induce optimal numbers of memory T cells in tissues as well as blood. For vaccines to be effective against many virus infections, they need to induce optimal T_{RM} responses at sites of viral entry. Using appropriate signals to "pull" memory cell precursors to specific sites could be one way of achieving this objective.[22]

IMMUNE EVASION AND IMMUNITY TO CHRONIC VIRAL INFECTIONS

Many, if not all, viruses employ immune blunting or delay tactics to circumvent aspects of the immune system, allowing them time to replicate further or escape detection (Table 25.3). One such mechanism may involve killing or infecting APCs. Viruses may also delay or prevent apoptosis induced by CTLs within infected cells. Other viral evasion measures aimed at the CD8 T cell–mediated antiviral defense system inhibit antigen processing, thereby minimizing effector CTL induction. To escape CTL killing, many viruses also downregulate the MHC molecules on the surface of infected cells. In addition, viruses may produce various mimics or modulators/inhibitors of cytokines, chemokines, or other components of the immune system or their receptors. Viruses also resort to antigenic hypervariability to escape antibody or T-cell recognition. This can occur during transmission from host to host (e.g., influenza virus) or within hosts during chronic infection through the generation of viral escape mutants. The latter is particularly important for HIV and HCV infections.

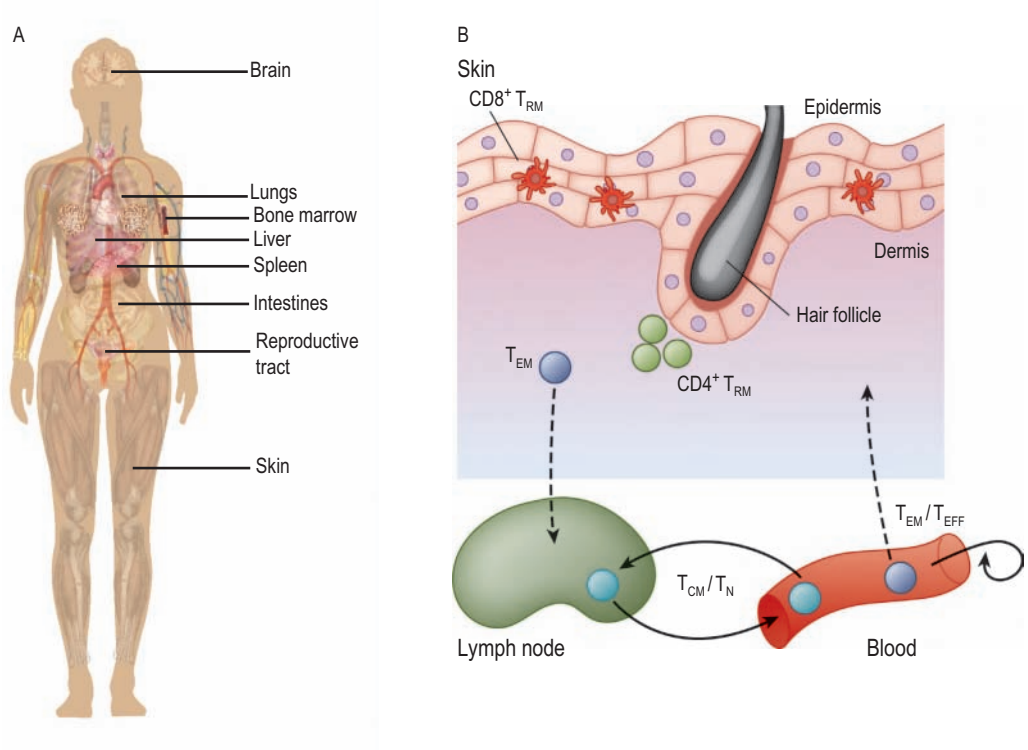


FIG. 25.3 Tissue-Resident Memory T Cells T_{rm} Reside Within Tissues and Sites of Previous Viral Infection and Provide Rapid Protection Against Reinfection. (A) T_{RM} cells have been found in most locations in the body. (B) Subsets of memory T cells in the skin. CD8 T_{RM} cells remain localized in the epidermis of the skin after herpes simplex virus (HSV) infection. Resident memory $CD4^+$ T_{RM} cells cluster in the dermis. Effector memory T cells ($CD4^+$ T_{EM}) continue to migrate through the dermal layers of skin with access to blood and lymphoid tissues. Central memory T cells (T_{CM}) and naïve T cells (T_N) circulate through lymphoid tissues via the blood.

TABLE 25.3 Mechanisms and Examples of Viral Immune Evasion

Mechanism	Example
Interference with viral antigen processing and presentation	HSV (ICP47), EBV (EBNA-1), HIV (Nef, Tat), HPV (E5), CMV (UL6)
Evasion of NK cell function	HIV (Nef), EBV (EBNA-1), CMV (UL40, UL18)
Inhibition of cell apoptosis	Adenovirus (RID complex and E1B), HIV (Nef), EBV (BHRF-1)
Destruction of T cells	HIV
Interference with antiviral cytokines and chemokines	EBV (IL-10 homolog), CMV(US28 chemokine receptor homolog), vaccinia virus (IL-18-binding protein), HIV (Tat chemokine activity)
Inhibition of complement action	HSV, pox viruses
Inhibition of DC maturation	HSV, vaccinia virus
Frequent antigenic variation	Influenza virus, HIV
Infection of immune-privileged site	Measles virus, VZV, and HSV (neurons)
Immune exhaustion	HIV, HCV, HBV

CMV, Cytomegalovirus; *DC*, dendritic cell; *EBV*, Epstein–Barr virus; *HBV*, hepatitis B virus; *HCV*, hepatitis C virus; *HIV*, human immunodeficiency virus; *HPV*, human papillomavirus; HSV, herpes simplex virus; *IL-18*, interleukin-18; *NK*, natural killer; *RID*, receptor internalization and degradation; *VZV*, varicella-zoster virus.

The success of many viral pathogens rests in their ability to subvert the host immune response. The most successful human viruses can escape the immune system and persist for the life of the host. Two well-studied examples of this are CMV and EBV. T-cell responses to these viruses are prominent and readily detectable in humans, and yet the immune system is unable to clear either pathogen completely. However, these viruses generally remain undetectable in immunocompetent individuals. Other viral infections, such as those caused by the herpes viruses HSV and VZV, are marked by periods of latency when no virus can be detected. Yet, periods of viral reactivation, often triggered by stress, can lead to episodes of disease. These are controlled by the immune response, which plays a central role in controlling herpes virus latency.

Many of the most medically important human viruses are associated with persistent viremia. These include those causing chronic infections, such as HIV, HCV, HBV, and human T-lymphotropic virus (HTLV), among others. Such chronic viral infections are marked by high levels of persisting antigen and can result in skewed T-cell immunodominance hierarchies, altered tissue localization of immune cells, and severely impaired T-cell function.[23] This altered T-cell function is hierarchical and results in functional T-cell defects ranging from reduced cytokine production and altered proliferative capacity (exhaustion) to death (deletion) of the responding T cells (Fig. 25.4, A).

Sustained viral antigen levels and inflammation are responsible for this immune dysfunction. This is in stark contrast to normal memory T-cell development, which occurs in the absence of persisting antigen (see previous section). Studies have demonstrated that signaling through multiple inhibitory receptors expressed on the cell surface contributes to exhaustion during chronic infections.[23] This includes the receptor programmed death (PD)-1, expression of which may be essential for preventing excessive

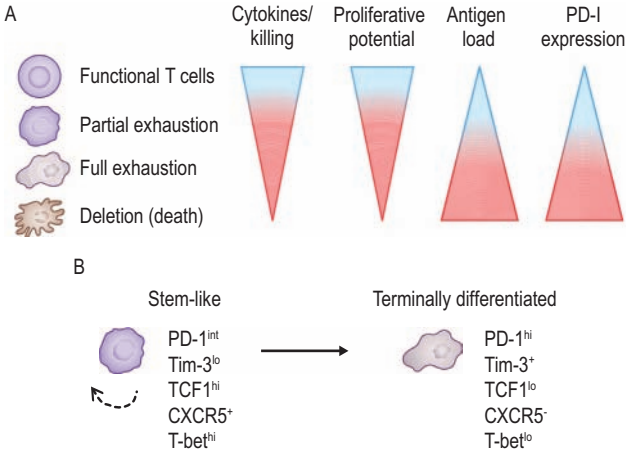


FIG. 25.4 Hierarchical Model of T-Cell Exhaustion During Persistent Viral Infection. (A) T-cell function (cytokine production, killing, and proliferative potential) is negatively influenced by increasing levels of antigen. Low levels of persistent antigen may lead to partial loss of function and intermediate levels of programmed death *(PD)*-1 expression. High, sustained levels of antigen over time can lead to full loss of function, high levels of PD-1, and eventually cell death (deletion). (B) A subset of stem-like exhausted T cells forms during chronic infection that can give rise to more terminally differentiated effector-like T cells.

immunopathology by effector T cells and yet appears to contribute directly to failed immunity to HIV infection and other chronic human viral infections. Although the molecular mechanisms of exhaustion remain incomplete, differential involvement of transcription factors (including Tbet, Blimp1, and Tox) and altered gene expression define exhausted T cells. Studies have implicated multiple inhibitory receptors as potential therapeutic targets, and although combinations of these checkpoint inhibitor blockade therapies are proving highly beneficial to the treatment of certain cancers,[24] similarly efficacious responses have yet to be fully demonstrated during chronic virus infection. Yet, the discovery of a stem-like subset of exhausted T cells that express Tcf1 and Tim3 and display enhanced proliferative capacity provides considerable promise for enhancing immunity against chronic viral infections (Fig. 25.4, B).

It is also important to note that the activation of immune cells involves considerable changes in their metabolic demands, in particular T cells that undergo phases of rapid expansion and production of antiviral cytokines. During chronic viral infection, exhausted T cells demonstrate marked changes in cellular metabolism, shifting from glycolysis to fatty acid oxidation, as well as impaired mitochondrial function.[23,25] Immune cells that are targets of viral infection, such as HIV-infected T cells or macrophages, also show marked metabolic changes. The complex interplay between immune cell functions and metabolic reprogramming in response to viral infections is yet to be fully appreciated but could reveal important new avenues for therapy and treatment of viral diseases.[26]

OUTCOMES OF VIRUS INFECTION: IMMUNITY OR IMMUNOPATHOLOGY

Typically, individual humans respond to a viral infection in different ways. When the common cold or even pandemic influenza infection occurs, only a small percentage of exposed persons may develop overt clinical disease. In the pre-vaccine days, poliomyelitis was a much-feared consequence of poliovirus infection, but only a very small percentage of infected persons developed paralyzing complications. Similarly, only an unfortunate few develop life-threatening meningoencephalitis following infection with the insect-transmitted West Nile virus. It is particularly characteristic of chronic viral infections that clinical expression is highly variable. With HCV, for example, in 70% to 80% of patients, some form of chronic liver disease develops, and the virus is not cleared. However, in up to 30%, the infection is controlled, the virus is cleared, and immunity to reinfection develops. The latter group of individuals make a type of immune response that includes protective antibodies along with an appropriate pattern of T-cell responsiveness.[27] This issue is particularly relevant during the recent SARS-COV-2 pandemic, where it seems that age and the presence of comorbidities, such as diabetes and hypertension, are variables that could greatly affect the outcome of infection.

We do not fully understand the reasons for the varying outcomes of virus infections in different persons, and almost certainly, multiple factors are involved. Many of these factors impact the response pattern made by the innate immune system, which, in turn, affects the magnitude and type of adaptive immune response that occurs. Some of the circumstances that do influence the outcome of infection include genetic susceptibility of the host, the age of the host when infected, the dose and route of infection, the variable induction in the host of anti-inflammatory cells and proteins, and the presence of concurrent infections and past exposure to cross-reactive antigens.

IMMUNOPATHOLOGY AND AUTOIMMUNITY

Immune responses against virus-infected cells often result in tissue damage, especially if cell killing is involved or if there is extensive recruitment and activation of inflammatory cell types, such as macrophages and sometimes neutrophils. If the response is brief and is quickly repaired, it is usually deemed an immunoprotective event. A prolonged tissue-damaging effect resulting from an immune reaction against viruses is considered immunopathology (Fig. 25.5). Such situations most commonly involve persistent viruses, which are themselves often mildly cytodestructive in the absence of an immune reaction. Chronic tissue damage initiated by viruses can also result in the development of an autoreactive and occasionally oncogenic response. For example, some autoimmune diseases may be initiated or exacerbated by viral infections, but no named virus has been regularly incriminated as a cause of human autoimmune disease.[28]

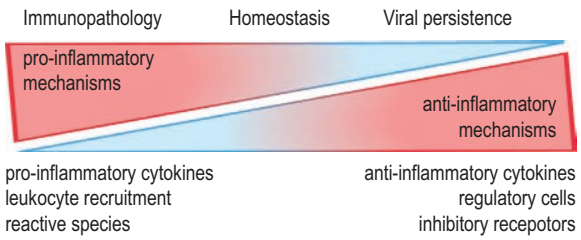


FIG. 25.5 A Balance of Signals is Required to Protect Against Viral Infection and Prevent Immunopathology. Multiple pro- and antiinflammatory mechanisms together decide the outcome of viral infection and the degree of tissue damage.

TABLE 25.4 Lesions Resulting From Immunopathology

Primarily involving CD8 T cells acting as cytotoxic T lymphocytes or sources of proinflammatory cytokines	Murine lymphocytic choriomeningitis virus
	Hepatitis B virus (HBV)–induced chronic hepatitis
	Coxsackie B virus-induced diabetes
	Coxsackie B virus-induced myocarditis
	Demyelination caused by some strains of mouse coronavirus and Theiler virus
Primarily involving CD4 T cells that produce Th1 cytokines	Demyelination caused by some strains of mouse coronavirus and Theiler virus
	Herpes simplex virus (HSV)–induced stromal keratitis
Involvement of CD4 T cells that produce Th2 cytokines	Respiratory syncytial virus (RSV)–induced pulmonary lesions
Involvement of antibody	Glomerulonephritis in chronic hepatitis B
	Dengue hemorrhagic fever

Circumstantial evidence exists for a virus link in multiple sclerosis (MS), insulin-dependent diabetes, and possibly systemic lupus erythematosus (SLE). In MS, many viruses have been isolated from patients, although no specific one has been tied to the disease etiology. The current hypothesis is that viral infections set up an inflammatory environment that may exacerbate or tip the balance toward disease in genetically susceptible individuals.

Immunopathological reactions involving viruses have several mechanisms, but T cells are usually involved as orchestrators of inflammatory events (Table 25.4). A clear example of immunopathology involving a virus is lymphocytic choriomeningitis virus (LCMV) in the mouse. This model has dominated ideas and has set several paradigms in viral immunology in general.[29] The first virus-induced immunopathological lesions recognized were glomerulonephritis and arteritis, noted in mice persistently infected with LCMV. The lesions were assumed to represent inflammatory reactions to tissue-entrapped immune complexes that activate complement. Similar immune complex–mediated lesions occur in other infections, including lung lesions found in severe influenza, RSV infection, viral hepatitis, and arthritis. However, only rarely have viral antigens been shown to contribute to the antigen component of the complex. An example where the inclusion of viral antigen in immune complexes has been demonstrated is chronic HBV infection of humans. Autoimmune diseases, such as SLE, also result from immune complex–mediated tissue damage. However, evidence linking viruses to the etiology or pathogenesis of SLE is scarce since the immune complexes in SLE do not appear to include viral antigens at any stage.

Thanks largely to the LCMV model, it is clear that CD8 T-cell recognition of viral antigens can result in tissue damage. In LCMV infection, damage occurs in the leptomeninges of immunocompetent mice, infected intracerebrally. Hepatitis can also occur in mice infected intravenously. Neither lesion becomes evident if the CD8 T-cell response is suppressed. CD8 T cell–mediated immunopathology can be a causative mechanism of chronic hepatitis associated with HCV and HBV infection, although the tissue damage also involves inflammatory CD4 T cells. Additional viral immunopathology models where lesions result primarily from CD8 T-cell involvement include myocarditis and insulin-dependent diabetes associated with coxsackie B virus infection. In both instances, CD8 T cells mainly orchestrate events, but tissue damage may result from the bystander effects of cytokines and other molecules, such as lipid mediators, metalloproteinases, and components of the oxygen burst. Although coxsackie virus can be a cause of diabetes in the mouse, attempts to relate viral infection directly to the etiology of human diabetes have so far failed.

CLINICAL RELEVANCE

Hypothesized Role of Viruses in Autoimmunity

Molecular mimicry: similar epitopes shared by virus and host

Bystander activation: chronic release of cytokines and host antigens activates local autoreactive lymphocytes

Viral persistence: chronic viral antigen presentation on host cells leads to prolonged immunopathology

Immunopathological reactions against viruses can also involve subsets of CD4 T cells, which can be either Th1 or Th17 or both. One well-studied example involves persistent infection with Theiler virus in mice.[30] This infection causes a demyelinating syndrome that resembles the autoimmune disease experimental allergic encephalomyelitis. In both situations, CD4 T cells that produce Th1 cytokines appear to serve as pathological mediators. Furthermore, in both models, an increase in the involvement of myelin-derived autoantigens occurs as the disease progresses. Once again, such observations indicate the possible role of a virus in autoimmune disease. With the Theiler virus model, the virus persists in the nervous system and chronically stimulates CD4 T cells to secrete an array of cytokines. The demyelinating events appear to result from cytokine action on oligodendrocytes. Myelin components, such as myelin basic protein, proteolipid protein, and myelin oligodendroglial glycoprotein, may be released and can participate as additional antigen in immunoinflammatory events. This scenario is referred to as *epitope spreading*.

Another model of virus-induced immunopathology that mainly involves the Th1 subset of CD4 T cells is stromal keratitis caused by HSV infection (Fig. 25.6).[31] The pathogenesis of this immunopathological lesion is unusual in that it occurs and progresses when viral antigens can no longer be demonstrated. The chronic immunoinflammatory lesions are mainly orchestrated by CD4 T cells, but multiple early events induce the subsequent pathology. Viral replication, the production of certain cytokines and chemokines (IL-1, IL-6, IL-12, and CXCL8), recruitment of inflammatory cells (e.g., neutrophils), and neovascularization of the avascular cornea all precede immunopathology. Recently, it has become evident that Th17 T cells participate in stromal keratitis lesions. The role of Th17 T cells as orchestrators of inflammatory reactions has been a major research focus, especially in lesions of autoimmune diseases.[32] When Th17 T cells are the principal mediators of tissue damage, abundant neutrophils are recruited to inflammatory sites, with such cells being mainly responsible for tissue damage.

A further mechanism of viral-induced immunopathology and autoimmunity is molecular mimicry. Molecular mimicry represents shared antigenic epitopes, either B- or T-cell antigen, between the host and virus. This concept originated with streptococci and their association with rheumatic fever. With human

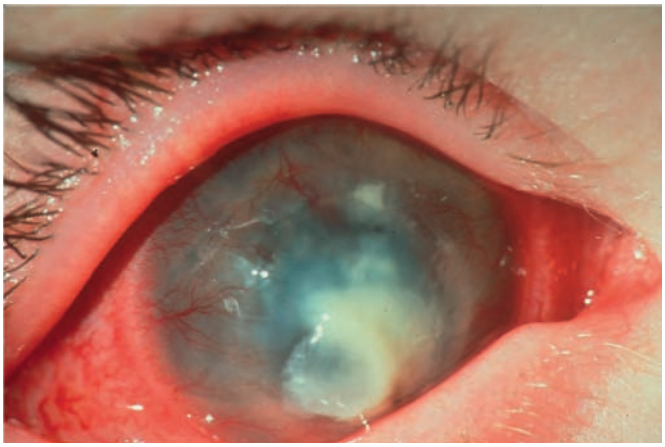


FIG. 25.6 Example of Herpetic Stromal Keratitis (Hsk) in the Human Eye After Herpes Simplex Virus-1 (HSV-1) Infection. Inflammation of the eye and eyelid can be observed, as well as neovascularization and substantial necrosis, ulceration, and opacity of the cornea.

autoimmune disease, there is little direct support for viral molecular mimicry; however, some animal models have been used to prove the theoretical case, where a viral antigen is expressed as a self-protein in the islet cells of the pancreas. In this model, subsequent infection with the virus induces diabetes. However, this is not true mimicry and may be more closely related to viral antigen persistence in a model such as Theiler disease.

As discussed previously, the outcome of a T-cell response to a virus may be critically dependent on the balance of the T-cell-type response. Thus, tissue damage is likely to be more severe and prolonged if CD8 or Th1 and Th17 CD4 T cells are predominant. Lesions become milder and may resolve when the balance favors Tregs. Accordingly, therapeutic approaches that can shift the balance of T cells are under trial.

KEY CONCEPTS

Phases of Immunity Affected by Regulatory T Cells

- Interference with antigen presentation by dendritic cells
- Inhibition of T-cell proliferation
- Inhibition of molecules involved in tissue-specific migration of effector cells
- Inhibition of T-cell effector functions in lymphoid and nonlymphoid tissues

TRANSLATIONAL RESEARCH OPPORTUNITIES

Reversing T-cell exhaustion in patients suffering from chronic infections or cancer will be a key clinical target in the near future. The discovery of multiple inhibitory receptors on exhausted T cells (e.g., PD-1, LAG-3, 2B4, TIM-3), as well as a stem-like subset of cells, has provided the opportunity to selectively improve T-cell function through blockade of these inhibitory receptors. This may be combined with a blockade of immunosuppressive cytokines (e.g., IL-10) or enhancement of signals stimulatory to the response (e.g., IL-7 therapy), as well as with more traditional antiviral therapies and vaccination. The challenge that lies ahead will be in determining which combination of inhibitory and stimulatory signals will need to be manipulated in different diseases and in different groups of patients.

The design of a new generation of vaccines to target diseases, such as HIV and influenza, may require tailor-made solutions for patients who respond poorly to vaccination or respond improperly, as with adverse effects, such as autoimmune reactions. High-throughput approaches now allow for the generation of a molecular signature of vaccination or infection.[33] Such systems biology approaches are expected to result in novel screening for immune protection parameters after vaccination. In the near future, this should also assist in the formulation of new vaccines containing key immune activators, such as those that stimulate certain subsets of T cells or induce appropriate homing molecule expression on these cells to direct them to tissues where they are required to mediate protection (e.g., mucosal sites, or skin).

ON THE HORIZON

Pressing Issues in Need of Solutions

- Design of new vaccines that induce broadly neutralizing antibodies
- Design of new vaccines that induce tissue-resident and circulating memory T-cell subsets
- Overcoming immune dysfunction during chronic viral infections for successful viral clearance
- Improving the efficacy of vaccines to viruses using systems biology approaches
- Therapies for reducing immunopathology during viral infections

In some individuals, viral infections cause mild, or sometimes debilitating, tissue damage. Factors that influence whether a viral infection results in immunopathology vary from individual to individual. These factors include age, the route of infection, preexisting immunity, host genetics, and the host's viral burden (or virome). The SARS coronaviruses, including the causative virus of the recent COVID-19 pandemic, SARS-CoV-2, are an excellent and timely example of viruses that can cause significant immunopathology. Pneumonia and pulmonary edema occur in patients with SARS due to excessive inflammatory cell recruitment to the lungs. Many patients also have multi-organ damage and vascular pathology due to a systemic cytokine storm.[34] Such responses usually involve strong stimulation of inflammatory cells and T cells which generate excessive amounts of cytokines, chemokines, and other damaging mediators. Patients with more severe COVID-19 disease are also typically lymphopenic, possibly due to substantial recruitment of lymphocytes into inflamed tissues. This presentation of severe disease is absent in most patients, and it remains unclear what factors govern the unrestrained inflammatory response in COVID-19. The many lessons learned from viruses provide a strong basis for understanding the immunopathogenesis of SARS-CoV-2 infection and future emerging diseases.

CONCLUSIONS

Humans are infected by many pathogenic viruses. In most cases, these infections are controlled by the immune system with limited damage to the host. However, certain viruses, particularly in cases where the host's immune system is impaired, can cause significant damage to the host's tissues. As our understanding of the mechanisms underlying innate immune defenses, antigen presentation, T- and B-cell responses, and Tregs continues to improve, so too does the ability to design better vaccines and

therapies to boost the immune control of viral infections. Although this remains a challenging goal, particularly for viruses such as HIV, and for novel viruses that can emerge swiftly, such as SARS-CoV-2, rapid scientific advances continue to provide many new avenues for novel therapies and vaccines.

ACKNOWLEDGMENTS

Scott N. Mueller is supported by the Australian National Health and Medical Research Council and Barry T. Rouse is supported by grants from the National Institutes of Health.

REFERENCES

1. Marsh M, Helenius A. Virus entry: open sesame. *Cell*. 2006;124(4):729–740.
2. Bachmann MF, Jennings GT. Vaccine delivery: a matter of size, geometry, kinetics and molecular patterns. *Nat Rev Immunol*. 2010;10(11):787–796.
3. Iwasaki A. A virological view of innate immune recognition. *Annu Rev Microbiol*. 2012;66:177–196.
4. Crouse J, Kalinke U, Oxenius A. Regulation of antiviral T cell responses by type I interferons. *Nat Rev Immunol*. 2015;15(4):231–242.
5. Jost S, Altfeld M. Control of human viral infections by natural killer cells. *Annu Rev Immunol*. 2013;31:163–194.
6. O'Sullivan TE, Sun JC, Lanier LL. Natural killer cell memory. *Immunity*. 2015;43(4):634–645.
7. Corti D, Lanzavecchia A. Broadly neutralizing antiviral antibodies. *Annu Rev Immunol*. 2013;31:705–742.
8. Walker LM, Burton DR. Passive immunotherapy of viral infections: "super-antibodies" enter the fray. *Nat Rev Immunol*. 2018;18(5):297–308.
9. Crotty S. T follicular helper cell differentiation, function, and roles in disease. *Immunity*. 2014;41(4):529–542.
10. Cyster JG, Allen CDC. B cell responses: cell interaction dynamics and decisions. *Cell*. 2019;177(3):524–540.
11. Taylor A, Foo SS, Bruzzone R, et al. Fc receptors in antibody-dependent enhancement of viral infections. *Immunol Rev*. 2015;268(1):340–364.
12. Ugur M, Mueller SN. T cell and dendritic cell interactions in lymphoid organs: more than just being in the right place at the right time. *Immunol Rev*. 2019;289(1):115–128.
13. Guilliams M, Ginhoux F, Jakubzick C, et al. Dendritic cells, monocytes and macrophages: a unified nomenclature based on ontogeny. *Nat Rev Immunol*. 2014;14(8):571–578.
14. Laidlaw BJ, Craft JE, Kaech SM. The multifaceted role of CD4(+) T cells in CD8(+) T cell memory. *Nat Rev Immunol*. 2016;16(2):102–111.
15. Veiga-Parga T, Sehrawat S, Rouse BT. Role of regulatory T cells during virus infection. *Immunol Rev*. 2013;255(1):182–196.
16. Antia A, Ahmed H, Handel A, et al. Heterogeneity and longevity of antibody memory to viruses and vaccines. *PLoS Biol*. 2018;16(8):e2006601.
17. Ahmed R, Akondy RS. Insights into human CD8(+) T-cell memory using the yellow fever and smallpox vaccines. *Immunol Cell Biol*. 2011;89(3):340–345.
18. Mueller SN, Gebhardt T, Carbone FR, et al. Memory T cell subsets, migration patterns, and tissue residence. *Annu Rev Immunol*. 2013;31:137–161.
19. Jameson SC, Masopust D. Understanding subset diversity in T cell memory. *Immunity*. 2018;48(2):214–226.
20. Tough DF, Rioja I, Modis LK, et al. Epigenetic regulation of T cell memory: recalling therapeutic implications. *Trends Immunol*. 2020;41(1):29–45.
21. Mueller SN, Mackay LK. Tissue-resident memory T cells: local specialists in immune defence. *Nat Rev Immunol*. 2016;16(2):79–89.
22. Iwasaki A. Exploiting mucosal immunity for antiviral vaccines. *Annu Rev Immunol*. 2016;34:575–608.
23. McLane LM, Abdel-Hakeem MS, Wherry EJ. CD8 T cell exhaustion during chronic viral infection and cancer. *Annu Rev Immunol*. 2019;37:457–495.
24. Baumeister SH, Freeman GJ, Dranoff G, et al. Coinhibitory pathways in immunotherapy for cancer. *Annu Rev Immunol*. 2016
25. Pallett LJ, Schmidt N, Schurich A. T cell metabolism in chronic viral infection. *Clin Exp Immunol*. 2019;197(2):143–152.
26. Varanasi SK, Rouse BT. How host metabolism impacts on virus pathogenesis. *Curr Opin Virol*. 2018;28:37–42.
27. Rouse BT, Sehrawat S. Immunity and immunopathology to viruses: what decides the outcome? *Nat Rev Immunol*. 2010;10(7):514–526.
28. Smatti MK, Cyprian FS, Nasrallah GK, et al. Viruses and autoimmunity: a review on the potential interaction and molecular mechanisms. *Viruses*. 2019;11(8).
29. Zehn D, Wherry EJ. Immune memory and exhaustion: clinically relevant lessons from the LCMV model. *Adv Exp Med Biol*. 2015;850:137–152.
30. Olson JK, Ercolini AM, Miller SD. A virus-induced molecular mimicry model of multiple sclerosis. *Curr Top Microbiol Immunol*. 2005;296:39–53.
31. Rajasagi NK, Rouse BT. The role of T cells in herpes stromal keratitis. *Front Immunol*. 2019;10:512.
32. Ma WT, Yao XT, Peng Q, et al. The protective and pathogenic roles of IL-17 in viral infections: friend or foe? *Open Biol*. 2019;9(7):190109.
33. Davis MM, Tato CM, Furman D. Systems immunology: just getting started. *Nat Immunol*. 2017;18(7):725–732.
34. Tay MZ, Poh CM, Renia L, et al. The trinity of COVID-19: immunity, inflammation and intervention. *Nat Rev Immunol*. 2020;20(6):363–374.

Host Defenses to Intracellular Bacteria

Stephen T. Reece and Stefan H.E. Kaufmann

The evolutionary relationship between humans and bacteria is so intimate that it is impossible to imagine the development of one without the other. Evolution of human immunity to bacteria has been accompanied by evolution of ingenious bacterial mechanisms to not only survive its onslaught but also to manipulate it for enhanced survival. These concepts are reflected in the lifestyle of intracellular bacteria. These pathogens actively seek out an environment inside human cells in which to flourish; yet, this is not an easy environment in which to survive. Human cells have developed an ability to differentiate bacterial from host components and to direct host cells to clear the invader. The most successful intracellular pathogens have adapted to the intracellular environment of a particular host cell, proliferate only slowly, and can live for long periods of time completely undetected by the immune system, as we see in the case of tuberculosis (TB). In other instances—for example, during listeriosis—intracellular infection is more explosive with the rich intracellular environment harnessed to rapidly amplify bacterial growth. More recently, bacteria that are not considered to typically adopt an intracellular lifestyle, but do so opportunistically, are being appreciated. *Staphylococcus aureus* can survive the intracellular environment of neutrophils and mononuclear phagocytes (MPs), and in doing so, spread infection around the human body, causing sepsis. In most cases, intracellular bacteria are long-lived in the human body, sometimes for the entire human lifetime. A wide spectrum of pathologies ensues from intracellular infection, making most intracellular bacteria highly clinically relevant. Moreover, new concepts on the influence of intracellular bacteria on host cell differentiation point to an ability to change infected cell phenotype to enhance survival.

This chapter dissects the current interpretation of this fascinating interplay between human and microbe, sheds light on how our immune system functions, and how cellular phenotype can be molded in cells whose fates were previously believed to be strictly predetermined. Finally, such insights can inform new therapeutic and prophylactic approaches to keep intracellular bacterial infections under control.

CLINICAL PEARLS

Distinguishing Clinical Characteristics of Infections With Intracellular Bacteria

- Nonsterilizing immunity
- Persistent bacteria, sometimes latent infection
- Formation of long-lasting tissue granulomas containing low numbers of viable bacteria
- Critical role of T cells in protection, role of antibodies less well established but likely to play an as-yet unappreciated role
- Critical role of immune response in pathology
- Lack of effective vaccines
- Host-directed therapies toward enhancing antimicrobial mechanisms while limiting host pathology

BALANCE OF PROTECTION AND PATHOLOGY DEFINES THE CHRONIC NATURE OF INTRACELLULAR BACTERIAL INFECTION

Some bacteria, such as *Listeria monocytogenes*, are fully eradicated once the host immune response has reached its peak activity. More often, the intracellular habitat provides a protective niche that promotes persistent infection in the face of an ongoing immune response. Here the bacteria can persist for long periods of time without causing clinical signs of illness, but bacterial growth can be reactivated and cause disease if the immune response becomes compromised. This occurs in *Mycobacterium tuberculosis* infection, resulting in disease years or even decades after primary infection. In fact, disease need not arise from infection at all; in many regions, for example, the vast majority of adults harbor *M. tuberculosis* without suffering from clinical disease. Alternatively, disease can develop directly after primary infection, during maturation of the immune response, or with regression once the immune response is sufficiently strong. Yet sterile eradication of the pathogen is rarely achieved; bacteria persist in a latent form, and illness may re-emerge at a later time. For example, *Rickettsia prowazekii* may persist for decades after convalescence from typhus to later cause Brill–Zinsser disease.

Several intracellular bacteria possess components that can profoundly influence the course of disease (such as the lipopolysaccharides [LPSs] of brucellae and salmonellae). Chronic persistence inside host cells, however, depends on the target cell remaining intact and physiologically active. Accordingly, many intracellular bacteria are of low toxicity and do not have dramatic direct effects on their host. Instead, pathogenesis is largely determined by the immune response. Classic examples of this concept include granuloma liquefaction in acute TB, which severely affects lung function, and eye scarring due to chronic or recurring *Chlamydia trachomatis* infection that ultimately leads to trachoma.

Intracellular bacterial survival and persistence have major consequences for pathology. Although many intracellular bacteria show some organ tropism, dissemination to other organs frequently occurs, resulting in different disease forms. For example, TB is generally manifested in the lung with 80% of cases, yet many other organs can be affected. In contrast to other *Salmonella enterica* serovars, the serovars Typhi and Paratyphi are not restricted to the gastrointestinal tract but are disseminated

to internal organs, primarily the liver and spleen. In these cases, the type of clinical disease depends markedly on the infected tissue type.

KEY CONCEPTS

Characteristic Features of Intracellular Bacterial Infections

- Persistence of bacteria inside mononuclear phagocytes (i.e., macrophages)
- Low to absent bacterial-mediated toxicity to the host
- Protection requires cytokine-mediated activation of infected phagocytes
- Interferon-gamma (IFN-γ) and tumor necrosis factor (TNF) produced by antigen-specific T cells are key cytokines for protection

INTRACELLULAR BACTERIAL INFECTIONS OF CLINICAL RELEVANCE

Granulomatous Infections

Tuberculosis

The major entry of tubercle bacilli into the human body is via inhalation into the lung (Table 26.1). These inhaled bacteria are then engulfed by alveolar macrophages (AMs), which transport the pathogens to the lung interstitia. Their exact fate after these events is enigmatic. Moreover, most infections in humans result in an asymptomatic carrier state referred to as latent TB infection (LTBI). Infection with *M. tuberculosis* starts with the so-called Ghon complex, characterized by a caseous lesion in the mid-lung as well as in the draining lymph nodes. These primary lesions can progress but their development rarely causes disease directly. Moreover, bacteria from these sites can disseminate to other regions of the lung and systemically, causing disease of the kidneys, liver, and central nervous system. Containment of the primary lesions, which leads to LTBI, is a function of an effective, predominately cellular anti-tubercular immune response. Infection of immunocompromised patients, notably newborns or those with acquired immunodeficiency syndrome (AIDS), frequently results in systemic disease (miliary TB). TB represents a major health problem worldwide, including an increasing incidence in many industrialized countries. In 2020, WHO estimated that 10 million active TB cases were diagnosed worldwide, with 1.5 million people dying of the disease. The much larger estimated number of 1.7 billion healthy individuals infected with *M. tuberculosis* well illustrates the dissociation of infection from disease. The emergence of multidrug-resistant strains and extremely drug-resistant strains has complicated treatment with currently available antibiotic therapy, and even when treatment is successful, recurrence of disease can occur. The currently available live whole cell vaccine BCG (bacille Calmette–Guérin), an attenuated strain derived from the etiological agent of bovine TB, *Mycobacterium bovis*, shows only low and variable protection against pulmonary TB; see (Table 26.1).

TABLE 26.1 Major Infectious Diseases Caused by Intracellular Bacteria

Disease	Pathogen	Prevalence	Incubation Time	Route of Infection	Target Cell
Granulomatous Intracellular Bacteria					
Tuberculosis	*Mycobacterium tuberculosis*	Worldwide	Years (latency after primary infection and disease reactivation) Weeks (miliary TB)	Inhalation of bacteria-containing microdroplets	Macrophage
Leprosy	*Mycobacterium leprae*	South America Africa India Southeast Asia	Years	Smear infection through mucosa/inhalation	Macrophage Schwann cell
Typhoid fever	*Salmonella enterica* serovars Typhi and Paratyphi	Worldwide	7–10 days	Fecal-oral	Macrophage
Brucellosis	*Brucella* spp.	Worldwide	Weeks to months	Zoonosis; cows, goats, pigs; inhalation, gut, skin abrasion	Macrophage
Listeriosis	*Listeria monocytogenes*	Worldwide	Days to months	Fecal-oral	Macrophage Hepatocyte
Nongranulomatous Intracellular Bacteria					
Legionnaires' disease	*Legionella pneumophila*	Worldwide	2–10 days	Inhalation	Macrophage
Rocky Mountain spotted fever	*Rickettsia rickettsia*	Western hemisphere	1 week	Tick bite	Vascular endothelial cell Smooth muscle cell
Urogenital infection	*Chlamydia trachomatis* serovars D–K	Worldwide	1–3 weeks	Sexual intercourse	Epithelial cell
Conjunctivitis, trachoma	*C. trachomatis serovars D–K*	Africa	Conjunctivitis: 1–3 weeks Trachoma: years	Eye	Epithelial cell
Cat-scratch disease	*Bartonella henselae* *B. quintana* *B. bacilliformis*	Worldwide	Bacillary angiomatosis Peliosis hepatis Endocarditis Bacteremia with fever Neuroretinitis: 1–3 weeks	Flea, sandfly or mosquito bite; animal scratch or bite	Erythrocyte Endothelial cell

Leprosy

Mycobacterium leprae is most likely transmitted by contact with patients who shed microorganisms in nasal secretions and lesion exudates. It primarily affects the nerves and the skin, frequently leading to stigmatizing deformities. In the skin, bacilli target keratinocytes and histiocytes as well as macrophages, whereas in the peripheral nerves, Schwann cells are the major target for entry. Leprosy is a spectral disease. The tuberculoid pole is characterized by rigorous T-cell responses, which succeed in restricting microbial growth in well-defined lesions containing few bacilli. In contrast, at the lepromatous pole, bacterial growth is unrestricted, and lesions contain abundant bacilli within macrophages lacking signs of activation. Several types of immunosuppression have been implicated in this latter type of disease. Infection of Schwann cells promotes nerve damage and anesthesia, which results in injuries and secondary infections that significantly exaggerate the disease. Despite the success of multidrug therapy in reducing the number of registered leprosy cases worldwide, there were some 200,000 new cases reported in 2018. This suggests that active transmission of *M. leprae* is still occurring and more effective interventions are required to prevent it.[1]

Nontuberculous Mycobacteria

Mycobacterial species present in the environment are typically unable to persist within activated macrophages, so rarely cause disease in individuals with competent immune status. Because of human immunodeficiency virus (HIV) infection, however, nontuberculous mycobacteria (NTM), primarily *M. avium/M. intracellulare*, have gained clinical importance, and these infections are recognized as one of the most common complications of AIDS in industrialized nations.

Mycobacterium scrofulaceum occasionally causes lymphadenitis in children, and *Mycobacterium kansasii* primarily causes infections in elderly individuals with pre-existing lung disease. Incidences of atypical mycobacteria, notably *Mycobacterium abscessus* in cystic fibrosis patients, have increased markedly over the last decades.[2] *Mycobacterium ulcerans* causes a severe subcutaneous infection characterized by chronic skin ulcerations known as Buruli ulcer. This pathology is caused—at least in part—by elaboration of a mycolactone toxin by the bacillus that exhibits highly cytopathic effects. Buruli ulcer is most predominant in West African countries, which accounted for most of the 2251 cases reported globally in 2014.[3]

Typhoid or Enteric Fever

S. enterica serovars Typhi and Paratyphi A, Paratyphi B, and Paratyphi C are leading causes of community-acquired bloodstream infections in low- and middle-income countries. The route of transmission is fecal-oral, and largely occurs via contaminated water sources. Bacteria are disseminated within MPs from the gastrointestinal tract to macrophage-rich organs, particularly the liver, spleen, and lymph nodes. Accordingly, typhoid is characterized by systemic symptoms such as prolonged fever and malaise with sustained bacteremia, although diarrhea or constipation may also be present. In some cases, an asymptomatic carrier state can persist as a result of chronic infection of the gallbladder, which maintains the environmental reservoir of infection in endemic areas. Typhoid fever remains a major cause of morbidity and mortality, with an estimated 11 million new cases and 117, 000 deaths per year worldwide.[4]

Gastroenteritis

S. enterica serovars Typhimurium and Enteritidis, often referred to as nontyphoidal salmonellae (NTS), are the major causes of salmonella gastroenteritis in humans, which occurs mainly as a result of the ingestion of contaminated food or water. The bacteria rapidly cross the intestinal epithelia and replicate in the lamina propria, inducing an influx of polymorphonuclear neutrophils (PMNs), which is generally sufficient to resolve the infection within a week. In rare cases, the bacteria enter the bloodstream and cause systemic bacteremia, most notably in AIDS patients, where death can occur as a result of septic shock.

Listeriosis

L. monocytogenes causes food-borne gastroenteritis. Clinical listeriosis affects mainly pregnant women, the elderly, unborn babies, and neonates. Disease manifestations are most severe in patients with a compromised immune system where the central nervous system becomes involved and fatal bacteremia can result. Additionally, as these bacteria can cross the placenta, listeriosis is a major cause of perinatal and neonatal disease, typically resulting in abortion. Listeria outbreaks are sporadic with low incidence but high fatality and affect high-income countries such as the United States. Outbreaks of *L. monocytogenes* due to contamination of meat products have increased over the last decades.[5]

Brucellosis

Brucellosis is the most common global zoonosis of humans, with approximately 500,000 cases per year.[6] It is caused by *Brucella abortus*, *Brucella melitensis* or *Brucella suis*, which primarily infect cows, goats and pigs, respectively. The bacteria are transmitted to humans via inhalation, abraded skin, or the gastrointestinal tract. Lesions are primarily found within macrophage-rich tissues, especially the spleen and bone marrow. Human brucellosis is characterized by systemic symptoms, particularly undulant fever. Although the disease often remains subclinical, in some patients it becomes chronic, and relapses and remissions may occur. Interest in brucellosis has increased in the last 5 years due to elevated levels of detection resulting from better surveillance.

Lymphogranuloma Venerum

Lymphogranuloma venereum (LGV), a sexually transmitted disease, is highly prevalent in Africa, Southeast Asia, and Latin America. LGV has recently emerged as an infection of sexually active homosexual men in Europe and the United States. It is caused by the L1, L2, and L3 serotypes of *C. trachomatis*, which are disseminated from the urogenital tract to local lymph nodes and to the skin. Accordingly, LGV is characterized by lymph node swelling and skin lesions.[7]

Melioidosis

Burkholderia pseudomallei is a gram-negative bacillus and the causative agent of melioidosis, endemic in Southeast Asia and northern Australia. The disease can be acquired through inhalation, ingestion, or through cuts in the skin. Susceptible hosts can suffer from abscess formation in multiple organs and, in some cases, disseminated infection resulting in septic shock accompanied by pneumonia. There are an estimated 165,000

cases of melioidosis per year globally, resulting in approximately 89,000 deaths.[8]

Tularemia

This rare zoonosis in humans caused by *Francisella tularensis* is mainly found in rabbits and has recently gained wider recognition due to its potential for dual use. Infection can be spread to humans via contaminated animals or tick bites. This gram-negative bacterium survives in macrophages and primarily causes acute pneumonia as well as sores of the skin, with subsequent involvement of the lymph nodes.[9]

Nongranulomatous Infections

Legionnaires' Disease or Legionellosis

Legionnaires' disease is caused by *Legionella pneumophila*, an environmental bacterium that persists within amoeba living in water reservoirs (e.g., air-cooling systems), from where it is spread aerogenically. Infection is exacerbated by a compromised immune status. Characteristically, legionnaires' disease presents as atypical pneumonia associated with general symptoms and is complicated by extrapulmonary infection, renal failure, and lung abscesses. Cases of legionnaires' disease in the United States increased from 0.39 to 1.36 per 100, 000 people from 2000 to 2011.[10]

Chlamydial Urethritis, Cervicitis, and Conjunctivitis

C. trachomatis serovars D–K enter and persist in epithelial cells of the urogenital tract, causing cervicitis and urethritis. In women, infertility can develop as a result of chronic or recurrent infection. In neonates, congenital infection during birth may result in conjunctivitis and pneumonia. Urogenital infections by chlamydiae occur worldwide and are now considered to be the most common bacterial sexually transmitted disease, with an estimated 100 million new infections occurring annually.

Trachoma

Smear infections of the eye with *C. trachomatis* serovars A, B, and C cause inclusion conjunctivitis. As a consequence of multiple chronic infections and of the resulting immune response, scars develop that eventually injure the cornea, leading to trachoma. *C. trachomatis* has been responsible for 1.9 million cases of visual loss worldwide; mass administration of azithromycin as part of a WHO public health strategy aims to eliminate trachoma by 2020.

Chlamydia pneumoniae

C. pneumoniae (formerly known as *C. trachomatis* TWAR strain) is the cause of mild respiratory disease in young adults and may cause serious infections in older, more debilitated patients. Atypical pneumonia may also be caused by *C. psittaci*, although this zoonosis, transmitted by birds, is relatively rare.[11]

Typhus

Rickettsia prowazekii, R. typhi, and *Orientia tsutsugamushi* cause diseases of varying severity. They are transmitted by arthropods and infect vascular endothelial cells at the site of an insect bite or scratch, causing skin reactions.[12] Subsequently, pathogens are disseminated to the central organs and more general symptoms develop. Globally, typhus is of minor importance.

Rocky Mountain Spotted Fever, Ehrlichiosis

Rocky Mountain spotted fever is caused by *Rickettsia rickettsii*. Infection of the vascular endothelium leads to systemic symptoms and skin manifestations that may be followed by shock and neurological complications. Worldwide, this disease, as well as Mediterranean spotted fever caused by *Rickettsia conorii*, is of minor importance, as is ehrlichiosis, a newly emerging zoonosis transmitted by ticks and caused by various *Ehrlichia* spp., mainly *E. chaffeensis*.[13] Disease manifestations include generalized symptoms such as fever and muscle pain.

Bartonella

Bartonella spp. represent gram-negative facultative intracellular pathogens transmitted by insect vectors such as fleas, sandflies, and mosquitoes. The most clinically relevant species are *B. henselae*, *B. quintana*, and *B. bacilliformis. B. henselae* causes cat-scratch disease (CSD), resulting in local lymphadenopathy in the lymph node draining the scratch site accompanied by fever, headache, and splenomegaly. Oculoglandular involvement (Parinaud syndrome), encephalopathy, neuroretinitis, or osteomyelitis can occur, albeit in rare cases. In immunosuppressed patients, bacillary angiomatosis and peliosis can occur, characterized by pseudotumoral proliferation of endothelial cells. Bacteria persist within erythrocytes with the intracellular location providing a protective niche.[14]

Sepsis, *Staphylococcus aureus*, and *Pseudomonas aeruginosa*

Sepsis is a life-threatening condition caused by a dysregulated response to infection. Bacteremia, or presence of bacteria in the blood, is a common trigger for sepsis. *S. aureus*, a gram-positive bacterium, often presents asymptomatically in the human nasal pharynx and on the skin. Yet, it causes up to 23% of healthcare-associated bacteremia infections in the United States.[15] Infection in hospital settings is likely via soft tissue lesions during surgical intervention or introduction of catheters and other medical prosthetic devices. Although *S. aureus* is considered a facultative intracellular pathogen, intracellular survival in a subset of neutrophils that are rendered defective in intracellular bacterial killing contributes to systemic dissemination of infection. Furthermore, reduced intracellular penetrance of the antibiotic treatment of choice for methicillin-resistant *S. aureus* (MRSA), namely vancomycin, requires extended treatment durations of up to 6 weeks, reflecting the clinical importance of clearing this intracellular niche.

Pseudomonas aeruginosa is an opportunistic gram-negative bacterial pathogen and prominent cause of ventilator-acquired pneumonia in immunosuppressed individuals in healthcare settings. In addition, chronic *P. aeruginosa* infection is associated with greater than 80% of adult cystic fibrosis sufferers and significantly shortens the life span of these individuals due to destructive bronchitis and bronchiolitis. Although, *P. aeruginosa* primarily forms extracellular biofilms during infection in the lung, persistence of *P. aeruginosa* in neutrophils and epithelial cells occurs via manipulation of intracellular killing mechanisms and could contribute to the chronic nature of infection and dissemination in the lung.

GRANULOMA PATHOLOGY AS HALLMARK OF INTRACELLULAR BACTERIAL INFECTION

KEY CONCEPTS

Balance of Protection and Host Pathology in Granulomas

- Macrophage activation results in bacterial death (protective)
- Intracellular bacterial killing by "killer molecules" from T cells (protective)
- Lysis of infected macrophages by T cells results in release of bacteria and killing by more effective effector cells (protective) or bacterial dissemination (pathogenic)
- Development of central necrosis in granulomas results in death of tissue and bacteria (protective/pathogenic)
- Fibrotic encapsulation of granuloma results in containment of infection (protective)
- Over-exuberant tissue fibrosis and necrosis (pathogenic)
- Liquefaction of central necrotic tissue in granulomas results in bacterial replication, cavity formation, and transmission of bacteria (pathogenic and contagious)

A characteristic feature of many infections caused by intracellular bacteria is the eventual need for tissue remodeling by the host at the site of infection. Granulomas are the result of an inability to rapidly clear host tissue of intracellular bacteria and are a striking locus of the host–pathogen interface (Fig. 26.1). The longevity of a granuloma depends directly on the continuous presence of the microbial pathogen, and the lesion generally disappears after its sterile eradication. Granulomas form the focus of the coordinated crosstalk between different types of T and B lymphocytes and infected and uninfected MPs and dendritic cells (DCs). Even if the immune system fails to eliminate bacteria inside the granuloma, the latter performs a protective function by containing microbes within distinct foci and preventing their dissemination. At the same time, the granuloma can be detrimental to the host because it can interfere with physiological organ functions.[16] More detailed study of cellular phenotype within granulomas is starting to establish how cellular differentiation is orchestrated and how the granuloma develops.

Granulomatous lesions are generally initiated by nonspecific inflammatory signals mediated by bacterial products, chemokines, and proinflammatory cytokines that are produced by endothelial cells and MPs at the site of infection. Inflammatory phagocytes (of both monocytic and granulocytic origin) are attracted to the site of microbial replication and an infiltrative, sometimes exudative, lesion develops. Following the accumulation and activation of increasing numbers of MPs and DCs, this lesion takes an increasingly structured granulomatous form. A significant number of B cells are also found, which seem to influence granuloma morphology. Once specific T cells have been attracted to the lesion, it transforms into a solid granuloma that provides the most appropriate tissue site for antibacterial protection. Here, activation of MPs by interferon-gamma (IFN-γ) and tumor necrosis factor (TNF) inhibits microbial growth. However, unbridled macrophage activation can have tissue-damaging effects and mechanisms within the granuloma tightly regulate these effects. Eventually, the granuloma is encapsulated by a fibrotic wall and its center becomes necrotic. Both tissue reactions are primarily protective, the former by promoting bacterial containment and the latter by reducing the nutrient and oxygen supply to the pathogen. The combined effects of chronic macrophage activation, persistence of intracellular bacteria, and hypoxia likely lead to enhanced cell death in the center of granulomas, resulting in the formation of a caseum. Caseation may also favor the local replication of normally facultative intracellular bacteria in the cellular detritus, as well as microbial dissemination to distant tissue sites and the environment to transmit infection. Hypoxia also has pronounced effects on enzyme functions that can dictate macrophage phenotype.

In direct contrast to the localized containment of bacterial intracellular infection mediated by granuloma formation in the host, more disseminated forms of TB also occur, particularly in children or those with HIV.[17] Multiple organs can be involved in extrapulmonary TB, including the spleen, liver, and kidney. Dissemination of infection to the brain tissue leads to TB meningitis, which is fatal if left untreated. Dissemination of bacteria typically occurs via the lymphatic system and is likely to involve intracellular carriage.

ROLE OF THE FACULTATIVE INTRACELLULAR NICHE IN SEPSIS AND PNEUMONIA

Dissemination of *S. aureus* and *P. aeruginosa*, triggering sepsis and causing pneumonia, is increasingly appreciated as involving intracellular carriage of bacteria. *S. aureus* infection is typically cleared by neutrophils, and neutropenia in both humans and mice causes high susceptibility to infection. Despite this, neutrophils that are defective for intracellular bacterial killing also allow dissemination of bacteria to other organs, thereby seeding new sites of infection. Recent studies indicate that *S. aureus* surviving within Kupffer cells in the liver also leads to dissemination to the peritoneal cavity, where in turn bacteria are taken up by tissue-resident macrophages able to disseminate infection further. Targeting these infected cells with novel therapies could stem dissemination of infection and safeguard against sepsis in high-risk patients.

Neutrophils are similarly vital for protection against *P. aeruginosa*. Intracellular survival can be enhanced by a type III secretion system (T3SS) encoded by *P. aeruginosa* that secretes bacterial proteins via a specific needle structure that projects from the bacterial surface. These secreted proteins then interfere with intracellular killing of bacteria by neutrophils. Antibodies that inhibit T3SS by *P. aeruginosa* could prove useful in treating lung infections in susceptible patients, such as those with cystic fibrosis.[18]

THE INTERDEPENDENCE OF INNATE AND ADAPTIVE IMMUNITY IN PROTECTION AGAINST INTRACELLULAR BACTERIA

Innate Immune Mechanisms as First-Line Defense

The interaction between host cell and pathogen that defines the intracellular fate is represented by multiple layers. The first layer that differentiates intracellular bacteria from other bacteria, notably commensal bacteria that colonize the host but do not cause infection, is that of host cell entry. Extracellular bacteria are typically engulfed by professional phagocytes, which include tissue macrophages, DCs, and PMNs. This uptake is enhanced by host components of the complement system and antibodies, which bind to complement receptors (CRs) and Fc receptors, respectively, on professional phagocytes.[19] *M. tuberculosis* actively

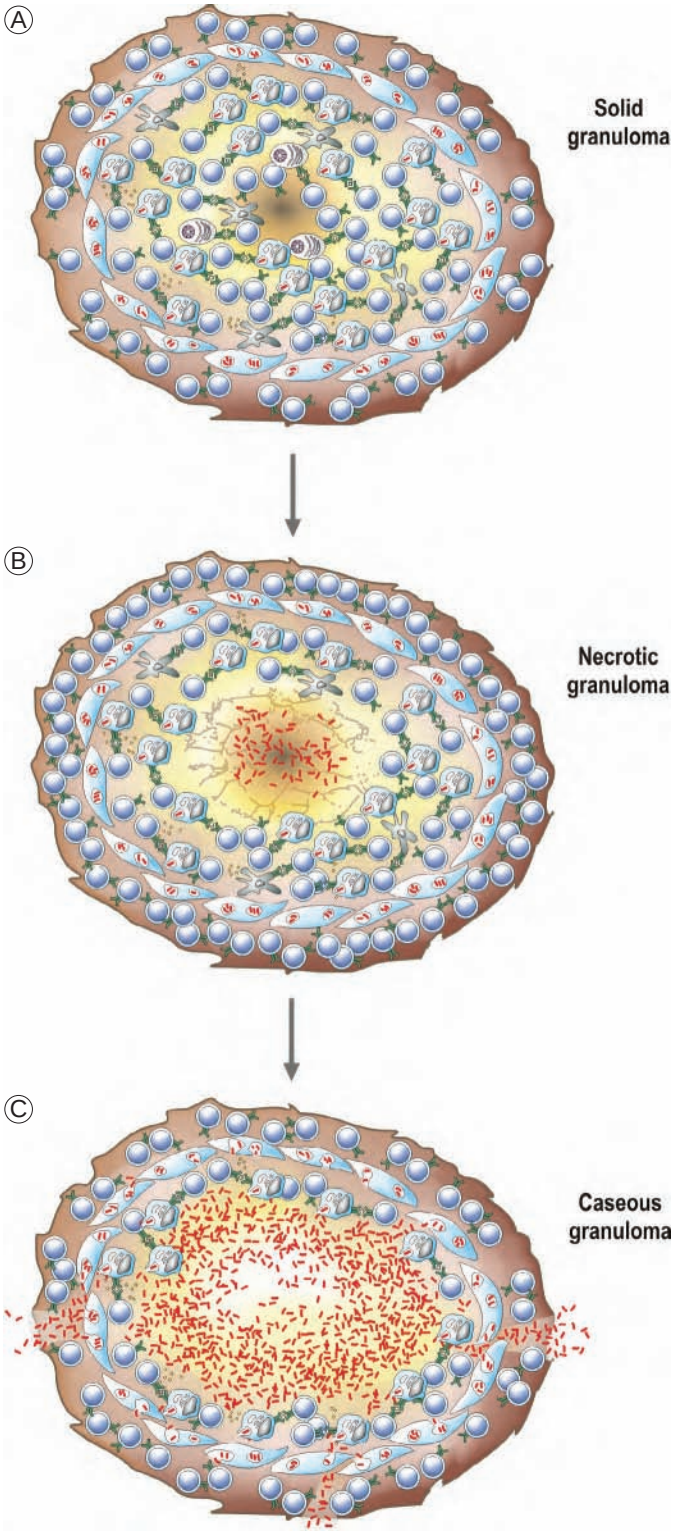


FIG. 26.1 Development of granuloma pathology and implications for tuberculosis (TB). This figure depicts three distinct yet continuous stages of granuloma pathology in the lung due to *Mycobacterium tuberculosis* infection. (A) Solid granuloma: composed largely of T cells and infected and uninfected macrophages (MPs). These granulomas are defined by a lack of central necrosis and are likely representative of an ability to control *M. tuberculosis* replication. (B) Caseous/necrotic granulomas: these structures contain a central region of demarcated necrotic cell death. Bacteria are often detected within the caseous necrotic region and in proximal cells, notably MPs. Since calcified caseous granulomas containing few bacteria have been observed, development of central necrosis may be a consequence of antibacterial mechanisms resulting in sacrifice of host cells to contain infection. (C) Cavity formation: these structures result from inability of caseous granulomas to contain bacterial replication. The acellular necrotic region, containing a large number of extracellular bacteria, increases in size and can liquefy and empty into the lung airways, resulting in transmission of viable bacteria via cough. Therefore, granuloma formation is central to human-to-human spread of TB. Dissemination of bacteria through the bloodstream results in disease manifestation in other organs, such as meninges and urinary bladder.

targets macrophages where it must counteract numerous antimicrobial mechanisms operative in these cells (see below). Intracellular bacteria also use elaborate mechanisms to enter nonprofessional phagocytes by which they must subvert host endocytic processes that are normally engaged in traffic of cellular cargoes. In some cases, this provides a less hostile environment owing to their inability to efficiently mobilize antibacterial effector mechanisms. *Bartonella* spp., unique among intracellular bacteria, can enter red blood cells, allowing transmission via blood-sucking insect vectors. This represents a particularly advantageous niche, as red blood cells lack the machinery to drive the adaptive immune responses required for protection. Entry into nonphagocytic host cells requires bacteria to induce their own internalization. Bacteria that colonize the gastrointestinal tract (i.e., *L. monocytogenes* or salmonellae) or mucosal membranes of the urogenital tract (i.e., *C. trachomatis*) must mediate tight adhesion to the host cell membrane and be capable of mediating the uptake process. Broadly, two processes are utilized by bacteria to induce uptake into a nonphagocytic cell. The "zipper" mechanism is mediated by binding of a bacterial cell surface protein to a cognate receptor on the host cell membrane. *L. monocytogenes* entry into intestinal epithelial cells depends on engagement of InIA to E-cadherin to mediate uptake.

Salmonellae and *C. trachomatis* use a "trigger" mechanism to induce internalization and inject multiple factors into the host cell cytoplasm to mediate uptake. These proteins are delivered by the needle-like structures that form part of bacterial T3SSs. These injected proteins target host proteins involved in host cell signaling and actin remodeling to induce bacterial entry. The *C. trachomatis*–secreted proteins Tarp, CT166, and CT694 reversibly stimulate the Rho-family GTPase Rac1 to trigger internalization. Similarly, salmonellae inject T3SS factors to stimulate Rho-family GTPases Cdc42 as well as Rac1. The success of these mechanisms of induced uptake enables intracellular bacteria to persist inside diverse cell types. *Rickettsia* spp., *C. trachomatis, M. leprae*, and *L. monocytogenes* ultimately target

vascular endothelial cells, epithelial cells, Schwann cells, and hepatocytes, respectively, as their preferred intracellular habitats.

To prevent intracellular infection, the host depends on its ability to discriminate between host and bacterial molecules. As already mentioned, bacteria targeting the intracellular environment often do so via mucosal surfaces already populated by commensal organisms (the microbiome) that do not alert host defenses. The host must therefore discriminate between commensal and pathogenic bacteria via recognition of conserved molecular motifs of bacteria, namely pathogen-associated molecular patterns (PAMPs). This occurs via host receptors broadly defined as pattern recognition receptors (PRRs, Table 26.2 Chapter 3).

The best-characterized group of PRRs are the so-called Toll-like receptors (TLRs). The TLR system constitutes an innate scanning mechanism of microbial pattern recognition to distinguish between a wide spectrum of bacteria and viruses. TLRs are present as homo- or heterodimers on the plasma membrane or within the intracellular endosome/phagosome compartment. PAMPs of bacterial origin comprise di- and tri-acylic lipoproteins, LPSs, and flagellin, which are recognized by TLR2/6, TLR2/1, TLR4/4, or TLR5/5, respectively. The vast array of mycobacterial cell wall lipids such as lipoarabinomannan (LAM), trehalose dimycolate (TDM), and phosphatidyl inositol mannosides (PIMs) bind either TLR2 or TLR4. Lipoteichoic acid (LTA) of gram-positive bacteria is recognized by TLR2. TLR9 binds low-methylated bacterial DNA containing CpG motifs within endosomes.

Scavenger receptors and C-type lectins are also PRRs and function at the cell membrane. Scavenger receptors were first defined by their ability to transport modified forms of low-density lipoproteins inside cells, indicating their ability to also interact with host molecules. However, receptors such as SR-A, MARCO, CD36, LOX-1, and SREC can bind a wide array of bacterial molecules such as lipids, CpG DNA, and proteins (see Table 26.2 for binding specificities). SR-A is important for clearance of extracellular bacteria from the spleen and liver. MARCO expressed on AMs is implicated in clearance of pneumococcal bacteria, preventing pneumonia. C-type lectins are similarly membrane-expressed and include DC-specific intercellular adhesion molecule-3-grabbing nonintegrin (DC-SIGN), mannose receptor, dectin-1, dectin-2, which chiefly recognize fungal components and MINCLE, which recognizes TDM, the cord factor of *M. tuberculosis*.

It has been suggested that whereas scavenger receptors and C-type lectins are required to bind and internalize the bacillus, it is primarily the TLRs that discriminate between the pathogens and initiate the necessary intracellular signaling events. It should, however, be noted that intracellular signaling events can also be triggered by other interactions, such as ligand binding to macrophage mannose receptor (MMR), dectin 1, or DC-SIGN. Far from a one-ligand, one-receptor binary mechanism of sensing and signaling, PRRs often collaborate to produce multiprotein complexes. CD14, MD2, and TLR4 collaborate for LPS sensing and signaling. Similarly, MARCO and TLR2 synergize to recognize TDM. To allow signaling, these complexes interact with adaptor proteins containing immunoreceptor tyrosine-based activation motif (ITAM)-like or Toll/interleukin-1 receptor (TIR) domain motifs. TLR signaling occurs via the adaptor proteins MyD88, TIRAP/Mal, and Trif. These molecules then orchestrate a downstream signaling cascade culminating in induced patterns of gene transcription that mediate innate and, ultimately, adaptive immune mechanisms that aim at combating the intracellular bacteria.

TABLE 26.2 Major Pattern Recognition Receptors Involved in Sensing of Intracellular Bacteria

PRR	Location	Ligand
Toll-like Receptors		
TLR1	Plasma membrane	Triacyl lipoprotein
TLR2	Plasma membrane	PGA, porins, LAM
TLR4	Plasma membrane	LPS
TLR5	Plasma membrane	Flagellin
TLR6	Plasma membrane	Diacyl lipoprotein
TLR7(human TLR8)	Endosome	ssRNA
TLR9	Endosome	CpG DNA
Scavenger Receptors		
SR-A	Plasma membrane	LPS, LTA, CpG DNA, proteins
MARCO	Plasma membrane	LPS, proteins
CD36	Plasma membrane	Diacyl lipoprotein
LOX-1	Plasma membrane	Protein
SREC	Plasma membrane	Protein
C-type Lectins		
DC-SIGN	Plasma membrane	LPS, ManLAM, capsular polysaccharide
MINCLE	Plasma membrane	Mycobacterial cord factor: TDM
NOD-like receptors		
NOD1	Cytoplasm	D-glutamyl-meso-diaminopimelic acid
NOD2	Cytoplasm	MDP
NLRP1	Cytoplasm	MDP
NLRP3	Cytoplasm	RNA, LPS, LTA, MDP
NLRC4	Cytoplasm	Flagellin
Naip5	Cytoplasm	Flagellin
AIM2-like Receptors		
AIM2	Cytoplasm	dsDNA
IFI16	Cytoplasm	dsDNA
STING/cGAS Pathway		
cGAS	Cytoplasm	dsDNA

We omit PRRs (e.g., TLR3, which binds viral-produced double-stranded RNA) not classically associated with intracellular bacteria.
AIM2, absent in melanoma 2; *CD36*, cluster of differentiation 36; *CpGDNA*, cytosine-phosphatidyl-guanine DNA; *DC-SIGN*, dendritic cell-specific intercellular adhesion molecule-3-grabbing nonintegrin; *dsDNA*, double-stranded DNA; *LAM*, lipoarabinomannan; *LOX-1*, lipoxygenase1; *LPS*, lipopolysaccharide; *LTA*, lipoteichoic acid; *ManLAM*, mannose lipoarabinomannan; *MARCO*, macrophage receptor with collagenous structure; *MDP*, muramyl dipeptide; *MINCLE*, macrophage-inducible C-type lectin; *NLR*, NOD-like receptor; *NOD*, nucleotide-binding oligomerization domain; *PGA*, peptidoglycans; *PRR*, pattern recognition receptor; *SR*, scavenger receptor; *SREC*, scavenger receptor expressed by endothelial cell-I; *ssRNA*, single-stranded RNA; *TDM*, trehalose dimycolate; *TLR*, Toll-like receptors.

The cellular cytoplasm is monitored for the presence of molecules of bacterial origin by a further group of PRRs, the nucleotide oligomerization domain protein-like receptors (NLRs). These molecules are characterized by a nucleotide-binding domain and leucine-rich repeat motifs. Molecules from this group recognizing bacterial components are nucleotide-binding oligomerization domain (NOD)-containing proteins NOD1 and NOD2, NOD-like receptors (NLR)P1, NLRP3, and Naip5. Other cytosolic PRRs include the absent-in-melanoma-2-like (AIM2) receptor family (ALR), cGMP-AMP synthase (cGAS), and stimulator of IFN genes (STING), all of which can be activated by bacterial DNA (see Table 26.2).

Engagement of NLRs and ALRs leads to activation of the multiprotein complex called the inflammasome, leading to cleavage of pro-IL-1β and pro-IL-18 to their active forms. In addition, activation of the NLRs NOD1 and NOD2 results in inflammatory cytokine secretion.[20] Certain PRRs are also receptive to certain endogenous "danger" signals produced by tissues undergoing stress, damage, or cell death. These signals are triggered by self-proteins, named danger-associated molecular patterns (DAMPs), and include endogenous heat shock proteins, host nucleotides, and the chromatin component HMGB1. Therefore, PRRs mediate signals not only emanating from intracellular bacteria but also from host cells damaged by the infection process. Understanding how PAMP and DAMP PRR signaling meshes to produce a coherent disease-specific output remains an exciting challenge for future research.

As already noted, the culmination of PRR collaborative sensing and signaling is the induction of inflammatory genes that induce innate immune mechanisms and, subsequently, the mobilization of the adaptive immune response. These include cytokines that act both locally and systemically and are important mediators of protection against intracellular bacteria via specific signaling through engagement of host cell surface receptors. Such engagement mobilizes both critical mechanisms of host protection and orchestration of adaptive immune responses.

Cytokines as Mediators of Defense Against Intracellular Bacteria

We have already mentioned that a range of cytokines are induced by the signaling mechanisms that result from engagement of PRRs. These serve by both enhancing intracellular mechanisms of bacterial killing and mobilizing adaptive immune responses, representing the next layer of host defense.[21] Because these responses allow an amplification of the initial innate immune responses, they must be carefully regulated by the host to prevent extensive tissue pathology. In fact, we might view the development of a granuloma as the sequela of a balance between bacterial killing mechanisms and the need to restrict tissue pathology orchestrated by adaptive immunity. At the onset of infection, initial cytokine secretion occurs in the cell type that first encounters intracellular bacteria and on initiation of signaling cascades by PRRs. These molecules can act locally and systemically to directly instruct cells, to produce antibacterial molecules, to combat intracellular infection, and both to increase numbers of immune cells and to direct the composition of the cellular infiltrate that will ultimately attempt to resolve intracellular bacterial infection.[22] Cytokines are ultimately produced by multiple cell types, including adaptive T cells, B cells, unconventional T cells, MPs, DCs, PMNs, and even epithelial and endothelial cells. We will first consider the hierarchy by which these cytokines act in the control of intracellular bacterial infection and the antibacterial mechanisms they regulate. We will then return to the generation and regulation of the cells that produce them.

KEY CONCEPTS

T-Cell-Mediated Mechanisms Underlying Protection

Interferon-gamma (IFN-γ)- and tumor necrosis factor (TNF)-mediated activation of phagocytes to kill bacteria by means of:

- Reactive oxygen intermediate (ROI) and reactive nitrogen intermediate (RNI)
- Delivery of lysosomal hydrolytic enzymes and antimicrobial peptides to the bacteria-containing phagosome
- Xenophagy
- Formation and maintenance of granulomas
- T-cell-mediated response controls, but does not eradicate, the pathogen

IFN-γ, TNF, IL-12, and IL-18

The cytokine with the clearest demonstrable potency against intracellular bacteria is IFN-γ. Extensive studies on the activation of antibacterial effector functions in macrophages have revealed a central role for IFN-γ. Accordingly, IFN-γ neutralization with antibodies or deletion of the IFN-γ gene by homologous recombination markedly exacerbates infectious diseases such as listeriosis, TB, or typhoid in experimental animals. For macrophages harboring intracellular bacteria, namely *M. tuberculosis*, signaling with IFN-γ is a game-changer, summoning infected macrophages to escalate an mechanisms. The action of TNF appears to augment IFN-γ and is also important in control of intracellular infection. This has been demonstrated in humans through use of blocking of TNF by antibodies as anti-inflammatory therapy. Such treatments can activate TB in individuals with LTBI. Despite this, these potent protective effects of IFN-γ and TNF come at a price. The need to kill intracellular bacteria often leads to death of the host cell as collateral damage. In part, the host manages this by controlling how the cell dies; excessive TNF leads to less-regulated necrotic cell death, benefiting *M. tuberculosis*. For this reason, elaborate host mechanisms have evolved to maintain TNF at optimum levels to control infection. The host enzyme leukotriene A4 hydrolase (LT4H) catalyzes synthesis of a highly proinflammatory lipid leukotriene B4. In the event of enzyme deficiency, an anti-inflammatory lipid lipotoxin A4 accumulates, which counteracts effects of TNF. Two common variant promoters control expression of *LT4H* in humans, and homozygotes are associated with either high or low inflammation. In contrast, the heterozygotes show a balanced response to TNF associated with resistance against TB. Such a finding strongly hints that genetic mechanisms can maintain an optimum level of TNF responsiveness of cells harboring intracellular bacteria, namely *M. tuberculosis*.

A central antimicrobial mechanism stimulated by IFN-γ and TNF is production of reactive nitrogen intermediates (RNIs) via the induction of nitric oxide synthase (NOS)2 and reactive oxygen intermediates (ROIs) via activation of reduced nicotinamide adenine dinucleotide phosphate (NADPH)-dependent oxidative burst. IFN-γ also promotes antimicrobial effects associated with vitamin D and induces autophagy, a mechanism that plays an important role in host defense. It is now clear that the production of IFN-γ depends on prior activation by IL-12 and/or IL-18. IL-12 in concert with TNF induces a cytokine loop resulting in the production of IFN-γ, which sustains the production of IL-12 and IL-18. These observations have been extended to humans where mutations that effect IFN-γ signaling cause susceptibility to *M. tuberculosis* and salmonellae, as well as BCG and commonly nonpathogenic mycobacteria termed Mendelian susceptibility to mycobacterial disease (MSMD). These mutations are in genes that include *IL12B* and *IL12RB*, which encode subunit β of IL-12 cytokine and its receptor respectively, and *IFNGR1* and *IFNGR2*, which encode the IFN-γ receptor.[23] Further unraveling of the molecular basis of human genetic susceptibility to intracellular bacterial infection will continue to illuminate our understanding of how immunity is similarly orchestrated across multiple infectious diseases.

Proinflammatory Cytokines and Phagocyte Attraction

The recruitment of more phagocytes to the site of infection represents a vital process in the resolution of infection. Phagocyte recruitment is achieved via the secretion by MPs and endothelial cells of cytokines of the IL-1 family, TNF-α, IL-6, and chemokines. Signaling via IL-1 cognates is considered closely related to that of the TLRs due to close homology of the cytoplasmic domains of TLRs and IL-1 family receptors. The most studied member is IL-1β, which in synergy with chemokines and TNF increases the expression of adhesion molecules on the vascular epithelium, thereby promoting extravasation of the inflammatory cell infiltrate into infected tissues. Chemokines are a family of structurally related proteins (Chapter 15). The positions of the first two cysteine residues in the protein sequence have been used to divide chemokines into four subfamilies: CC (MIP-1β, MCP-1, -2, -3), CXC (MIP-2, IL-8), C (lymphotactin), and CX3C chemokines (fractalkine), where C represents cysteine and X represents any amino acid other than cysteine. These molecules are critical in controlling the migration of PMNs (IL-8) and monocytes (MCP-1, also known as CCL2) from the bloodstream to infected tissue. Recently, the role of chemokines in intracellular infections has been increasingly appreciated, e.g., with mice lacking the receptor for CCL2 being deficient in their ability to clear listeria infection. It has been suggested that in the early stages of infection, *M. tuberculosis* exploits a delay in the mobilization of T-cell immunity to recruit MPs to the site of infection, which preferentially serve as habitat due to a lack of local IFN-γ from T cells. Moreover, *M. tuberculosis* is thought to infect the lower airways with a limited microbiome. This scarcity could mean that *M. tuberculosis* uses cell surface phenolic glycolipid (PGL) to signal epithelial cells to produce the chemokine CCL2 in the absence of signaling via other PAMPs. Phthiocol of *M. tuberculosis* is sensed by the aryl hydrocarbon receptor (AhR) to produce CCL2 and other chemokines for attraction of blood-derived phagocytes.[24] This mechanism then recruits MPs that are more permissive for bacterial growth than those recruited by a more "global" MyD88-dependent signaling of TLRs, requiring co-engagement of PAMPs on commensal bacteria that are more abundant in the upper airways. The initial MP infiltrates could play an important role in early granuloma development.

Cytokine-Induced Host-Protective Mechanisms

Effector Molecules

Activation of a membrane-bound NADPH oxidase by stimulation with IFN-γ or IgG initiates an oxidative burst that generates the ROI, O_2^-, H_2O_2, OH^-, $^1O_2$, and •OH radical (Table 26.3). In human PMNs and blood monocytes that possess myeloperoxidase, ROI activity is further augmented by the formation of hypochlorous acid. Oxidation and/or chlorination of bacterial lipids and proteins results in their inactivation and subsequent bacterial killing. The importance of ROIs in antibacterial defense is underlined by recurrent infections in patients whose phagocytes fail to generate an oxidative burst (Chapter 39). NOS2 is an inducible cytosolic enzyme in professional phagocytes that delivers NO to the phagolysosome-harboring bacteria while consuming O_2 and L-arginine. NO is further oxidized to NO_2^- and NO_3^-. Nitrification and/or oxidation then inactivates bacterial molecules needed for bacterial growth.[25] The formation of NO is catalyzed by NOS2, which is promoted by both immunological stimuli such as IFN-γ and TNF, and microbial products such as LPS, LTA, and mycobacterial lipids. RNIs exert their bactericidal activity by destroying iron-/sulfur-containing reactive centers of bacterial enzymes and by synergizing with ROIs to form highly reactive peroxynitrite ($ONOO^-$). Despite being highly effective in killing intracellular bacteria, NO production relies on a continuous supply of L-arginine, which becomes limited due to competition with another macrophage enzyme, Arginase-1 (Arg-1). Arg-1 metabolizes L-arginine to produce urea and ornithine and demonstrates anti-inflammatory activity. The competitive function of Arg-1 likely regulates collateral tissue damage caused by over-exuberant RNIs. The final downstream product of NOS2 activity is citrulline, which is recycled to L-arginine by the enzymes argininosuccinate synthase (Ass1) and argininosuccinate lyase (Asl). A mouse deficient in macrophage Asl activity is unable to control mycobacterial infection, highlighting the importance of this recycling pathway. The central role of NOS2 in protection against intracellular bacteria is well established in murine models of infection. Whether NOS2 plays a similarly central role in humans is still unclear. Defensins are small lysosomal polypeptides that are microbicidal at basic pH and are particularly abundant in phagocytes. These include granulysin, present in granules of human natural killer (NK) and cytolytic T lymphocytes (CTLs), and cathelicidin, which is regulated by vitamin D in a TLR-dependent manner and is converted by cleavage to the antimicrobial peptide LL-37.

Apoptosis and Autophagy

Apoptosis is a highly regulated form of cell death that is critical for control of cell turnover, which is a vital process for tissue homeostasis.[26] Macrophage apoptosis also constitutes a defense

TABLE 26.3 Antibacterial Effector Mechanisms of Activated Macrophages and Corresponding Microbial Evasion Strategies

Macrophage Effector Mechanism	Microbial Evasion Strategy
Production of ROIs	Uptake via complement receptors; production of ROI detoxifying molecules (superoxide dismutase, catalase); bacterial ROI scavengers (phenolic glycolipids, sulfatides, lipoarabinomannans)
Production of RNIs	Inhibition of phagosome maturation via blockage of H^+ATP pump; indirect effect of ROI-detoxifying molecules
Autophagy, intraphagolysosomal killing	Egression into cytoplasm; resistant cell wall
Phagosomal acidification, phagosome–lysosome fusion	Inhibition of phagosome maturation
Defensins	Modification of cell wall lipid A to resist defensins
Reduced iron supply (transferrin receptor downregulation, lipocalins)	Expression of microbial siderophores to increase iron uptake
Tryptophan degradation	Upregulation of bacterial tryptophan synthesis

RNI, Reactive nitrogen intermediate; ROI, reactive oxygen intermediate.

mechanism, allowing removal of phagocytes containing intracellular bacteria without the need to generate significant inflammation. Apoptosis, in contrast to cellular necrosis, results in cell death without permeabilization of the host cell membrane (Chapter 17). The process can be triggered by TNF signaling and augmented by IFN-γ, resulting in activation of cellular caspases, mitochondrial membrane permeability, and cytochrome *c* release. These processes result in cellular disintegration and generation of apoptotic bodies that are engulfed and digested by neighboring phagocytic cells. Apoptosis is protective against *L. monocyogenes* and *Salmonella* spp. and is inhibited by *M. tuberculosis*, which promotes necrotic cell death of infected cells to its benefit via mitochondrial membrane damage and by caspase-independent mechanisms during conditions of high bacterial burden in macrophages. Noninfected cells engulf bacterial antigens associated with vesicles produced by apoptotic cells. Apoptosis as a prerequisite for this pathway is induced by many intracellular bacteria, such as salmonellae, mycobacteria, and listeriae. This cross-presentation pathway in infections with intracellular bacteria adds an essential function to the physiological role of apoptosis in maintenance of tissue integrity and growth.

Upon signaling via IFN-γ, autophagy (a process common to all cells for removal of dysfunctional or damaged cellular organelles) can be harnessed to dispose of intracellular *L. monocytogenes* and *M. tuberculosis* in a process termed xenophagy. Signaling via members of the immunity-related GTPase family (IRG family) and the guanylate-binding protein family, TLR2 and 4 engagement and the active form of vitamin D_3 all act to augment xenophagy. This process is also triggered via the cGAS-STING (cyclic GMP-AMP sythase-stimulator of interferon genes). STING senses cyclic GMP-AMP generated by cGAS on binding bacterial DNA and leading to engagement of the autophagic machinery. Formation of double-membrane autophagosomes that mature analogously to the phagosomal pathway fuse with lysosomes that degrade bacteria contained within. The importance of this process is highlighted by polymorphisms in one of the three IRG family genes in humans, *IRGM*, being associated with susceptibility to TB. Recently, a host-encoded microRNA, miRNA-155 was shown to potentiate xenophagy during intracellular mycobacterial infection by targeting an endogenous inhibitor of autophagy, Ras homologue enriched in brain (Rheb).

Nutrient Deprivation

Deprivation of required nutrients to intracellular bacteria is also a strategy employed by the host, markedly so within infected macrophages. Tryptophan degradation is achieved by the enzyme indolamine 2, 3-deoxygenase (IDO), which degrades tryptophan to kynurenine (see Table 26.3). This reaction is induced by IFN-γ in both MPs and IFN-γ-responsive nonprofessional phagocytes and inhibits the growth of *C. psittaci* and *C. trachomatis* inside human macrophages and epithelial cells. Similarly, augmentation of NOS2 by IFN-γ and TNF depletes intracellular L-arginine, also required for growth of intracellular bacteria.[19]

Evasion From, Interference With, and Resistance to Microbial Killing

Strategies Against Toxic Effector Molecules

Many intracellular bacteria have exploited successful strategies against macrophage effector mechanisms (see Table 26.3). One mechanism of evasion is determined by the receptor that is used for pathogen entry into the host cell. Internalization via complement receptors (CRs) inhibits the production of IL-12, a cytokine critical in facilitating macrophage activation. Engulfment by this receptor also bypasses activation of the oxidative burst, thereby avoiding ROI production. Similarly, engaging MMR and DC-SIGN for uptake triggers secretion of the suppressive cytokines IL-10 and TGF-β. Several intracellular bacteria also produce ROI detoxifiers, including superoxide dismutase and catalase, which nullify O_2 and H_2O_2, respectively. Finally, a number of small bacterial products, such as phenolic glycolipid, phthiocol, and LAM of mycobacteria or phenazines of *P. aeruginosa*, scavenge ROIs. Many of the strategies used to counteract the effects of ROIs also overlap in their effects on RNIs. A modification of lipid A renders gram-negative bacteria, including salmonellae, resistant to the effects of host antimicrobial peptides.

Intraphagosomal Survival

Inhibition of phagolysosome fusion represents a major intracellular survival strategy for a number of intracellular bacteria, including *M. tuberculosis*, *Francisella* spp., *Brucella* spp., and *L. monocytogenes* (Fig. 26.2). After engulfment, these pathogens manipulate the endocytic fate of the phagosome that contains them. This is achieved in part by manipulation of Rab GTPases (proteins required for normal endocytic trafficking) positioned in the phagosome membrane. Rab GTPases associated with phagosome maturation of the pathogen-containing phagosome are Rab 3, 4, 5, 9, 7, 11, and 14. These proteins are associated with different maturation stages of the phagosome and chiefly orchestrate membrane fusion events to allow delivery of vesicular protein cargo to the phagosomal compartment. The mycobacteria-containing phagosome acquires Rab5a, but not the late endosomal marker Rab7a, which ultimately mediates fusion of the bacterial-containing phagosome with lysosomes that contain proteolytic enzymes active at low pH. By enabling arrest of this maturation, *M. tuberculosis* maintains its compartment at an early endosomal stage. This compartment does not acidify, due in part to a paucity of vacuolar H^+ ATPase; at the same time, it exchanges molecules with the plasma membrane such as the transferrin receptor to access iron. Activation of macrophages with IFN-γ restores the normal maturation of the mycobacterial phagosome, resulting in a drop in mycobacterial viability. Francisellae and brucellae are engulfed by phagosomes that acquire the early endosomal markers EEA1 and Rab5a. The Francisella-containing vacuole acquires late endosomal markers, but the pathogen escapes into the cytosol by perforating the late endosomal membrane. After a transient phagosomal stage, brucellae enter compartments enclosed by endoplasmic reticulum (ER) membranes to escape delivery to phagolysosomes.

Phenotypic Plasticity of the Infected Cell

The ability of intracellular bacteria to influence the phenotypic fate of both the cell in which it dwells and cells within lesions that form due to unresolved infection is becoming increasingly appreciated. This has already been highlighted by the tendency of *M. tuberculosis* to use its own membrane lipids to exploit host chemotactic pathways to recruit bacterial growth-permissive macrophages to the site of infection, subsequently allowing a proliferative head start before adaptive immunity kicks in, and amplifies intracellular defense by the action of cytokines such as

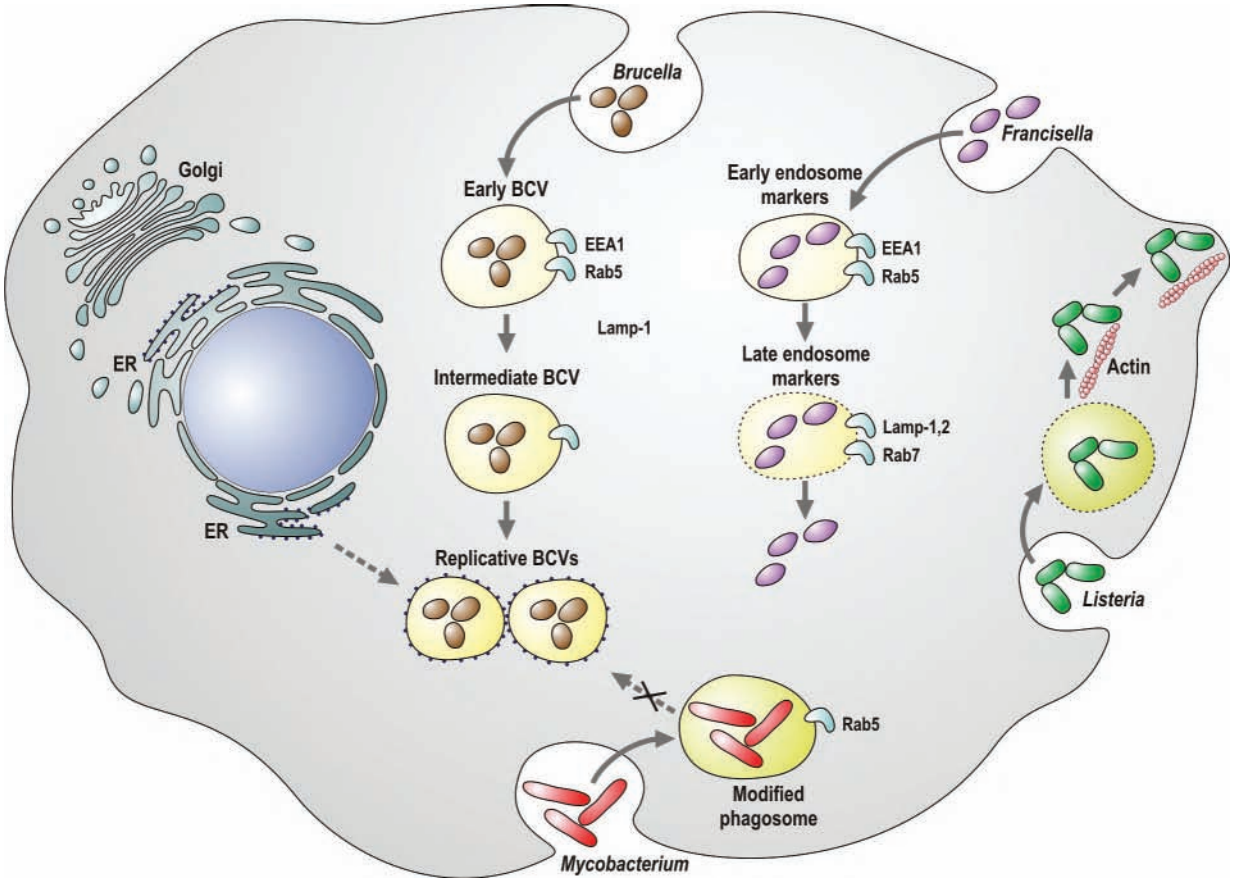


FIG. 26.2 Inhibition of Phagolysosome Fusion Represents a Major Survival strategy for a number of Intracellular Bacteria, Including *Mycobacterium tuberculosis*, *Francisella* spp., *Brucella* spp., and *Listeria monocytogenes*. The mycobacteria-containing phagosome acquires Rab5, but not the late endosomal markers Lamp-1 and 2, enabling arrest of maturation of this compartment at an early endosomal stage. *Francisella* spp. and *Brucella* spp. are engulfed by phagosomes that acquire the early endosomal markers EEA1 and Rab5. The *Francisella*-containing vacuole acquires late endosomal markers, but escapes into the cytosol by perforating the late endosomal membrane. A similar strategy is adopted by *L. monocytogenes*. After a transient phagosomal stage, brucellae enter compartments enclosed by endoplasmic reticulum (*ER*) membranes to escape delivery to phagolysosomes. Brucella-containing vacuole (*BCV*).

IFN-γ and TNF. Moreover, bacterial killing must be tempered inside the granuloma to prevent destruction of host tissue. This is achieved by balancing the macrophage phenotype ranging from highly bactericidal, termed "classically" activated (also termed M1 macrophages), to a phenotype that is more suppressive of inflammation and associated with wound healing, tissue remodeling, and fibrosis, termed "alternatively" activated (also called M2 macrophages). Tipping the balance one way or the other is detrimental for the host in terms of disease.[27]

Myeloid-derived suppressor cells (MDSCs) represent a certain stage of development of myeloid cells (both of monocytic and granulocytic lineage). Although most of our knowledge stems from their suppressive role in cancer, recent evidence suggests that they play a role in control of chronic infection such as TB. They can be distinguished from canonical MPs and granulocytes by means of distinct surface markers. The granulocytic MDSCs are $CD11b^{+}$ $LY6G^{hi}$ $Gr1^{int}$, whereas the monocytic MDSCs are $CD11b^{+}$ $LY6G^{neg}$ $LY6C^{+}$ $Gr1^{hi}$.

During intracellular infection of Schwann cells, *M. leprae* downregulates genes active for the Schwann cell phenotype and upregulates genes that orchestrate differentiation to a "stem-cell-like" phenotype. This stem-cell-like property allows the infected cell to differentiate further to multiple mesenchymal cell states, such as skeletal cells or smooth muscle cells.[28] This ability to regress and then re-program infected cell phenotype could play a role in spreading infection throughout the host during leprosy.

Recently, both mesenteric stem cells (MSCs) and hematopoietic stem cells (HSCs) have been identified as intracellular niches of *M. tuberculosis* in mice and humans. Because these cells predominantly reside in hypoxic niches in the bone marrow and most antimycobacterial therapies are inactive in these conditions, it is possible that MSCs and HSCs maintain intracellular bacteria during long-term infection and represent a protective niche from drug therapy. *M. tuberculosis* was also detected in long-term repopulating pluripotent HSCs, and these cells could also be capable of resuscitating active disease.

Training of Innate Immunity

Vaccination with BCG induces IFN-γ-dependent effects on HSCs and their more immediate progeny, multipotent progenitors (MPPs), leading to production of myeloid cells that contain epigenetic changes and that are more resistant to intracellular infection by *M. tuberculosis*.[29] Moreover, BCG vaccination

was shown to induce changes in patterns of methylation of the NOD2 gene in humans, and these epigenetic changes led to increased production of proinflammatory cytokines. Furthermore, this enhanced innate immunity caused increased resistance not only against *M. tuberculosis* but also against other bacterial pathogens. Indeed, it has been speculated that epigenetic alterations in MPs following vaccination are responsible for the nonspecific reduction of mortality in BCG-vaccinated infants in resource-poor regions.

Effects of bacterial infection on myelopoiesis can also mediate resistance to sepsis in neonates. Neonates produce high systemic levels of the alarmins S100/A8 and S100/A9. In model systems, these alarmins effect myelopoiesis by inducing a more tolerogenic monocyte phenotype that attenuates hyperresponsiveness to bacterial colonization while not affecting intracellular bacterial killing mechanisms. Therefore, alarmins in neonates could reduce the risk of sepsis in the first year of life by generation of monocytes less likely to trigger a hyperinflammatory response to a bacterial infection.

Escape into Cytoplasm

A successful strategy for survival inside activated macrophages is egression from the phagosome into the cytoplasm, which has been exploited by *L. monocytogenes* and the various pathogenic *Rickettsia* spp. (see Fig. 26.2). This has the advantage of both avoiding the cellular defense mechanisms within the phagosome and providing the bacteria with a nutrient-rich environment. *L. monocytogenes* possesses several virulence factors to facilitate its escape from the phagolysosome, a pore-forming hemolysin (listeriolysin, LLO) that acts together with a metalloproteinase, a lecithinase, and two phospholipases to efficiently promote the rupture of the phagosomal membrane, as well as spreading to other cells. To avoid collateral damage, LLO contains an amino acid sequence that initiates its own destruction soon after it has entered the cytosol. *M. tuberculosis* and *M. leprae* can also egress from the phagosome into the cytoplasm of macrophages and DCs, a behavior that is mediated by mycobacterial protein secretion system ESX-1. Bacterial virulence factors secreted by ESX-1 may also contribute to increased cell death and are not expressed by the vaccine BCG, which does not escape from the phagosome.

T Lymphocytes as Specific Mediators of Acquired Resistance

Whereas activated macrophages act as the nonspecific executors, T lymphocytes are the specific mediators of acquired resistance against intracellular bacteria. The dramatic increase in the incidence of TB and other intracellular bacterial infections in AIDS patients illustrate the central role of T lymphocytes in protection. For instance, 15 million individuals are co-infected with HIV and *M. tuberculosis*, and HIV increases the risk of developing TB by several orders of magnitude, resulting in one million TB cases annually. At the site of microbial growth, T lymphocytes not only initiate the most potent defense mechanisms available but also focus this response to the site of encounter, thus minimizing collateral damage to the host. Although protective T-cell responses are multifactorial, they can be reduced to a few principal mechanisms (Fig. 26.3).

As previously mentioned, T cells inevitably also produce pathology through cytotoxic antimicrobial defense mechanisms. Moreover, pathogenesis of intracellular bacterial infection is highly influenced by T cells. It is therefore important that the T-cell response be tightly controlled and downregulated when necessary. Regulatory mechanisms, including regulatory T (Treg) cells, are in place to limit immunopathology.[16]

Protective immunity involves so-called conventional T-cell sets, CD4 αβ T cells, and CD8 αβ T cells, as well as unconventional T cells, such as γδ T cells, CD1-restricted αβ T cells, and T cells that recognize antigen in the context of other nonclassical MHC class I molecules, such as mucosal-associated invariant T (MAIT) cells (see Fig. 26.3). Although these T-cell sets perform different tasks, substantial redundancy exists. Furthermore, these T-cell populations act in a coordinated way in close interaction with other leukocytes. Depending on the etiological agent and the stage of disease, the relative contribution of the different T-cell subsets to acquired resistance may vary. The conventional αβ T cells make up more than 90% and γδ T cells less than 10% of all lymphocytes in the blood and peripheral organs of humans and mice. However, γδ T cells represent a significant proportion of the intraepithelial lymphocytes in mucosal tissues, suggesting a particular role at this important port of microbial entry.[30]

CD4 T Cells

The CD4 T-cell population can be further subdivided into distinct subsets, according to their pattern of cytokine production and expression of unique transcription factors that control patterns of gene expression. At least four major subsets exist: Th1, Th2, Th17, and Treg. The first two subsets were discovered several decades ago and are identified in both mice and humans; Th1 cells overwhelmingly produce IFN-γ and IL-2, and Th2 cells produce IL-4, -5, and -13. The Th1 subset can also be defined based on the T-bet transcription factor and the signal transducer STAT4, while Th2 classification is consistent with expression of the transcription factor GATA-3 and signal transducer STAT5.

Th17 cells express the retinoid orphan receptor γT (RORγT) transcription factor and the signal transducer STAT3. They produce the cytokines IL-17, IL-22, and GM-CSF. Cytokines of the IL-17 family are strong inducers of granulopoiesis, proinflammatory mediators such as IL-6, and the chemokines CXCL1, CXCL8, and CXCL6, which attract neutrophilic and eosinophilic granulocytes and prolong their survival.[31] Th17 cells have limited importance for protection in murine models against primary infection with mycobacteria, salmonellae, and listeriae. However, Th17 cells can drive more rapid Th1 responses against pulmonary TB in mice after vaccination, resulting in enhanced protection. IL-17 is also required for optimally protective Th1 responses during murine *F. tularensis* infection.[31]

Despite the convenience of defining T-cell populations in terms of subsets, recent evidence suggests considerable plasticity in cytokine production by T cells. This was first suggested by demonstration that all subsets could produce IL-10, which regulates potency of T-cell responses to limit host collateral damage during immune responses. IL-10 expression might be an intrinsic control mechanism common to all T cells. However, reduction in T-cell potency also favors chronic intracellular bacterial infection. T-cell subsets may acquire the ability to produce additional cytokines by expression of additional transcription factors or by remodeling chromatin structure.

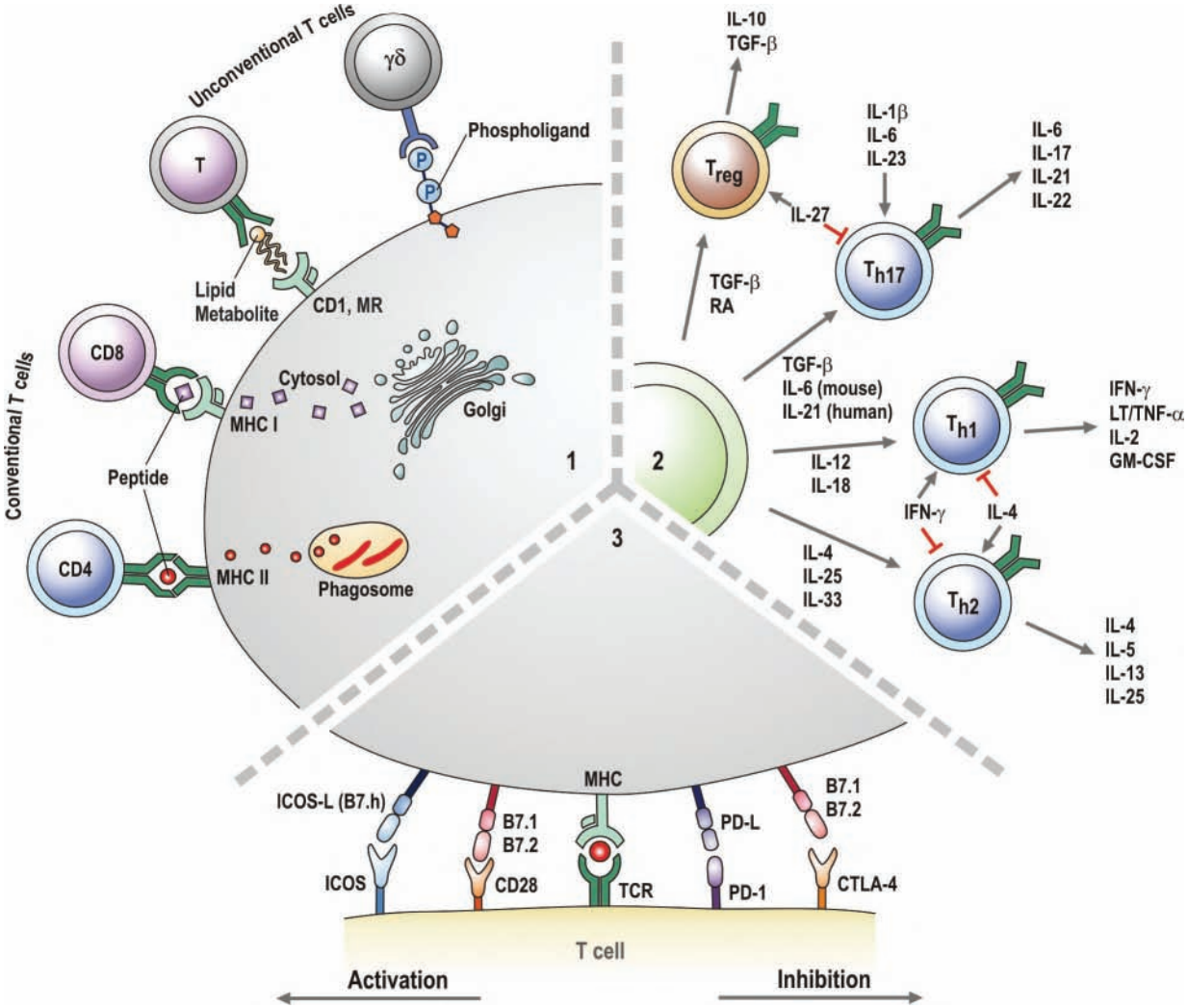


FIG. 26.3 T-Cell Stimulation During Infection. Recognition of bacterial antigen by T cells (A). Antigen originating from intracellular bacteria is presented to conventional CD4 and CD8 T cells. Unconventional T cells including gamma-delta (*γδ*) T cells, mucosal-associated invariant T (MAIT) cells, and CD1-restricted T cells are also activated. Human γδ T cells recognize small molecules containing pyrophosphate residues; MAIT cells recognize bacterial metabolites such as vitamin B_2 derivatives in the context of major histocompatibility complex (*MHC*)-related (*MR*) gene products; CD1-restricted T cells recognize glycolipids in the context of CD1 molecules. (B) CD4 T cells can be subdivided into different T helper (*Th*) cells according to their cytokine expression pattern. Th1 cells are critical for protection against intracellular bacteria; they typically produce interferon-gamma (*IFN-γ*), tumor necrosis factor-alpha (*TNF-α*), lymphotoxin (*LT*), interleukin (*IL*)-2, and granulocyte–macrophage colony-stimulating factor (*GM-CSF*). Th2 cells stimulate humoral immune responses via secretion of IL-4 and IL-5. Other cytokines produced by Th2 cells include IL-13 and IL-25. Th17 cells produce IL-6, IL-17, IL-21, and IL-22, which probably contribute to early protection. Regulatory T cells (*Treg*) produce transforming growth factor beta (*TGF-β*) and IL-10, which suppress immune responses. Additional abbreviations: inducible co-stimulatory molecule (*ICOS*); programmed death 1 (*PD-1*); program death ligand (*PD-L*); retinoic acid (*RA*); T-cell receptor (*TCR*). See Chapters 6, 10, and 14 for details. (Modified from Kaufmann SHE, Parida SK. Tuberculosis in Africa: learning from pathogenesis for biomarker identification. *Cell Host Microbe*. 2008;4[3]:219–228, with permission of Elsevier.)

CD8 T Cells

Infection of mice deficient in specific T-cell subsets has conclusively demonstrated a role for CD8 T cells during listeriosis and TB. Furthermore, CD8 effector T cells have been identified in granulomas of tuberculoid leprosy patients that contain low numbers of bacteria. The cytolytic potential of these T cells can serve two roles in infection with intracellular bacteria: namely, target cell killing or lysis of cells that are unable to control the infection, thus releasing the bacteria for phagocytosis by more efficient cells. In humans, CD8 T-cell-mediated killing is cell-contact-dependent and based on production of perforin, granzymes, and granulysin (Chapter 12).[32] Finally, CD8 T cells are also a potent source of IFN-γ and TNF, thus contributing to direct activation of infected macrophages to enhance protective mechanisms (see Fig. 26.3).

CD8 T cells recognize antigenic peptides in the context of MHC class I gene products, which are responsible for presentation of antigens residing in the cytosol (Chapter 9).

Initially, therefore, it was mysterious how CD8 T cells were stimulated by intracellular bacteria, which were thought to have a uniquely restricted phagosomal residence. However, with the knowledge that many intracellular bacteria egress into the cytoplasm, one major mechanism for MHC class I processing became obvious: proteins secreted by bacteria in the cytoplasm undergo antigen processing and presentation similar to newly synthesized proteins of viral or host origin.[33] Yet, alternative contact points for MHC class I molecules and bacterial peptides exist. Cross-presentation by noninfected antigen-presenting cells (APCs) of antigens engulfed within apoptotic blebs from infected cells represents a critical pathway to induce CD8 T cells by phagosomal bacteria (see below). One major advantage of CD8 T cells over CD4 T cells is their recognition of antigen bound by MHC class I gene products, which are expressed by almost all host cells. Thus, CD8 T cells recognize professional and nonprofessional phagocytes equally well.

Unconventional T Cells

The relevance of the γδ T cells to antibacterial immunity is not fully understood. Several studies indicate that γδ T cells are rapid producers of IL-17 at sites of bacterial implantation. Transient participation of γδ T cells in protection and a unique requirement for γδ T cells in granuloma formation have been described for murine listeriosis and TB. While murine γδ T cells appear to recognize peptides presented by nonpolymorphic MHC class I molecules, human γδ T cells respond to nonpeptidic phosphorylated metabolites, notably from the isoprenoid pathway of bacterial and host origin.

MAIT cells are primarily localized at mucosal sites and current evidence suggests that they play a role in control of bacterial infections in mucosal tissues such as lung (*M. tuberculosis*) and gut (gram-negative bacteria). Antigenic ligands include derivatives of vitamin B_2 (riboflavin), produced by many intracellular bacteria, including salmonellae and mycobacteria.[34]

CD1 comprises a group of nonpolymorphic MHC-related molecules that can present glycolipid antigens to unconventional T cells. In humans, group 1 CD1-restricted T cells respond to a variety of microbial glycolipids, including LAM, PIMs, mycolic acids, sulfatides, sulfoglycolipids, and lipopeptides. Group I CD1 molecules are absent in mice. The group II CD1 molecule CD1d is present in both humans and mice and controls development of NKT cells that express the NK cell marker NK1.1 and an invariant T-cell receptor. Upon antigen activation, these T cells rapidly produce cytokines and are capable of producing both IL-4 and IFN-γ. Bacterial antigens recognized by NKT cells, PIMs from mycobacteria, and glycosphingolipids from *Ehrlichia* and *Sphingomonas* spp. have been identified. NKT cells also respond to host endogenous lysosomal lipids loaded onto CD1d.[35]

In sum, unconventional T cells often recognize nonpeptidic ligands of bacterial origin, emphasizing that they play a particular role in immunity against bacteria, including intracellular bacteria. Because of the highly skewed T-cell receptor, these T cells are specific for a limited variety of bacterial ligands. As a consequence of their less-demanding antigen recognition and activation requirements, unconventional T cells may fill a gap between prompt innate resistance and the delayed conventional T-cell response.

T-Cell Memory and Regulation of Immune Responses

Long-term protective immunity against infectious agents relies on immune memory, which principally forms the basis for the success of vaccines (Chapter 87). Memory T cells can be divided into central memory T cells (TCM) and effector memory T cells (TEM), based on differential surface phenotypic and tissue migration patterns.[36] TEM accumulate in peripheral tissues where they express effector functions, while TCM persist in lymph nodes where they rapidly develop into TEM after secondary antigen encounter. The tissue-resident memory T cells (TRM) are localized to mucosal sites where they provide efficient protection against invading pathogens. Although it is generally accepted that memory T cells can survive in the absence of persistent antigen, little is known about the induction and maintenance of long-lasting T-cell memory in chronic infections with intracellular bacteria.

B Cells

Historically, antibody is presumed to be protective largely when the bacillus is extracellular and, as such, has been perceived unlikely to play a central role in intracellular bacterial infection. However, recent evidence is beginning to challenge this assumption and suggests a role in protection against TB. Antibodies from healthcare workers exposed to *M. tuberculosis* have been shown to be protective, and this protection was abrogated in the absence of CD4 cells.[37] Furthermore, the Fc portion of *M. tuberculosis*-specific antibodies from individuals with latent TB showed a specific pattern of glycosylation, resulting in enhanced binding of FcγRIIIa. This receptor mediates antibody-dependent cell cytotoxicity by NK cells and indicates that this mechanism may play a role in preventing development of active TB. Moreover, individuals that are exposed but do not become infected show IgA and class-switched IgG to major antigens of *M. tuberculosis* with characteristic patterns of Fc glycosylation. This indicates that these antibodies may prevent infection completely, leading to a so-called "resistor" immune status.[38] Aside from antibody production, B cells are potent APCs for soluble antigens (including lipids presented by CD1c) and secrete many cytokines otherwise associated with T cells, DCs, and macrophages. B-cell signaling via MyD88 during *S. typhimurium* infection has been associated with B-cell production of IL-10, and mice with B-cell-specific deficiency in MyD88 were more resistant to infection demonstrating that, like T cells, a subset of B cells can perform regulatory functions. These B cells, termed regulatory B cells (Breg), function via production of IL-10, IL-35, and TGF-β, and have been shown to suppress immunity to intracellular bacteria through cognate interactions between Breg and CD4 T cells via MHC-II.[39]

From the therapeutic perspective, there is increasing interest in using antibody therapies as checkpoint inhibitors to potentiate T-cell responses to clear chronic bacterial infection. Inducible T-cell co-stimulation (ICOS) and programmed cell death protein 1/PD ligand 1 (PD-1/PDL-1) are molecular checkpoint targets whose inhibition has been therapeutically useful in cancer settings. In the case of TB, there remains controversy as to whether expression of these targets is associated with protection against or enhancement of disease.

In humans, the proportion of both $PD\text{-}1^+CD4^+$ and $PD\text{-}L1^+CD4^+$ T cells in TB patients are significantly increased compared to healthy controls, and blockade of PD-1 significantly enhances $CD4^+$ T-cell proliferation. Phagocytosis and

intracellular killing activity of macrophages increased significantly with PD-1/PD-L pathway blockade.[40] Despite this, recent case studies on patients enrolled in trials of anti-PD1 therapy for cancer showed exacerbation of latent TB accompanied by increased numbers of TB-specific CD4 T cells in the blood.[41]

KEY CONCEPTS

How Might a Vaccine Work?

- Activation of innate immunity for instruction of appropriate acquired immune response (pattern recognition receptors)
- Activation of the appropriate array of T-cell populations
- T-cell secretion of appropriate cytokine combination
- Efficient development of T-cell memory
- Efficient activation of antibacterial effector mechanisms
- Best-case scenario: sterile pathogen eradication
- Second best-case scenario: driving the infection "deeper" into latency, thus efficiently preventing reactivation of disease
- Alternative scenario: prevention of stable infection by antibodies and T cells

Regulatory T Cells

Intracellular bacteria can cause detrimental inflammation and tissue damage, e.g., due to IFN-γ- and TNF-mediated Th1 responses. Under normal circumstances, control mechanisms are in place to limit immunopathology; such countermeasures are elicited as part of the ongoing immune response during infection. The main cytokines, which limit inflammation and control IFN-γ production, are IL-10 and TGF-β. Although macrophages and DCs produce these cytokines, their main producers are Treg cells, which are the prime cells involved in immune regulation. Natural Treg cells are responsive to IL-2 due to high constitutive CD25 expression and are characterized by expression of the transcription factor FOXP3 (Chapter 13).[42] Expansion of Treg cells appears to be both antigen-dependent and -independent. In addition, Treg cells selectively express TLRs and can be activated, for example, by LPS and possibly other TLR ligands. This makes their immediate activation during bacterial infection a probable scenario. Although Treg cells limit CD8 T-cell responses in experimental listeriosis, their general role in infections with intracellular bacteria has not been fully elucidated. Suppression of T-cell responses and anergy have been clinically documented for TB and leprosy. Although Treg functions can limit detrimental T-cell responses and immunopathology, they can also prevent elimination of bacteria, and hence are likely a key factor in promoting the persistent chronic state of infection with intracellular bacteria.

Innate Lymphoid Cells

Innate lymphoid cells (ILCs) are a family of cell types that share features of both innate and adaptive immune cells. They typically locate to mucosal tissues where they orchestrate early barrier-protective responses to infection.[43] There are three main subsets: ILC1, which produce IFN-γ and include NK cells and non-cytotoxic, non-NK type I ILCs; ILC2, which produce IL-4, IL-5, and IL-13 and are involved in inflammatory-linked airway hyperactivity, tissue repair, and helminth clearance; and ILC3, which produce IL-17 and/or IL-22 and are involved in the structure of ectopic lymphoid tissue. Circulating ILC precursors can differentiate into all ILC types, rapidly developing into a diverse ILC population in response to environmental cues. ILCs are depleted in the blood during TB, but ILC1 and ILC3 rebound in numbers after treatment. In mouse models of TB, ILC3s accumulate in the lung early after infection with TB concomitantly with AMs, which precedes infiltration of monocytes and macrophages. Increased numbers of ILC3 in the lung are associated with early-stage protection against TB via IL-17 and IL-22 production, accumulation of AMs, and orchestration of ectopic lymphoid structures in the lung.[44]

CONCLUDING REMARKS

ON THE HORIZON

- Design of predictive biomarkers to prognose risk of progression to active tuberculosis (TB).
- Development and clinical testing of new drugs for MDR/XDR-TB and dormant *Mycobacterium tuberculosis*.
- Developing host-directed therapy (HDT) in adjunct to conventional drugs, notably for highly drug-resistant pathogens.
- Development and clinical testing of new vaccines that protect against pulmonary TB in adults.
- Reduction of the unequal burden of TB in developing and industrialized countries by public health measures, education, and socioeconomic advances.

Our deepening understanding of the molecular events governing intracellular bacterial infections is allowing development of novel therapeutic and preventive approaches. The need for such interventions is becoming all the more pronounced in the face of increasing levels of antibiotic resistance of bacteria, such as *M. tuberculosis*, rendering canonical drugs that specifically target bacterial molecular processes ineffective. Host-directed therapy (HDT), an approach for which there is accumulating activity, aims to develop new drugs or repurpose previously approved ones directed at host molecular processes.[45] Such approaches include monoclonal antibodies to neutralize cytokines such as TNF or IL-6 to abrogate tissue destructive inflammation, repurposed use of licensed drugs such as ibuprofen and verapamil to modulate inflammation and enhance antibiotic effectiveness, respectively, and use of immunostimulatory molecules such as vitamin D_3 to enhance bacterial killing by xenophagy. Metformin was developed as an oral treatment for type 2 diabetes mellitus and functions as an insulin anti-sensitizer. It can also enhance treatment of TB by augmenting formation of ROS species, and it shows promise as an adjunct to TB chemotherapy. These approaches range from preclinical and early-stage clinical development to late-stage clinical development and offer promise to shorten the traditional duration of therapy.

It is becoming increasingly recognized that the interaction between intracellular bacteria and the immune system is not of the "all or nothing" type but is instead a "continuous struggle." This realization has far-reaching implications for preventive and therapeutic strategies against intracellular bacterial infections. First, vaccination against intracellular bacteria has not yet been effected satisfactorily because of the involvement of several distinct T-cell subsets with different modes of stimulation and activity profiles. Second, chemotherapy has frequently proved suboptimal for the sterile eradication of bacteria hidden in cellular niches. A better understanding of the complex crosstalk between cytokines, T lymphocytes, macrophages, and infected host cells will no doubt directly promote the development of improved control measure.

ACKNOWLEDGMENTS

We are grateful to Souraya Sibaei for excellent assistance and Diane Schad for the figures.

Due to space considerations, we have been unable to include several pertinent references for some of the concepts we have described. We apologize to those researchers whose important work we were unable to cite for this reason.

REFERENCES

1. Sarode G, Sarode S, Anand R, et al. Epidemiological aspects of leprosy. *Dis Mon*. 2020;66(7):100899. https://doi.org/10.1016/j.disamonth.
2. Horne D, Skerrett S. Recent advances in nontuberculous mycobacterial lung infections. *F1000Res*. 2019;8:F1000.
3. Simpson H, Deribe K, Tabah EN, et al. Mapping the global distribution of Buruli ulcer: a systematic review with evidence consensus. *Lancet Glob Health*. 2019;7:e912–e912.
4. GBD 2017 Typhoid and Paratyphoid Collaborators. The global burden of typhoid and paratyphoid fevers: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet Infect Dis*. 2019;19:369–381.
5. Desai AN, Anyoha A, Madoff LC, Lassmann B. Changing epidemiology of *Listeria monocytogenes* outbreaks, sporadic cases, and recalls globally: a review of ProMED reports from 1996 to 2018. *Int J Infect Dis*. 2019;84:48–53.
6. Yagupsky P, Morata P, Colmenero JD. Laboratory diagnosis of human brucellosis. *Clin Microbiol Rev*. 2019;33(1): e00073–19.
7. Rawla P, Limaiem F. Lymphogranuloma venereum. In. *StatPearls*. StatPearls Publishing; 2020.
8. Limmathurotsakul D, Golding N, Dance DAB, et al. Predicted global distribution of *Burkholderia pseudomallei* and burden of melioidosis. *Nat Microbiol*. 2016;1:15008.
9. Snowden J, Simonsen KA. Tularemia. In *StatPearls*. StatPearls Publishing; 2020.
10. Dooling KL, Toews KA, Hicks LA, et al. Active bacterial core surveillance for legionellosis—United States, 2011–2013. *MMWR Morb Mortal Wkly Rep*. 2015;64:1190–1193.
11. Di Pietro M, Filardo S, Romano S, Sessa R. *Chlamydia trachomatis* and *Chlamydia pneumoniae* interaction with the host: latest advances and future prospective. *Microorganisms*. 2019;7(5):140.
12. Blanton LS. The rickettsioses: a practical update. *Infect Dis Clin North Am*. 2019;33:213–229.
13. Winslow GM, Bitsaktsis C. Immunity to the ehrlichiae: new tools and recent developments. *Curr Opin Infect Dis*. 2005;18:217–221.
14. Deng H, Pang Q, Zhao B, Vayssier-Taussat M. Molecular mechanisms of bartonella and mammalian erythrocyte interactions: a review. *Front Cell Infect Microbiol*. 2018;8:431.
15. Holland TL, Arnold C, Fowler VG. Clinical management of *Staphylococcus aureus* bacteremia: a review. *JAMA*. 2014;312:1330–1341.
16. Dorhoi A, Reece ST, Kaufmann SHE. For better or for worse: the immune response against *Mycobacterium tuberculosis* balances pathology and protection. *Immunol Rev*. 2011;240:235–251.
17. Esmail H, Riou C, du Bruyn E, et al. The immune response to *Mycobacterium tuberculosis* in HIV-1-coinfected persons. *Annu Rev Immunol*. 2018;36:603–638.
18. Horcajada JP, Montero M, Oliver A, et al. Epidemiology and treatment of multidrug-resistant and extensively drug-resistant *Pseudomonas aeruginosa* infections. *Clin Microbiol Rev*. 2016;32(4):e00031–19.
19. Kaufmann SHE, Dorhoi A. Molecular determinants in phagocyte-bacteria interactions. *Immunity*. 2016;44:476–491.
20. Tan X, Sun L, Chen J, Chen ZJ. Detection of microbial infections through innate immune sensing of nucleic acids. *Annu Rev Microbiol*. 2018;72:447–478.
21. Slight SR, Khader SA. Chemokines shape the immune responses to tuberculosis. *Cytokine Growth Factor Rev*. 2013;24:105–113.
22. Netea MG, Schlitzer A, Placek K, et al. Innate and adaptive immune memory: an evolutionary continuum in the host's response to pathogens. *Cell Host Microbe*. 2019;25:13–26.
23. Bustamante J. Mendelian susceptibility to mycobacterial disease: recent discoveries. *Hum. Genet*. 2020;139(6–7):993–1000. https://doi.org/10.1007/s00439-020-02120-y.
24. Moura-Alves P, Faé K, Houthuys E, et al. AhR sensing of bacterial pigments regulates antibacterial defence. *Nature*. 2014;512:387–392.
25. Bogdan C. Nitric oxide synthase in innate and adaptive immunity: an update. *Trends Immunol*. 2015;36:161–178.
26. Nagata S. Apoptosis and clearance of apoptotic cells. *Annu Rev Immunol*. 2018;36:489–517.
27. Murray PJ. Macrophage polarization. *Annu Rev Physiol*. 2017;79:541–566.
28. Masaki T, Qu J, Cholewa-Waclaw J, et al. Reprogramming adult Schwann cells to stem cell-like cells by leprosy bacilli promotes dissemination of infection. *Cell*. 2013;152:51–67.
29. Kaufmann E, Sanz J, Dunn JL, et al. BCG Educates hematopoietic stem cells to generate protective innate immunity against tuberculosis. *Cell*. 2018;172(176–90):e19.
30. Godfrey DI, Uldrich AP, McCluskey J, et al. The burgeoning family of unconventional T cells. *Nat Immunol*. 2015;16:1114–1123.
31. Lin Y, Ritchea S, Logar A, et al. Interleukin-17 is required for T helper 1 cell immunity and host resistance to the intracellular pathogen *Francisella tularensis*. *Immunity*. 2009;31:799–810.
32. Walch M, Dotiwala F, Mulik S, et al. Cytotoxic cells kill intracellular bacteria through granulysin-mediated delivery of granzymes. *Cell*. 2014;157:1309–1323.
33. Cruz FM, Colbert JD, Merino E, et al. The biology and underlying mechanisms of cross-presentation of exogenous antigens on MHC-I molecules. *Annu Rev Immunol*. 2017;35:149–176.
34. Coulter F, Parrish A, Manning D, et al. IL-17 Production from T helper 17, mucosal-associated invariant T, and γδ cells in tuberculosis infection and disease. *Front Immunol*. 2017;8:1252.
35. Darmoise A, Teneberg S, Bouzonville L, et al. Lysosomal alpha-galactosidase controls the generation of self lipid antigens for natural killer T cells. *Immunity*. 2010;33:216–228.
36. Lanzavecchia A, Sallusto F. Understanding the generation and function of memory T cell subsets. *Curr Opin Immunol*. 2005;17:326–332.
37. Li H, Javid B. Antibodies and tuberculosis: finally coming of age? *Nat Rev Immunol*. 2018;18:591–596.
38. Lu LL, Smith MT, Yu KKQ, et al. IFN-γ-independent immune markers of *Mycobacterium tuberculosis* exposure. *Nat Med*. 2019;25:977–987.
39. Maerz JK, Trostel C, Lange A, et al. Bacterial immunogenicity is critical for the induction of regulatory B Cells in suppressing inflammatory immune responses. *Front Immunol*. 2019;10:3093.
40. Shen L, Gao Y, Liu Y, et al. PD-1/PD-L pathway inhibits M.tb-specific CD4+ T-cell functions and phagocytosis of macrophages in active tuberculosis. *Sci Rep*. 2016;6:38362.
41. Barber DL, Sakai S, Kudchadkar RR, et al. Tuberculosis following PD-1 blockade for cancer immunotherapy. *Sci Transl Med*. 2019;11(475):eaat2702.
42. Belkaid Y, Rouse BT. Natural regulatory T cells in infectious disease. *Nat Immunol*. 2005;6:353–360.
43. Masopust D, Soerens AG, Tissue-resident T Cells and other resident leukocytes. *Annu Rev Immunol*. 2019;37:521–546.
44. Ardain A, Domingo-Gonzalez R, Das S, et al. Group 3 innate lymphoid cells mediate early protective immunity against tuberculosis. *Nature*. 2019;570:528–532.
45. Kaufmann SHE, Dorhoi A, Hotchkiss RS, Bartenschlager R. Host-directed therapies for bacterial and viral infections. *Nat Rev Drug Discov*. 2018;17:35–56.

27

Host Defenses to Extracellular Bacteria Including Spirochetes

Sarah W. Satola, Marcos C. Schechter, Shyra Wilde, and David S. Stephens

The human host has developed protective mechanisms to interact with the multitude of bacterial species encountered in nature. These host defenses include nonspecific mechanisms of clearance as well as innate and specific adaptive immune responses. Partly because of these mechanisms, most bacterial species do not cause human disease. Many bacterial species have established symbiotic or commensal relationships with the human host and colonize skin and mucosal surfaces. These commensals are generally of low virulence except in individuals whose host defenses are compromised. Despite the diversity of the microbial world, a relatively few pathogenic bacterial species or subpopulations of those species have evolved virulence factors or strategies that can overcome or circumvent intact human host defense mechanisms to cause localized or systemic disease.

Bacterial pathogens of clinical importance reside mostly extracellularly (Table 27.1) and can be transmitted from one individual to another by close contact or transmitted through food, water, animal, or other environmental contact. Acquisition of these pathogenic bacteria may be transient, resulting in intervals of colonization varying from asymptomatic carriage to localized or systemic disease. Extracellular bacterial pathogens can produce acute inflammatory and purulent infectious diseases, such as meningitis, septicemia, pneumonia, urethritis, pharyngitis, inflammatory diarrhea, cellulitis, and abscesses, and/or produce disease by the release of toxins. Disease associated with some extracellular bacteria (e.g., *Helicobacter pylori*) results from chronic colonization. Susceptibility to extracellular bacterial pathogens is enhanced by hereditary, acquired, or age-related defects in innate or adaptive host defenses. Resistance to extracellular bacterial pathogens or their toxins can be accentuated by chemoprophylaxis vaccines, and other immune modulation processes (e.g., passive immune globulin administration). Caution is urged in the interpretation of the term "extracellular." The classification of bacteria as "extracellular" and "intracellular" is primarily based on observations in vitro and has been challenged by some authors, as some "extracellular" bacterial species invade host cells as a part of their normal life cycle and during steps in the disease process (e.g., *Staphylococcus aureus, Streptococcus pneumoniae, Streptococcus pyogenes, Neisseria meningitides*).[1] Conversely, bacteria typically classified as "intracellular" can have an extracellular component to their life cycle (e.g., *Mycobacterium tuberculosis* in cavitary lesions).

Spirochetes constitute a unique group of extracellular bacteria that inhabit many different environments, such as soil, arthropods, and mammals. Spirochetes share a typical spiral shape and a distinctive flat-wave morphology (Fig. 27.1). They are motile organisms with a multilayered outer membrane that encapsulates a peptidoglycan layer surrounding their inner membrane. The viability of the organism is dependent on an intact outer membrane, which can be damaged by variations in osmolarity, antibodies, or complement, resulting in the loss of intracellular components and ultimately death of the bacterium.

CLINICAL PEARLS

Distinguishing Clinical Characteristics of Infections With Extracellular Bacteria

- Sterilizing immunity
- Colonization of mucosal surfaces often precedes disease
- Causes of pyogenic infections
- T-helper (Th)17 response critical in generating a neutrophilic response
- Antibodies are protective for some of the major pathogens
- Effective vaccines available for many of the major pathogens

Several spirochetal species can induce disease (Table 27.2) including syphilis (*Treponema pallidum* subspecies *pallidum*, hereafter *T. pallidum*) and Lyme disease (*Borrelia burgdorferi* sensu stricto, hereafter *B. burgdorferi*). Syphilis is primarily sexually transmitted, whereas Lyme disease is transmitted by *Ixodes* complex ticks and is the most common tickborne disease in the United States.[2] *T. pallidum* cannot be cultured in the laboratory and *B. burgdorferi* culture is not routinely available. Thus, diagnoses for syphilis and Lyme disease are usually based on clinical presentation and serological tests.

Although dark-field microscopy can be used for the identification of *T. pallidum*, infection with *T. pallidum* leads to the production of nonspecific antibodies to cardiolipin-cholesterol-lecithin antigen, which are the basis for nontreponemal serological tests, including the Venereal Disease Research Laboratory (VDRL) and rapid plasma reagin (RPR) tests. Because these tests are nonspecific, false-positive reactions can occur as a result of pregnancy, autoimmune disorders, or infections. Treponemal-specific tests such as the *T. pallidum* hemagglutination test (TPHA) and fluorescent treponemal antibody-absorption test (FTA-ABS) are more specific and often used as confirmatory tests following a positive nontreponemal test. Of note, some laboratories are now using reverse-screening algorithms that start with a treponemal-specific test.

A two-tiered approach is standard for the serodiagnosis of Lyme disease: serum is first tested for *B. burgdorferi*–specific antibodies by enzyme-linked immunosorbent assay (ELISA) or immunofluorescent assay (IFA), followed by more specific immunoblotting for immunoglobulin M (IgM) and IgG antibodies. However, a positive serological test, particularly IgG, is evidence of exposure to *B. burgdorferi*, but not necessarily an active infection.

Despite similar ancestry and morphological features, these spirochetes have striking differences at the genetic level, which may account for the differences in their life cycles, environmental adaptations, and the diseases they cause. *B. burgdorferi* has one

TABLE 27.1 Examples of Clinically Relevant Pathogenic Extracellular Bacteria

Species	Examples of Human Disease	Selected Mechanisms of Pathogenesis	Special Features Key to Host Infection	Examples of Susceptible Populations/Risk Factors
Staphylococcus aureus	Cellulitis, abscesses, bacteremia, endocarditis, toxic shock syndrome, osteomyelitis, pneumonia, wound infections	Protein A: promotes fibronectin binding PVL: cytotoxic α-toxin: membrane damage eTSST-1: superantigen	Asymptomatic colonization, resistant to dehydration	Injection drug users, patients on hemodialysis, defects on Th17 response (Job syndrome), surgical procedures and skin trauma
Streptococcus pneumoniae (pneumococcus)	Pneumonia, otitis media, meningitis	Capsule: prevents phagocytosis, antigenic variation Pneumolysin: cytotoxic PspA & C: inhibition of complement Neuraminidase, hyaluronidase: spread and colonization IgA_1 protease	Asymptomatic colonization, readily acquires new genes through transformation	Smokers, cerebrospinal fluid leak, asplenia, hypogammaglobulinemia, human immunodeficiency virus/acquired immunodeficiency syndrome (HIV/AIDS), unvaccinated children
Streptococcus pyogenes (group A *streptococcus*)	Pharyngitis, cellulitis, erysipelas, toxic shock syndrome, necrotizing fasciitis, scarlet fever, rheumatic fever	Hyaluronic acid capsule, M protein: prevents phagocytosis Streptolysin O & S: cytotoxic Streptococcal pyrogenic exotoxins C5a peptidase	High diversity of M-proteins, molecular mimicry of human antigens	School-age children, crowded-conditions (e.g., military barracks), injury to lymphatic system (e.g., surgical harvest of saphenous vein)
Streptococcus agalactiae (group B *streptococcus*)	Neonatal sepsis, pneumonia and meningitis, perinatal infections, bacteremia	FbsA: fibrinogen receptor, promotes adherence Capsule β-Hemolysin C5a peptidase β Protein: downregulates complement	Asymptomatic colonization, acquisition by infants during birth	Neonates and infants (immunity dependent on passive transfer of maternal antibodies), diabetes mellitus
Neisseria meningitidis (meningococcus)	Meningitis, bacteremia (purpura fulminans)	Capsular polysaccharide: promotes adherence and prevents phagocytosis Type IV pili: promote attachment to host cells LOS: analogues to LPS, activates TLR4 pathway IgA_1 protease fHbp: downregulates the host alternative complement pathway	Molecular mimicry of human antigens, phase and antigenic variation, asymptomatic carriage	Terminal complement deficiencies, hypogammaglobulinemia
Neisseria gonorrhoeae (gonococcus)	Urogenital infections, disseminated gonococcal infection, pharyngitis	Type IV pili: promote attachment to host cells Opa protein adhesion IgA_1 protease LOS: analogs to LPS, activates TLR4 pathway	Phase and antigenic variation, molecular mimicry of human antigens	Terminal complement deficiencies, women during menstrual period (increases risk of dissemination)
Escherichia coli	Urinary tract infections, gastroenteritis, sepsis, neonatal meningitis	Capsular polysaccharide Tissue-specific fimbriae Heat-labile enterotoxins: increases intestinal chloride secretion LPS: activation of TLR4	Antigenic heterogeneity of LPS and capsule	Bladder instrumentation, pregnancy
Pseudomonas aeruginosa	Ventilator-associated pneumonia, bronchiectasis	Pili and flagella: attachment to the host and formation of biofilms LPS Exotoxin A Lipases, lecithinases, elastase	Considerable adaptability to changes in environment, large genome size, biofilms	Orotracheal intubation, cystic fibrosis
Clostridium difficile	Colitis	Toxin B: cytotoxic Flagella	Endospore formation, asymptomatic carriage	Antibiotics and other disruptions of the microbiota
Haemophilus influenzae	Otitis media, pneumonia, epiglottitis, bacteremia, meningitis	LPS with phosphorylcholine Pili: adherence Capsule High-molecular-weight adhesins IgA_1 protease	Phase variation of pili, asymptomatic carriage	Unvaccinated children, immunocompromised, sickle cell disease, smoking
Helicobacter pylori	Peptic ulcer disease	Urease: colonization of gastric mucosa Flagella: motile in gastric mucus CagA; bacteria-derived carcinogen	Polymorphism of CagA	Crowded living conditions, unreliable source of clean water, living with someone who has *H. pylori*
Bordetella pertussis	Whopping cough (children), chronic cough (adults)	Pertussis toxin: inhibits neutrophils, macrophages, lymphocytes Pertactin and filamentous hemagglutinin; mediates attachment	Antigenic variation of adhesins	Nonvaccinated adults and children, infants who have not completed vaccine series, adults and adolescents whose immunity has diminished

CapA, Cytotoxin-associated gene A; *fHbp*, factor H–binding protein; *Ig*, immunoglobulin; *LOS*, lipooligosaccharide; *LPS*, lipopolysaccharide; *Psp*, pneumococcal surface protein; *PVL*, Panton-Valentine leukocidin; *TLR*, Toll-like receptor; *TSST-1*, toxic shock syndrome toxin 1.

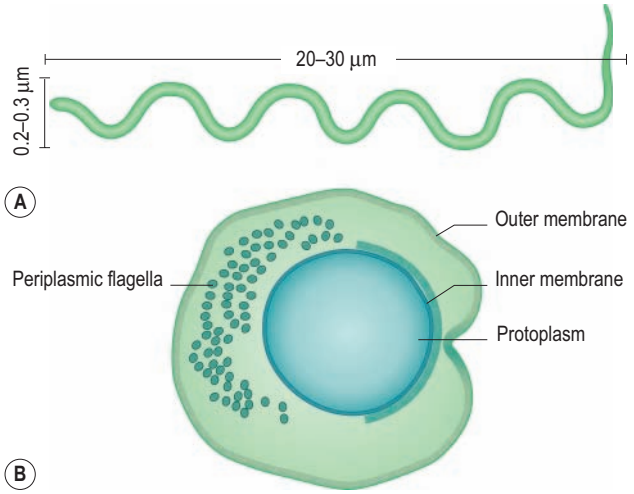


FIG. 27.1 The *Borrelia burgdorferi* structure is characterized by a distinctive flat-wave morphology consisting of approximately 18 bends and a length of 20 to 30 µm (A). A cross-section of this spirochete reveals the endoflagella, which are responsible for the unique morphology and motility of this organism (B).

CLINICAL PEARLS

Lyme Disease

- Erythema migrans at inoculation site (early stage)
- Joint inflammation in untreated individuals (often synovium of the knee)
- Chronic arthritis, neuroborreliosis, or cutaneous lesions (late stage)

Venereal Syphilis

- Hardened and painless ulcer (*chancre*) at initial infection site
- Secondary eruption often accompanied with rash on the palms of hands/soles of the feet. (4–6 weeks post infection); resolution of secondary manifestation weeks to a year
- Long periods of latency, late lesions of skin, bone, and viscera, cardiovascular system and CNS
- Susceptibility to *T. pallidum* universal; only 30% of exposures with lesions result in infection
- Infection results in gradual development of immunity against *T. pallidum* and often against heterologous treponemes as well

of the most complex genomes known among prokaryotes, with a single linear chromosome and 21 plasmids, the largest number of plasmids of any characterized prokaryote, and less than 10% of *B. burgdorferi* plasmid-coding regions are found in other microorganisms, including spirochetes.[3,4] In contrast to

TABLE 27.2 Major Diseases Caused by Spirochetes[a]

Disease	Agents	Distribution	Transmission	Symptoms
Lyme disease	*Borrelia burgdorferi*	North America, Europe	Tick engorgement	Development of a skin rash known as *erythema migrans*, accompanied by other symptoms, such as malaise, myalgia, and/or arthralgia. Symptoms can progress to include carditis and arthritis. Persistent infection can result in chronic arthritis, neuroborreliosis, or cutaneous symptoms (acrodermatitis chronica atrophicans)
	B. garinii	Asia, Europe		
	B. afzelii	Asia, Europe		
	B. andersonii	North America		
	B. japonica	Japan		
	B. lusitaniae	Southern Europe		
	B. valaisiana	Europe, Ireland, UK		
	B. mayonii	North America		
	B. miyamotoi	North America		
Relapsing fever	*B. hermsii*	Western USA	Tick engorgement	Clinical manifestations of infection include high-density spirochetemia, high fever, myalgias, and arthralgias and can even include cerebral hemorrhage and fatality
	B. turicatae	Southwestern USA, Mexico		
	B. parkeri	Western USA		
	B. mazzotti	Central America		
	B. venezuelensis	Central America		
	B. duttonii	Sub-Saharan Africa		
	B. crocidurae	North Africa, Middle East		
	B. persica	Middle East, Central Asia		
	B. hispanica	Iberian peninsula, North Africa		
	B. latyschewii	Iran, Iraq, Eastern Europe		
	B. caucasia	Iraq, Eastern Europe		
Venereal syphilis	*Treponema pallidum pallidum*	Worldwide	Sexual contact	Disease progresses from a primary lesion (chancre) to a secondary eruption and then to a latent period, and if left untreated, tertiary symptoms may appear
Endemic syphilis or Bejel syphilis	*T. pallidum endemicum*	Eastern Mediterranean region, West Africa	Nonsexual skin contact	Symptoms begin with a slimy patch inside the mouth, followed by blisters on the trunk and limbs. Bone infection in the legs soon develops, and in the later stages, lumps may appear in the nose and on the soft palate of the mouth
Yaws	*T. pertenue*	Humid equatorial countries	Nonsexual skin contact	Destructive lesions of the skin and bones, which is rarely fatal but can be debilitating
Pinta	*T. carateum*	Mexico, Central America, South America	Nonsexual skin contact	Dark-colored skin lesions found on those areas of the body that are exposed to sunlight. Eventually, the skin lesions become discolored
Leptospirosis	*Leptospira interrogans*	Worldwide	Urine from an infected animal	Symptoms include fever, headache, chills, nausea and vomiting, eye inflammation, and muscle aches. In more severe cases, the illness can result in liver damage and kidney failure

[a]Spirochetes are the causative agents of many diseases, which can have social as well as lasting health-related consequences.

other spirochetes, *B. burgdorferi* and *T. pallidum* do not contain lipopolysaccharide (LPS). Lipoproteins are the major immunogens of *B. burgdorferi* and most likely *T. pallidum*, and, thus, they are their dominant proinflammatory agonists.

CLEARANCE AND NONSPECIFIC HOST DEFENSES AT MUCOSAL EPITHELIAL SURFACES

Bacteria first encounter a physical barrier, which comprises skin, mucus and mucosal surfaces, and the normal microbiota as well as nonspecific factors, such as nutrient limitation (e.g., iron) and antimicrobial proteins or peptides (AMPs). Intact skin and mucosal surfaces provide complex chemical and biological obstacles to bacteria and are an important line of defense preventing the invasion of these pathogens and their products. As such, the human epithelium has evolved to prevent colonization and invasion.[3] Skin is a relatively dry, acidic (pH 5–6) barrier that contains growth-inhibiting fatty acids and AMPs, characteristics that are detrimental to many bacteria. The constant desquamation of stratified epithelial surface of skin helps in the removal of microorganisms.[4] Disruption of these physical barriers can augment pathogen tissue colonization and invasion. Infections by *S. aureus* and *S. pyogenes*, bacteria that can colonize skin, are often preceded by skin damage. Repeated trauma to skin (e.g., dialysis and intravenous drug use) also enhances skin colonization with pathogens, including that by *S. aureus*.

Mucosal surfaces have additional nonspecific antibacterial defenses. The mucociliary blanket of the respiratory tract and the female urogenital tract (fallopian tube) move bacteria away from epithelial surfaces, as does the flushing of the urinary tract with urine, intestinal peristalsis, and the bathing of the conjunctiva with tears. Lysozyme is found in most mucosal secretions and lyses bacterial cell walls by splitting muramic acid β(1–4)-*N*-acetylglucosamine linkages. The acid pH of the stomach, intestinal peristalsis, and the antibacterial effect of proteolytic enzymes present in intestinal secretions are important gastrointestinal (GI) tract host defenses against many pathogenic bacteria. The GI mucosa has a layer of mucus that acts as a physical shield to bacteria. Mucus is rich in mucin, glycoproteins that limit pathogen binding to other host molecules necessary for mucosal adhesion. Additionally, the mucus layer may function as more than a physical barrier by acting as a diffusion barrier to concentrate antimicrobial proteins at the appropriate epithelial cell surface. The glycocalyx, an extracellular layer of the apical surface of mucosal cells composed of carbohydrates, also protects cells against bacterial attachment.

Bacterial attachment and colonization of mucosal surfaces can be inhibited by bacterial binding to human cellular antigens present in secretions, such as ABO blood group antigens. Cell adhesion and extracellular matrix molecules, such as fibronectin and proteoglycans, can also inhibit or enhance bacterial binding to epithelial surfaces. The Tamm-Horsfall glycoprotein, found in urine, can bind avidly to a variety of bacteria and facilitate clearance. Proteins, such as lactoferrin, present at mucosal surfaces, bind iron, which is an important requirement for bacterial growth. This action may reduce microbial proliferation, but some mucosal pathogens bind lactoferrin and remove iron from the molecule for growth.

To colonize human epithelial and mucosal surfaces, extracellular bacteria must overcome the local host defense mechanisms described above. After navigating these defenses, adhesion to host cells is usually the first important step for extracellular bacterial pathogens (Fig. 27.2). Initial attachment of bacteria to human epithelial cells is, in part, mediated by pili, fimbriae, or other bacteria ligands or adhesins, and close adherence of bacteria to the human cell-surface receptors involves the cell wall, outer membrane proteins, lipopolysaccharide (LPS), and other bacterial surface structures. The attachment of bacteria to human epithelial cells prevents elimination of bacteria from the host. Attachment can also induce host cell pathways leading to cytoskeletal rearrangements, such as elongation and branching of the microvilli, the accumulation of actin, and calcium efflux, which facilitates close adherence and invasion of epithelial cells by normally "extracellular" bacteria, especially at sites with fluid movement. Strains of *Escherichia coli* that successfully colonize the bladder and cause renal infection possess pili that allow adhesion to the renal epithelium.[5,6] Type IV pili are fundamental for attachment of gonococci to the male reproductive tract and play a role in the attachment of meningococci to vascular endothelial cells.[5,7] Meningococcal pili also facilitate twitching motility and microcolony formation, which allows the penetration of mucus and provides initial attachment. The pneumococcal CbpA surface protein promotes mucosal adhesion and dissemination.[6] *B. burgdorferi* produce many adhesive surface proteins that collectively recognize diverse host substrates and cell types and are likely to promote dissemination and chronic infection in a variety of tissues. *B. burgdorferi*-endothelial interactions are mediated by the adhesin BBK32. *T. pallidum* protein Tp0751 (pallilysin) binds to host laminin, fibrinogen, fibronectin, and collagen, important constituents of blood and endothelia.[8]

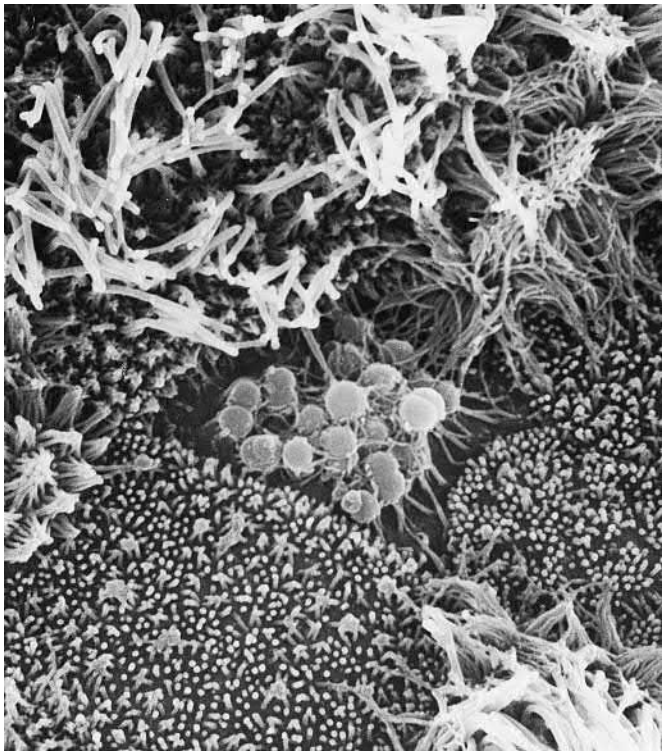


FIG. 27.2 Colonization and Adherence of Extracellular Bacteria at Mucosal Surfaces. Scanning electron micrograph of *Neisseria meningitidis* adherence and microcolony formation of a human upper respiratory mucosa (×16, 250).[46]

Bacteria utilize several mechanisms to avert the host immune response to bacterial surface antigens (see Table 27.1). Phase variation of adhesins is a mechanism of immune evasion common to

pathogenic *Neisseria* spp. Meningococcus, for example, utilizes phase variation of the adhesion protein Opa and type IV pili during the process of colonization of human upper respiratory mucosal surfaces.[9] Sialylation of LPS, a potent inducer of host inflammatory response, is an example of bacterial "hiding" of surface antigens. Sialylation of lipooligosaccharide, a molecule analogous to LPS, in meningococci has been shown to increase resistance to classical pathway (CP) and alternative pathway (AP) complement-mediated killing by decreasing the deposition of C3b and IgM on the cell surface, irrespective of capsular phenotype.

Normal Microbiota as Host Defense

The human microbiome is now recognized as a major host defense against bacterial pathogens by providing "colonization resistance," maintaining a balance of commensals to pathogens, and by priming the immune system.[10] Altering or disrupting the normal microbiota by antibiotics facilitates the expansion of enteric pathogens as *Clostridioides difficile* and *Salmonella typhimurium* or selection of antibiotic-resistant members of the microbiome. Similarly, changes in human physiology—for example, exposure of skin to elevated temperatures and humidity, chronic stress, host immune suppression, or active behavioral changes, such as smoking—can cause a commensal-to-pathogen switch. Recent studies have demonstrated that certain resident microbiota can resist pathogen colonization and infection. For example, matched volunteers were inoculated with *Haemophilus ducreyi* into the arms, and the subsequent infection either resolved or resulted in formation of abscesses; characterization of the skin microbiome before, during, and after the experimental inoculation showed that the microbiomes of those with pustule formation and of those with resolved infection were distinct and influenced the course of the *H. ducreyi* infection.[11]

The interaction of the microbiome with the immune system is also important for defense against extracellular pathogens. Normal microbiota prime the immune system by maintaining high levels of major histocompatibility complex (MHC) class II molecule expression on macrophages and other antigen-presenting cells. Pathogen recognition receptors (PRRs; see below) are traditionally known to recognize microbial molecules during infection; however, ligands for PRRs are abundantly produced by the resident microbiota during normal colonization. The integrity of the intestinal epithelial layer is dependent on activation of Toll-like receptors (TLRs; see below) by normal microbiota.[4] Stimulation of TLR-5 has been shown to increase resistance to *Enterococcus faecium* infection in a murine model. Activation of nucleotide-binding oligomerization domain 1 (NOD1) receptors by gut resident microbiota is necessary for priming of the innate immune system. Additionally, resident microbiota produce such factors as bacteriocins, lantibiotics, and phenol-soluble modulin (PSM), which function in a similar manner to that of host-derived AMPs (see below), suggesting an important host defense strategy against pathogen colonization.[12] Importantly, members of the resident microbiota can cause disease, particularly with loss of epithelial integrity and translocation to a different host tissue.

KEY CONCEPTS

Host Defenses and Immune Response at Epithelial Surfaces to Extracellular Bacteria

- Clearance and nonspecific host defenses at skin and mucosal surfaces:
  - Epithelial barriers
  - Antibacterial factors (fatty acids, antimicrobial peptides, lysozyme, phospholipase A_2)
  - Mucociliary activity
  - Normal microbiota
  - Adherence blocking molecules
- Specific immune defenses at mucosal surfaces:
  - Innate immune mechanism
  - Immunoglobulins
  - Phagocytosis at mucosal surfaces
- Mucosa-associated lymphoid tissue (MALT), gut-associated lymphoreticular tissue (GALT), bronchus-associated lymphoid tissue (BALT)

HOST DEFENSES TO EXTRACELLULAR BACTERIA INCLUDING SPIROCHETES

Early pathogen recognition by the innate immune system is key for rapidly responding to pathogens, allowing for the development of more advanced immunity mediated by T and B cells. Spirochete and extracellular bacterial pathogens' virulence is attributed, in part, to the evolution of their sophisticated tactics to evade killing mechanisms during all stages of the immune response, including serum complement and cellular immunity, as well as pathogen-specific antibodies. Both the innate and adaptive immune responses elicited by spirochetes and other extracellular bacteria are discussed, with the supposition that these responses are required for efficient bacterial clearance, while acknowledging that unnecessarily prolonged or intense responses may contribute to pathology arising from infection. Indeed, predisposition to infection could be the result of one or more monogenic traits that confer primary immunodeficiencies; whether or not this is the case remains to be determined, but human studies have shown that responses to *B. burgdorferi* are diminished in individuals with specific mutations in or diminished expression of innate immune cell receptors (nucleotide-binding oligomerization domain 2 [NOD2] and Toll-like receptor 1 [TLR1]). It has also been observed that individuals on anti-IL-1β immunotherapy experience an increased frequency of invasive infections caused by *S. pyogenes*.[13]

Similarly, *T. pallidum* is known colloquially as the "stealth pathogen" because of its denuded outer membrane, which comprises mostly nonimmunogenic transmembrane proteins, whereas the highly immunogenic lipoproteins are contained within the periplasmic space.[14] This molecular architecture, coupled with the ability to generate antigenic variants, is responsible for the treponeme's remarkable ability to cause persistent infection with relatively few organisms.[15] Consequently, our understanding of the immune responses to this pathogen is not nearly as detailed as our knowledge of those elicited in response to infection with *B. burgdorferi* and other bacterial pathogens. Here, we will discuss both the innate and adaptive immune responses to bacterial pathogens, as well as physical and genetic barriers extracellular bacteria and spirochetes use to evade host killing.

Innate Immune Responses

Early Pathogen Recognition and Antimicrobial Peptides

The initial recognition of pathogens by host cells relies on a complex interplay between PRRs and bacterial constituents, initiating a cascade of responses leading to the upregulation of chemokines and cytokines, adhesion molecules, and other effector

molecules. This response is often initiated by endothelial and/or epithelial cells, resulting in the recruitment and activation of innate immune cells. Each PRR recognizes a specific structure that is present in a group or groups of microorganisms. The recognition of patterns instead of specific antigens provides the innate immune system with a rapid way to respond to infecting organisms until the more specific response mediated by T and B cells develops. Immune pattern recognition molecules are a major arm of the innate immune system and are released or expressed by a range of host cells, including lymphocytes, macrophages, dendritic cells (DCs), polymorphonuclear leukocytes (PMNs), and epithelial cells. The discovery and characterization of specific pattern recognition molecules (see below) has revolutionized our understanding of the initial specific events occurring between microbes and human cells.

Innate immune recognition relies on the detection of unique molecular structures found on microorganisms by host PRRs.[16,17] Toll-like receptors (TLRs) and NOD-like receptors (NLRs) are the best studied PRRs (Table 27.3). TLRs (TLR1–11) are found on macrophages, PMNs, and other host cells. These receptors recognize a variety of microbial ligands or pathogen-associated molecular patterns (PAMPs), including lipoproteins, lipopolysaccharide (LPS), flagellin, and nucleic acids produced by gram-negative and/or gram-positive bacteria. For instance, the interaction of *B. burgdorferi* lipoproteins with complexes formed by TLRs 1 and 2 initiates a series of signaling cascades that results in the production of proinflammatory cytokines (IL-1β, tumor necrosis factor [TNF], IL-12, and IL-18, among others), chemokines (IL-8, monocyte chemoattractant protein [MCP]-1, keratinocyte chemoattractant [KC]), metalloproteinases, adhesion molecules (E-selectin, vascular cell adhesion molecule-1 [VCAM-1] and intercellular adhesion molecule-1 [ICAM-1]),[17] and type I interferons (IFNs). Extracellular bacterial pathogens such as *S. pyogenes* signal through an unknown TLR but—like spirochetes—elicit the production of proinflammatory cytokines and chemokines.

TABLE 27.3 Pattern Recognition Receptors Recognize Pathogen-Associated Molecular Patterns From Various Bacteria

PRRs	PAMPs	Microbes
Toll-like receptor (TLR)2/1	Triacyl lipoproteins lipoprotein	Bacteria *B. burgdorferi, T. pallidum*
TLR2/6	Diacyl lipoproteins	Mycoplasma
	Lipoteichoic acid	Gram-positive bacteria
TLR2	Peptidoglycan	Gram-positive bacteria
	Porins lipoprotein	Bacteria (*Neisseria, Treponema denticola*)
TLR4	Lipopolysaccharide (LPS)	Gram-negative bacteria
TLR5	Flagellin	Flagellated bacteria (*Helicobacter pylori, Salmonella*)
TLR7/8	RNA	Group B *streptococcus*
TLR9	CpG-DNA	Bacteria (*Salmonella*)
	DNA	Bacteria (*Staphylococcus* at low MOI)
TLR11	Not determined	Uropathogenic bacteria
Nucleotide-binding oligomerization domain (NOD)1	Meso-diaminopimelic acid	*H. pylori, Bacillus* spp., *Campylobacter jejuni, Pseudomonas aeruginosa*
NOD2	Muramyl dipeptide (MDP)	*Streptococcus pneumoniae, Staphylococcus aureus, Salmonella typhimurium*
NOD-like receptor (NLR)P3	Whole pathogens	*Staphylococcus aureus*
	Toxins, LPS, MDP, and RNA	Bacteria
NLRP1	MDP	Bacteria
NLRP1b	Microbial toxin	*Bacillus anthracis*
NLRC4	Flagellin	*P. aeruginosa*

MOI, Multiplicity of infection; *PAMP*, pathogen-associated molecular pattern; *PRR*, pattern recognition receptor.

Both spirochetes and *S. pyogenes* use MyD88-dependent mechanisms to stimulate cytokine induction. Mice deficient in TLR1, TLR2, or MyD88—an adaptor molecule—have significant increases in *B. burgdorferi* burdens after infection, underlining the importance of TLRs in early pathogen detection; similar to *B. burgdorferi* lipoproteins, treponemal lipoproteins appear to be the major proinflammatory agonists during treponemal infection through engagement with their cognate receptors, TLR1/2, and CD14. The inflammatory milieu established by treponemal lipoproteins is a principal driving force for immune cell recruitment to *T. pallidum*-infected tissues. The importance of this immune system compartment during the response to *T. pallidum* is further supported by the systemic upregulation of innate immune cells during treponemal dissemination and by demonstration that macrophages are principal effectors of treponemal clearance during infection.[18] Other extracellular pathogens such as *S. pyogenes* lack LPS but do contain surface proteins such as M protein that elicit a highly inflammatory immune response largely mediated by macrophages, dendritic cells, and PMNs.

The NLRs are a family of intracellular receptors, some of which also function as PRRs.[16] NOD1 and NOD2 are well characterized as PRRs to extracellular pathogens, such as *S. pyogenes*. Importantly, in concert with TLR signaling, NLR can respond to a variety of PAMPs by forming the inflammasome complex. Inflammasome activation generates interleukin (IL)-18 and activates IL-1 through caspase-1, an important step in the immune response to many bacteria. Inflammasome activation is critical to the restriction of many pathogens. Interestingly, *S. pyogenes* can directly activate IL-1β independently of the NLRP3 inflammasome, leading to a highly inflammatory and often detrimental host immune response.[13] Similarly, *B. burgdorferi* stimulates the production of IL-17 and IFN-γ by a caspase-1-dependent mechanism leading to a highly inflammatory immune response critical to Lyme disease progression and pathology, but unlike *S. pyogenes*, does not rely on the NLRP3 inflammasome.

Upon recognition of bacterial pathogens by PRRs, immune cells produce antimicrobial peptides (AMPs) in addition to cytokines and chemokines that either kill invading microbes or modulate and dampen host inflammatory responses (Table 27.4).[12] Cells that are present on skin and mucosal surfaces constitutively express AMPs as a first line defense against invading pathogens, but AMPs can also be induced by infection or injury.[12] Electrostatic interactions between AMPs and negatively charged cell membranes result in either membrane disruption or entry into the cell, consequently inhibiting critical cell functions. Humans produce two main classes of AMPs: defensins and cathelicidins. Defensins, for example, are expressed in skin, intestines, and the respiratory tract and have activity against gram-positive and gram-negative bacteria. Cathelicidins such as LL37 are produced by macrophages and PMNs and are important for rapid bactericidal activity against both spirochetes and extracellular bacteria. Keratinocytes of inflamed psoriatic lesions produce increased levels of certain AMPs, and

TABLE 27.4 Antimicrobial Proteins Against Extracellular Bacteria and Spirochetes

AMP	Tissue/Cell Sources	Mechanism of Action	Target Organisms
α-Defensins	Small intestines, Paneth cells	Membrane disruption; inhibits complement activation; chemoattracts dendritic cells	Gram-positive bacteria Gram-negative bacteria
β-Defensins	Large intestines, skin, respiratory tract epithelial cells	Membrane disruption; lipid II binding	Gram-positive bacteria Gram-negative bacteria
Cathelicidin (LL37)	Large intestine, skin, lung, urogenital tract	Membrane disruption	Gram-positive bacteria Gram-negative bacteria
RNases	Skin, intestine, respiratory epithelia, placenta	Unknown	Gram-positive bacteria Gram-negative bacteria
Psoriasin (S100A7)	Skin, urogenital tract	Unknown	*Escherichia coli*
Calprotectin (S100A8-A9)	Abscesses/neutrophils	Metal chelation	*Staphylococcus aureus*
C-type lectins	Small intestines	Peptidoglycan recognition	Gram-positive bacteria
Bactericidal/permeability-increasing protein (BPI)	Neutrophils, epithelial cells	Neutralizes lipopolysaccharide	Gram-negative bacteria
Lysozyme	Skin, body fluids, tears, intestinal Paneth cells	Degrades peptidoglycan	Gram-positive bacteria Some gram-negative bacteria activity
Dermcidin	Sweat glands	Membrane disruption	Gram-positive bacteria Gram-negative bacteria
Peptidoglycan recognition proteins	Liver, intestines, skin, neutrophils	Activates bacterial two-component systems; targets peptidoglycan	Especially active against gram-positive bacteria Gram-negative bacteria
Phospholipase A_2	Tears, intestines	Hydrolysis of bacterial phospholipids	Gram-positive bacteria

AMP, Antimicrobial proteins.

patients with such lesions rarely have secondary bacterial infections. Not surprisingly, successful pathogens have developed several mechanisms to counteract AMPs; invasive clones of *S. pyogenes*, for example, uses the M surface protein to sequester the human cathelicidin LL37 to evade killing by AMPs and cause diseases such as necrotizing fasciitis.[19]

The interaction of several cell types with spirochetes and other extracellular bacteria also involves several integrins. Integrins are involved in the adhesion of cells to a variety of ligands and mediate essential cellular processes, including attachment and cell migration. Some integrins have also been associated with the phagocytosis of microorganisms. For the most part, studies of the interaction between the spirochete and integrins have focused on their role in aiding the adhesion of *B. burgdorferi* to host cells and the colonization of tissues and as receptors that contribute to signals that induce the production of proinflammatory factors (Fig. 27.3). Integrins are also important for restricting extracellular bacteria. Some serotypes of *S. pyogenes* encode a CD11b homologue that is able to modulate host immune responses and prevent phagocytosis; conversely, *S. pyogenes* can also encode collagen-like surface proteins that bind to $\alpha_2\beta_1$ integrins to promote phagocytosis, which allows for the bacterium to later re-emerge as an extracellular pathogen.

Complement

Destruction of microorganisms via complement involves the formation of a pore in the microbial cell membrane by the membrane attack complex (MAC), which results in the lysis of the organism and leads to the release of opsonins and chemoattractant molecules for phagocytic cells.[20] There are three different pathways that elicit complement activation: the classical (antigen/antibody–mediated) pathway (CP), the lectin pathway, and the alternative (pathogen surface) pathway (AP). These pathways converge at the level of C3 convertase, a protease that cleaves complement component C3 into C3a and C3b. As a

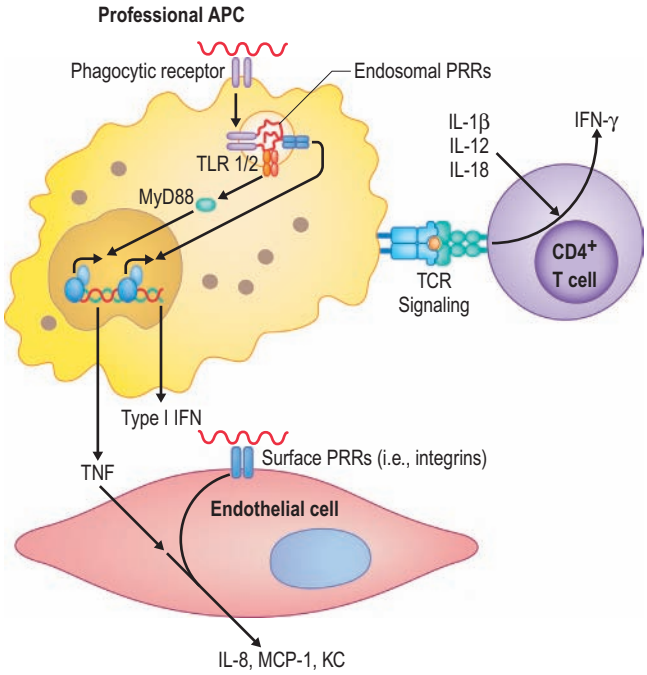


FIG. 27.3 The Interaction of Borrelia burgdorferi With Pattern Recognition Receptors (PRRs) in Innate Immune Cells and endothelial Cells Mediates the Inflammatory Response to the Spirochete. Phagocytosis induces Toll-like receptor (*TLR*)–driven proinflammatory cytokine production as well as antigen presentation by professional antigen-presenting cells (*APCs*), which leads to the activation of CD4 effector T cells, marked by the production of interferon-gamma (*IFN-γ*). Likewise, PRR- and tumor necrosis factor (*TNF*)–receptor signaling lead to the upregulation of chemokines by endothelial cells. Overall, these responses lead to increased activation and recruitment of innate immune cells in sites of infection.

result, C3b can (1) bind to the surface of the bacteria and facilitate internalization of the bacterium via opsonization; or (2) it can bind C3 convertase and facilitate the deposition of downstream components onto the surface of the bacterium resulting in the formation of MAC and lysis of the cell. Spirochetes such as *B. burgdorferi* and *T. pallidum* activate the CP and AP of the complement cascade. Moreover, the activation of complement has been associated with dramatic decreases in spirochetal numbers in different tissues of infected mice, indicating the importance of the complement system in early infection, particularly for *B. burgdorferi*. The immune protection afforded by human syphilitic serum is, in large part, a result of the activation of the complement cascade by bactericidal antibodies and spontaneous hydrolysis of C3, and there is a considerable amount of evidence for an important role of these pathways in syphilitic lesion resolution during human infection with treponemes.

Bacteria have evolved several ways to evade complement-mediated killing. Gram-positive extracellular pathogens resist the bacteriolytic action of the MAC as a result of a thick peptidoglycan layer, which impedes the insertion of the MAC C5b-9 complex; similarly, bacteria such as *Streptococcus pyogenes* and *Staphylococcus aureus* can evade host immune detection by producing a thick capsule that precludes opsonization. *S. pyogenes* M protein can also directly bind C4b and fibrinogen, thus inhibiting complement-mediated opsonization. Gram-negative bacteria can resist the MAC through structural alterations in their LPS (the possession of O antigen keeps the MAC at a distance from the bacterial surface) or by masking or deleting the epitope(s) responsible for binding bactericidal antibodies. Binding of human factor H (hfH) by meningococci factor H-binding protein (fHbp) downregulates the host AP and helps the organism to evade host innate immunity and is now included in the new serogroup B vaccines.[21] Spirochetes like *B. burgdorferi* have evolved a variety of mechanisms enabling them to escape complement-mediated lysis, including the expression of complement regulator-acquiring surface proteins (CRASPs).[22] Of these CRASPs, the Erp (OspEF-related protein) family of outer membrane proteins serve as binding sites for the complement inhibitor factor H and factor H–like protein 1 (FHL-1).[23] The interaction of factor H with these proteins recruits a protease (factor I) that cleaves and inactivates the complement serum proteins C3b and C4b. Cleavage of these two complement proteins prevents the deposition of downstream components onto the surface of the spirochete, thereby halting the formation of MAC. *B. burgdorferi* also express a CD59-like molecule on the outer membrane that can inactivate MAC and prevent complement-mediated lysis.[24] In contrast to *B. burgdorferi*, there is no evidence indicating that *T. pallidum* has evolved mechanisms to evade complement-dependent killing, suggesting that these complement pathways have a larger role in controlling treponemal infection than in the case of *B. burgdorferi*. During experimental syphilis, immunization with purified outer membrane vesicles (OMVs) isolated from *T. pallidum* results in complement-dependent bactericidal activity.[25] More recently, immunization with OMVs led to the isolation of a bactericidal monoclonal antibody (mAb) M131 that provides partial protection to experimental syphilis.

Phagocytic Cell Recruitment and Pathogen Clearance

The recruitment of phagocytic cells and other cell types into sites of infection is mediated by the production of chemokines, increased vascular permeability, and upregulated expression of cell adhesion molecules in endothelial cells. Macrophages, PMNs, and other phagocytic cells are also present at mucosal surfaces. These cells express PRRs and migrate to mucosal surfaces by chemotaxis and diapedesis between epithelial cells. Macrophages are also encountered after crossing the epithelial barrier. Specialized epithelial M cells of mucosal surfaces are key sites for antigen sampling, including viruses, and bacteria and macrophages surround these sites.[26] Dendritic cells (DCs; Langerhans cells) sample live bacteria at the mucosal surface, traffic to mucosal lymphoid tissue, and induce B cells to produce bacteria-specific immunoglobulins.[4] In most tissues, DCs are at a low level of activation and are immature, but upon activation, they take up and process antigens. DCs are rich in PRRs (e.g., TLRs), and microbe–PRR interactions have a key role in shaping the T-cell response.[27] DCs play a central role in controlling *S. pyogenes* infection by the production of TNF, and mice deficient in CD11c+ DCs fail to control infections caused by *S. pyogenes*. Skin contains a major supply of tissue DCs, and their involvement in combating skin and soft tissue infections must be considered along with their function and contribution to stimulating immunity during vaccination.

Chemokine production at sites of pathology in disease-susceptible C3H/HeJ mice and disease-resistant C57BL/6J mice shows that inflammation is related to increased production of neutrophil and monocyte–macrophage chemokines, KC and MCP-1, respectively. In humans, the production of chemokines, especially IL-8, during the initial response to *B. burgdorferi* correlates well with the onset of symptoms associated with the early stages of infection. This suggests cytokine production is increased during the early stages of infection to recruit phagocytic cells, which are involved in the initial clearance of the spirochete.

In areas of epithelial inflammation, PMNs can be recruited to mucosal and skin surfaces. PMNs are more effective in the presence of other immune defenses, such as antibody and complement components. PMNs express PRRs and have both oxygen-dependent and oxygen-independent mechanisms of killing (Fig. 27.4). Activated neutrophils can release granule proteins with direct antibacterial action (e.g., bactericidal/

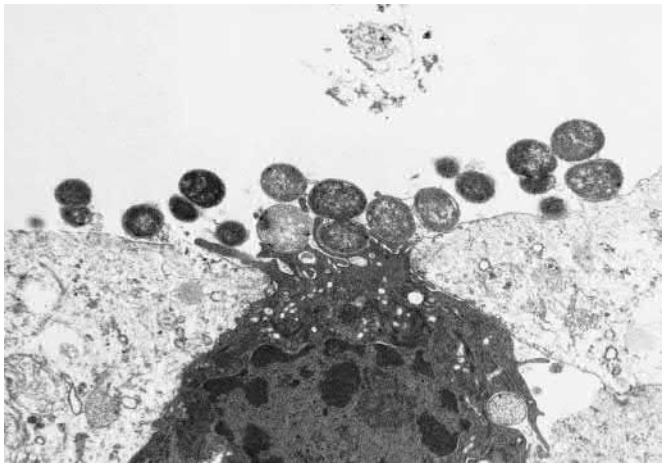


FIG. 27.4 Bacterial Phagocytosis at Mucosal Surfaces. Transmission electron micrograph of phagocyte engulfing *Neisseria meningitidis* at a human respiratory epithelial mucosal surface (×19,000).

permeability-increasing [BPI]) or degradative activity (e.g., elastase) and chromatin containing the antibacterial histone H2A.[28] These released compounds work together to form extracellular fibers, termed *neutrophil extracellular traps* (NETs), which can trap and kill gram-positive and gram-negative bacteria and degrade their virulence factors as well. NETs have been observed in instances of acute inflammation (experimental dysentery and spontaneous appendicitis) and provide a mechanism for reducing bacterial spread at sites of acute infection. Strains of *S. pyogenes* that encode the DNase Sda1 that degrades NETs are significantly more virulent, highlighting the importance of NETs in restricting extracellular bacterial pathogens. PMNs are important to restricting infection by both spirochetes and other extracellular bacteria due to their production of reactive oxygen species (ROS) including superoxides; consequently, neutrophil influx may also contribute to the onset of arthritis observed during Lyme disease. The importance of PMNs in host defense against extracellular pathogens can best be highlighted by the increased frequency of bacteremia and other life-threatening infections in patients with neutropenia or those individuals with neutrophil deficits (e.g., chronic granulomatous disease, Chediak-Higashi syndrome, or specific granule deficiency, to name a few).

Phagocytosis plays a major role in the pathogenesis of spirochetes and other bacteria not only through the control of bacterial numbers but also through modulation of the potency and quality of proinflammatory cytokine induction. However, little is known about the molecular events or receptors that mediate *B. burgdorferi* phagocytosis despite the importance of this cardinal mechanism of pathogen clearance. MyD88-mediated signals substantially mediate the phagocytosis of *B. burgdorferi* and *T. pallidum;* however, MyD88-mediated phagocytosis occurs independently of any known *B. burgdorferi*-recognizing TLRs. Similarly, it is unclear which TLRs are required for phagocytosis of *S. pyogenes*, but MyD88 deficiency leads to impairment of phagocyte recruitment and significantly decreased production of cytokines and chemokines. The analysis of MyD88-deficient macrophages shows that although reduced, phagocytosis of *B. burgdorferi* is not absent, indicating that uptake seems to be mediated by more than one receptor. Indeed, as opposed to MyD88-mediated phagocytosis—which is proinflammatory—the internalization of *B. burgdorferi* by CR3 tempers the inflammatory response of macrophages; therefore the presence of alternative phagocytic mechanisms has nonredundant physiological consequences during infection with the spirochete. Similarly, activated macrophages readily phagocytose antibody-opsonized *T. pallidum* and other treponemes through Fc receptor–mediated uptake. These results revealed a previously unsuspected link between MyD88 signaling and FcR-mediated phagocytosis.

Once bacteria are taken up by phagocytic cells, they must endure the intracellular phagolysosome. Phagolysosomes form after bacteria are taken up by professional phagocytes, forming small vacuoles that fuse with lysosomes containing ROS and other hydrolases used to kill bacteria. The acidification of the phagolysosome is a critical step in the clearance of bacterial pathogens. The presence of the phagolysosome initiates a cascade of intracellular signaling, leading to the increased production of proinflammatory cytokines; for example, spirochetes like *B. burgdorferi* induce interferon production from within the phagolysosome through a MyD88-TLR8-dependent pathway, resulting in killing of the spirochete. Very little is known about killing of *T. pallidum* through phagolysosomes, but it has been shown that both spirochetes are able to induce production of IFN-γ.[29,30] Bacteria have developed a variety of ways to evade both phagocytosis and phagolysosome-mediated killing. These can involve physical barriers such as the hyaluronic acid capsule and M protein of *S. pyogenes* that bind the Fc region of antibodies, or the denuded outer membrane of *T. pallidum* that results in diminished phagocytosis; other mechanisms include the degradation of phagolysosomes through pore-forming toxins such as streptolysin O produced by *S. pyogenes*.

Several extracellular bacteria possess polysaccharide-rich capsules that resist phagocytosis. A number of pyogenic bacteria (e.g., *S. aureus*) secrete leukocidins, which lyse phagocytes. Other pathogens (e.g., group A streptococci) inhibit chemotaxis of neutrophils through the elaboration of enzymes (e.g., C5a peptidase) that proteolytically cleave chemotactic signals. Some bacteria possess mechanisms to prevent opsonization by changing surface antigens.[7] Many bacteria form biofilms, which shield these microorganisms from host defense molecules and antibiotics.[31] Leukocytes that invade *S. aureus* biofilms exhibit impaired phagocytosis and decreased ability to kill bacteria. In addition, biofilm matrices can protect bacteria from antibody-mediated phagocytosis.

Adaptive Immune Responses to Extracellular Bacteria and Spirochetes

T Cell–Mediated Responses

Upon antigen presentation by macrophages, dendritic cells (DCs), or B cells, naïve CD4 T cells are activated and differentiate into effector T cells. Effector CD4 T cells are classified on the basis of their cytokine production profile, which determines their mode of action and downstream effects, and include T-helper [Th]1 (IFN-γ-producing), Th2 (IL-4, IL-5, IL-13), Th17 (IL-6, IL-17), or regulatory T cells (Tregs; IL-10). Interaction of *B. burgdorferi* antigen with TLRs induces the production of IL-12, which drives the differentiation of CD4 T cells into Th1 effector cells. Th1 cells are regulators of the cell-mediated inflammatory reactions, which are characterized by macrophage activation, including phagocytosis, or the induction of opsonizing IgG antibodies. In the case of *T. pallidum*, the infiltration of T cells and macrophages into the primary and secondary syphilitic lesion facilitates local clearance of the majority of treponemes via a vigorous cell-mediated immune response characteristic of a delayed-type hypersensitivity or Th1 response.[32,33] CD4 T cells are the principal T-cell subset found in the lesions and are believed to promote macrophage activation and subsequent treponeme clearance through IFN-γ secretion.

Although IFN-γ and Th1 CD4 T cells have been shown to be protective during cardiac inflammation with *B. burgdorferi*, joint inflammation is independent of this effector cell type in mice. However, in patients with Lyme disease, Th1 cells dominate in the synovial fluid, and the severity of arthritis directly correlates with increased levels of Th1 cells in the synovium.[34] Whether neutrophilic infiltration during joint infection with the spirochete is influenced by Th effector cells that more directly affect this cell type, such as Th17 cells (through the production of IL-17), remains to be elucidated.

Th1 cells are characterized by IFN-γ production and function to activate macrophages to phagocytize and kill pathogens. While this mechanism of pathogen elimination is primarily directed against pathogens with a predominant intracellular life

cycle, Th1 cells are relevant for typical extracellular bacteria such as *Streptococcus pyogenes* and *pneumoniae*.[35] The neutrophilic response to extracellular bacteria is primarily coordinated by Th17 cells.[36] Animal models have suggested that the Th17 response is central for protection against a wide variety of gram-positive and gram-negative bacteria. For example, Th17 response has been shown to induce nasopharyngeal clearance of Pneumococcus in both animal models and in children. Differentiation toward the Th17 subtype appears to be favored by strong antigenic signals and broad activation of PRRs. IL-17 and IL-22, the signature ILs of the Th17 response, promote AMP secretion by epithelial cells, neutrophil migration, and epithelial integrity. *S. pyogenes* induces a strong inflammatory response, leading to the activation of Th1 and Th17 cells, which is largely mediated by IFN-γ and IL-6 production. Invasive infections by *S. pyogenes*, particularly toxic shock syndrome, are a result of aberrant activation of large populations of T cells that produce fatal levels of inflammatory cytokines.

B Cell–Mediated Responses

Antibodies are specific and powerful effector molecules of the adaptive immune response. Once antibodies bind to their specific foreign antigen, they confer protection to the host by using a variety of effector mechanisms. However, in response to *B. burgdorferi*, T cell–independent humoral responses also confer protection to the host. This is supported by studies that indicate unusually large numbers of B cells residing in lymph nodes with minimal activation of CD4 T cells; follicular helper T cells that are commonly associated with germinal center B-cell responses, were also present.[37] Mice that are deficient in CD4 T cells were still able to resolve infection, but mice that lack both B and T cells developed severe arthritis and carditis in response to infection with *B. burgdorferi*, indicating that B cells are also important for clearance of spirochetes. In addition, mice deficient in CD40L and MHC class II infected with *B. burgdorferi* mount a protective antibody response, which, upon passive serum transfer, affords protection to severe combined immunodeficient (SCID) mice from homologous challenge.

The role of antibodies in controlling *B. burgdorferi* infection may be more important during the hematogenous dissemination phase, when they are easily accessible, than after the spirochetes have colonized tissues. Once the spirochete is in the joints and the heart, its clearance may be more dependent on cellular responses (e.g., macrophages in the heart) than on antibodies. In fact, the lack of IFN-γ–mediated activation of macrophages has profound consequences on murine cardiac inflammation, even in the presence of strong antibody responses. Furthermore, the bacterial clearance potential of infected mouse sera administered in newly infected mice is lost when administered 4 to 8 days after infection, potentially the result of the colonization of tissues into which antibodies are less able to penetrate. Like *B. burgdorferi*, many functional activities of human syphilitic serum originate from B-cell responses to *T. pallidum*. Infection with *T. pallidum* invokes a humoral immune response early in the course of infection, which strengthens as the number of recognizable antigens increases during the progression of infection.

Both *B. burgdorferi* and *T. pallidum* can avoid clearance by antibodies through antigenic diversity. *B. burgdorferi* differentially expresses outer membrane antigens under pressure from the immune response, which might contribute to the ability of the spirochetes to persist in the host. A mechanism that is potentially essential for spirochetal immune escape is the recombination that takes place at the *vls* locus,[38] located near the right telomere of the linear plasmid lp28-1. TprK, an immunogen found in *T. pallidum* has seven discrete variable regions differing among isolates of the spirochete.[39] In fact, the antibody response to *T. pallidum* is directed against these variable regions, which leads to immune selection of new TprK variants; thus, antigenic variation is involved in the reinfection of hosts by spirochetes despite robust immune responses.[39]

Producing an effective adaptive immune response to *S. pyogenes* has been challenging due to the development of rheumatic heart disease, an autoimmune disorder in which cross-reactive antibodies against *S. pyogenes* bind to epitopes on heart tissue and antagonize heart valves. The mechanism by which this occurs is still unclear. Other studies have indicated that *S. pyogenes* is capable of evading B cell–mediated responses via superantigen SpeA, resulting in the production of abnormal follicular helper T cells, leading to B-cell death.[40] In contrast, B cells can produce antigen-specific IgG or IgM toward other extracellular bacterial pathogens such as *Staphylococcus aureus*. Neonates often receive IgG antibodies against *S. aureus* from their mother during pregnancy. The importance of B cells in resolving *S. aureus* infection can be seen in HIV-infected patients where the decline of CD4 T cells, responsible for assisting B-cell germinal centers, results in increased incidence of *S. aureus* infection.

Immunoglobulins

Immunoglobulins (Igs), principally secretory IgA and IgG, are present at mucosal surfaces and in mucosal secretions. Dissemination of IgA and IgG class–committed B- and T-helper (Th) cells with specificity to an antigen encountered and processed at a mucosal site to the mucosal site where the antigen was encountered and to distant mucosal sites is important in the generation of mucosal immunoglobulins. Protective mucosal antibodies against bacteria may be derived from prior colonization, vaccines, or shared cross-reactive antigens on normal flora. Mucosal immunoglobulins may neutralize bacterial toxins, facilitate phagocytosis or bactericidal activity, inhibit bacterial adherence ligands, or sterically hinder other events necessary for bacterial colonization and invasion. Many extracellular bacterial pathogens (*N. meningitidis*, *N. gonorrhoeae*, *H. influenzae*, certain streptococci) colonize and/or infect mucosal surfaces where protective IgA_1 antibodies could become available. These pathogens secrete an IgA_1 protease that cleaves IgA_1, thereby inactivating the molecule. IgA_1 protease can also recognize other substrates, notably lysosomal-associated membrane protein 1 (LAMP-1), which are important in host defense.

HOST RISK FACTORS FOR LOCAL AND SYSTEMIC INVASION BY EXTRACELLULAR PATHOGENS AND SPIROCHETES

Bacteria that breach mucosal and skin barriers and reach submucosal tissues of sites, such as pulmonary alveoli, the middle ear, and the bloodstream, induce immune responses, including cytokine release, phagocytosis, complement activation, antibody release or production, and other local or systemic induction of the inflammatory cascade (Fig. 27.5). The survival of bacteria following colonization of the epithelium and access to the bloodstream depends on the integrity of the host immune

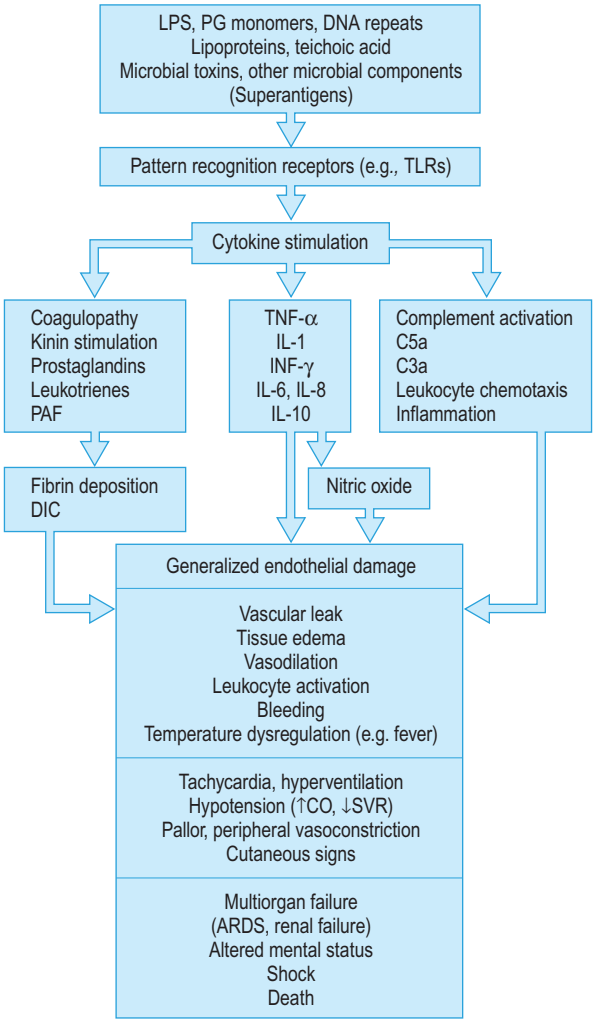


FIG. 27.5 Inflammatory Cascade Initiated During Sepsis.

response (including variability caused by genetic polymorphisms) and on the ability of the bacteria to resist this host immune response. Host factors that increase the risk for the development of systemic disease as a result of extracellular bacteria include polymorphisms in innate immune mechanisms, the absence of bactericidal or opsonizing antibodies, deficiencies in the complement pathways, and an absence of or reduction in neutrophil function or levels (see Table 27.1).

Complement deficiencies, either congenital or acquired, increase the risk for invasive bacterial diseases. Because C3 plays a critical role in the complement cascade, congenital C3 deficiency or conditions that reduce C3 (e.g., systemic lupus erythematosus, cirrhosis, nephritis, C3 nephritic factor) increase the risk for invasive disease due to pyogenic bacteria, such as *S. pneumoniae* and *N. meningitidis*. Mannose-binding lectin (MBL) is a plasma opsonin that initiates complement activation. *MBL* gene polymorphisms are associated with meningococcal and pneumococcal sepsis during childhood. Properdin deficiency, leading to defective AP killing, is also associated with severe and recurrent meningococcal infections. Terminal complement deficiencies (C5–C8) are also associated with recurrent invasive bloodstream meningococcal and gonococcal infections, indicating an important role for insertion of the complement MAC in the bactericidal activity of human serum against pathogenic *Neisseria*.

In addition to defects in innate immunity, immunoglobulins, and complement deficiencies, human genetic polymorphisms are associated with an increased risk or severity of bacterial diseases. For example, FcγIIa (CD32) receptor polymorphisms, Fcγ-receptor III (CD16), MBL, TLR4, TNF promoter region polymorphisms, plasminogen activator and inhibitor expression, and hereditary differences in cytokine induction influence susceptibility to meningococcemia. Each of these polymorphisms can influence the course of invasive bacterial infection by influencing the response of the inflammatory cascade.

DELETERIOUS HOST RESPONSES

Inflammation and Autoimmunity

The host immune response can be the leading cause of tissue injury in the acute phase of an infection. Brain edema and infarcts, which are devastating consequences of pyogenic meningitis, occur as a result of the host inflammatory response. Corticosteroids are currently recommended as an adjunctive therapy to pneumococcal meningitis, an acute pyogenic infection. The use of small molecules targeting specific immunological pathways is an area of ongoing research. As acute infections are initially characterized by an inflammatory response followed by an anti-inflammatory response, the timing of use of these compounds is of the utmost importance.[41] Molecular mimicry between a bacterial antigen and a host protein can lead to autoimmunity. Examples include rheumatic fever and glomerulonephritis after *S. pyogenes* infections, reactive arthritis following *Chlamydia trachomatis* urethritis, and the Guillain-Barré syndrome following *Campylobacter jejuni* enteritis. Molecular mimicry can also limit selection of epitopes for vaccine development.

Sepsis

Septicemia remains a leading cause of death in the United States and accounts for several billion dollars in healthcare expenditure.[41] Both gram-negative and gram-positive bacteria can rapidly multiply in the bloodstream and trigger sepsis and septic shock. Septic shock is a result of an initial and widespread systemic proinflammatory response, resulting in hypotension, organ failure, and death. The later phase of sepsis is also characterized by an antiinflammatory response. Although survival of patients with the acute phase of sepsis has improved, advances in treatment and prevention of death, which can be secondary to nosocomial infections, have been slower. These secondary infections are often caused by less virulent organisms, likely as a result of "immunoparalysis" (as a result of an exaggerated anti-inflammatory reaction) and also breaching of the physical barriers to infection as a result of invasive medical procedures (e.g., intravenous lines, intubation, and bladder catheterization). The systemic inflammatory cascade of sepsis is initiated by recognition of PAMPs by PRR, both in extracellular (e.g., LPS) and intracellular environments (e.g., DNA fragments). The severity of sepsis is also influenced by polymorphic alleles of genes involved in the inflammatory cascade.[41]

The morbidity and mortality of bacteremia and sepsis have been directly correlated with the initial levels of proinflammatory cytokines and the amount of circulating bacterial components. Indeed, the severity of gram-negative sepsis has been equated with high levels of endotoxin, increased levels of cytokines, and excessive activation of the AP. Disseminated intravascular coagulation, which often accompanies gram-negative sepsis, is

caused by excessive activation of the coagulation cascade and downregulation of the fibrinolytic system associated with high levels of LPS. Levels of natural anticoagulants in the vasculature, such as antithrombin and protein C, are often low in gram-negative sepsis. The onset and severity of disseminated intravascular coagulation may be influenced by genetic polymorphisms in plasminogen activation or inhibition. The generalized, altered vascular endothelial lining facilitates thrombosis and thrombocytosis. Although much remains to be learned about the mechanisms by which gram-negative and gram-positive bacteria and microbial products trigger sepsis, significant advances have been made recently, particularly with endotoxin-mediated sepsis. Advances during the past decade include the identification of certain LPS–host protein interactions that result in delivery of LPS to host cell receptors and gene activation events that result in elevated expression of a diverse array of proinflammatory and antiinflammatory mediators (Fig. 27.6).

For example, TLR4 signaling requires an accessory protein, myeloid differentiation protein-2 (MD-2), which binds directly to endotoxin. The key point, however, is that the LPS engagement of MD-2/TLR4 on host cells, particularly macrophages, triggers intracellular signaling events that ultimately, through nuclear factor kappa B (NF-κB) and other pathways, result in cytokine gene activation and production of cytokines (TNF, IL-1, IL-6, IL-8, IFNs). Other TLRs (e.g., TLR2) play a critical role in the recognition of lipoproteins, and the recognition of these components is a likely key determinant in the development of septic shock seen with gram-positive infections. TLR5 recognizes bacterial flagella, and such recognition is of importance in the host response to motile bacteria. Some human pathogens (e.g., *H. pylori*) produce flagellin molecules that do not engage TLR5. TLR9 has been shown to recognize bacterial DNA CpG dinucleotides. Taken together, the clinical syndrome of septic shock represents a series of interactions of bacterial products in the vascular space with pattern recognition molecules on serum proteins (lipopolysaccharide-binding protein [LBP] and soluble CD14) and with host cell receptors (MD-2/TLR4 and TLR2, other TLRs), leading to signaling events and release of transcriptional factors that modulate cytokine gene expression. These events also trigger other events in the inflammatory cascade, leading to activation of the coagulation, complement, and kinin pathways. Superantigens can activate large pools of non–antigen-specific T cells by binding to MHC class II molecules and the TCR Vβ region outside the peptide-binding domain. The result is a cytokine storm with clinical manifestations of sepsis.

Early and effective antimicrobial therapy is the primary goal in the treatment of sepsis. In contrast to the initial phase of sepsis characterized by the release of TNF, IL-1, IL-6, and IFN-γ, an antiinflammatory response may predominate during the latter phase. The clinical failures of antiinflammatory therapeutic mediators (antiendotoxin antibodies, TNF, antagonists of IL-1, or platelet-activating factor) in sepsis suggests that this hypoinflammatory state encountered in many patients at presentation could be an additional target of immune modulation.

THERAPEUTIC PRINCIPLES

Sepsis

- Early effective antibiotic therapy associated with improved outcomes
- Lack of "source control" (e.g., removal of infected lines, drainage of abscesses) linked to poor outcomes even in the presence of effective antibiotics
- Intensive and supportive care, management of fluid, electrolytes, and respiratory function
- Insulin for glucose control—unknown mechanism of protection; neutrophils have impaired function in the presence of hyperglycemia, insulin can have antiapoptotic effects

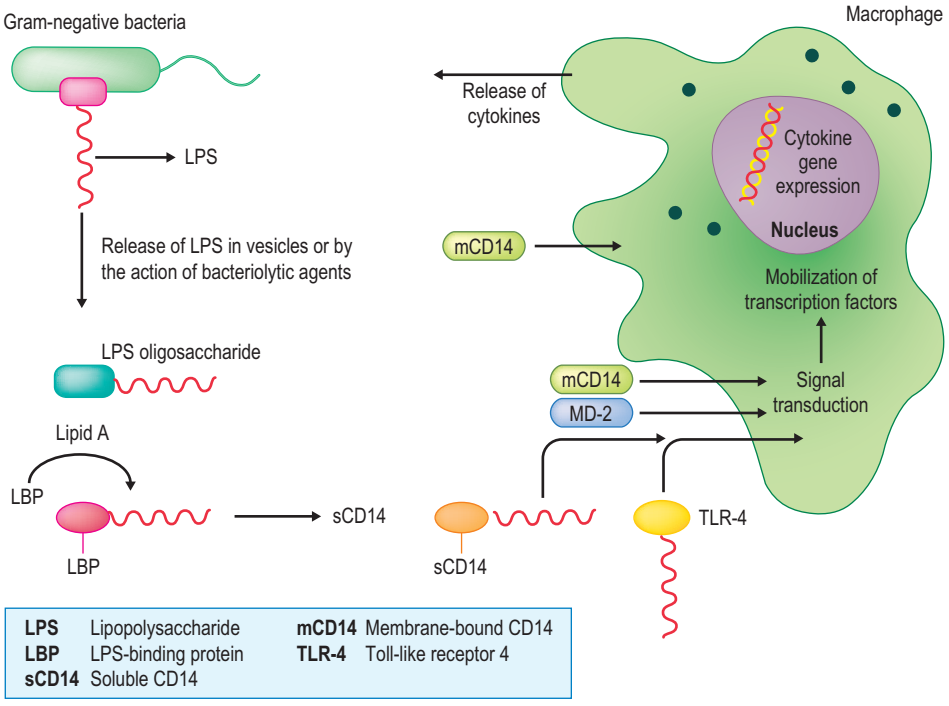


FIG. 27.6 Lipopolysaccharide *(LPS)* Triggering of Cytokine Production by Macrophages. Steps known or presumed to be necessary for LPS triggering of proinflammatory cytokines.

TRANSLATIONAL RESEARCH OPPORTUNITIES

An important challenge for the next decade will be to take the rapidly expanding basic discoveries in innate immunity, systems biology, and response to bacterial antigens into clinical applications. The design and use of bacterial vaccines through the assessment of innate immune molecular signatures after vaccination, both for general use and for subpopulations of nonresponders, is one example. A second is the continued development of small-molecule inhibitors or enhancers that target innate immune pathways to modulate bacterial immune responses. A third is the control of mucosal immune responses to prevent or eliminate colonization by bacterial pathogens. A fourth is the understanding of the role of the microbiome in shaping the immune response to pathogens and vaccines and its therapeutic potential for both infections and noninfectious diseases. Fecal microbiota transplantation for *C. difficile* colitis is an early example of therapeutic use of the microbiome. Finally, the development of new therapies for acute bacterial sepsis may be based on improved understanding and control of the immune responses in sepsis.[42]

The challenge in the next 5 to 10 years is to develop better diagnostic methods for both Lyme disease and syphilis and to devise new preventive measures. For Lyme disease, advances in genetic manipulation, as well as other means to study structure/function of key components of the spirochete, could lead to development of single or combination vaccines. Furthermore, the identification of host factors that mediate a protective response and those that contribute to inflammation may result in a better understanding of the disease. Overall, the identification of the elements that mediate specific tissue tropisms for the bacterium and their interaction with local/infiltrating cellular components can permit the design of targeted therapies in conjunction with antimicrobial treatments.

A significant effort is being made to find new antigenic determinants that can be the base for a vaccine to prevent infection with *B. burgdorferi*. While the search for spirochetal antigens continues, a new wave of research has focused on finding antigenic determinants that can prevent the efficient attachment of the tick vector to the mammalian host, which, in turn, could dampen the ability of the arthropod to transmit *B. burgdorferi* and perhaps other pathogens.

For syphilis, the study of the microorganism is hampered by the difficulties associated with its culture and manipulation. Although advances have been made in our understanding of the pathology associated with infection, a significant challenge in the near future is to fully understand this pathogen. Despite availability of a safe and effective therapy since 1943 (penicillin), syphilis control remains elusive and there has been a resurgence over the last two decades, including in high-income countries. Over the next decade, increasing uptake of existing preventive methods (e.g., male condoms), developing more effective patient-centered education methods, and new prevention methods (e.g., pre-exposure prophylaxis) could curb the current syphilis epidemic.

KEY CONCEPTS

Protective Versus Pathological Responses

Borrelia burgdorferi

- An early immune response to *B. burgdorferi* is necessary to control spirochetal burden; however, by itself, it is not sufficient to resolve infection.
- Phagocytosis is a key element of the innate immune response, which is involved in the elimination of the bacteria while also contributing to the proinflammatory output of macrophages.
- A T cell–mediated response appears to be involved in pathology arising from infection.
- A T cell–independent B-cell response is sufficient to resolve infection with *B. burgdorferi*.

Treponema pallidum

- The role of the early innate response to *T. pallidum* is poorly understood because:
- A cell-mediated immune response to *T. pallidum* is likely involved in the development of pathology following infection with the spirochete and resolution of infection.
- The humoral response is unclear; no definite antigens have been isolated due to inability to cultivate organism in vitro.

ON THE HORIZON

- Tailoring vaccine design based on assessment of innate immune molecular signatures
- New-generation vaccines for Lyme disease that target tick vectors
- Small-molecule inhibitors or enhancers specifically targeting innate immune pathways
- Identification of immune responses that prevent or eliminate mucosal bacterial pathogen colonization
- Development of new therapies modulating immune response in sepsis
- Defining microbial community and metagenome changes after antibiotic treatment
- Managing disease based on the human microbiome
- Efficient diagnostics for spirochete infection
- Investments in prophylaxis and treatments for communities at risk for syphilis

REFERENCES

1. Silva MT. Classical labeling of bacterial pathogens according to their lifestyle in the host: inconsistencies and alternatives. *Front Microbiol.* 2012;3:71.
2. Center for Disease Control and Prevention. 2018. Available at: https://www.cdc.gov/lyme/stats/index.html.
3. Moens E, Veldhoen M. Epithelial barrier biology: good fences make good neighbours. *Immunology.* 2012;135:1–8.
4. Nestle FO, Di Meglio P, Qin JZ, et al. Skin immune sentinels in health and disease. *Nat Rev Immunol.* 2009;9:679–691.
5. Ribet D, Cossart P. How bacterial pathogens colonize their hosts and invade deeper tissues. *Microbes Infect.* 2015;17:173–183.
6. Pan X, Yang Y, Zhang JR. Molecular basis of host specificity in human pathogenic bacteria. *Emerg Microbes Infect.* 2014;3:e23.
7. Criss AK, Seifert HS. A bacterial siren song: intimate interactions between *Neisseria* and neutrophils. *Nat Rev Microbiol.* 2012;10:178–190.
8. Kao WA, Pětrošová H, Ebady R, et al. Identification of Tp0751 (Pallilysin) as a *Treponema pallidum* vascular adhesin by heterologous expression in the Lyme disease spirochete. *Sci Rep.* 2017;7:1538.
9. Rouphael NG, Stephens DS. *Neisseria meningitidis:* biology, microbiology, and epidemiology. *Methods Mol Biol.* 2012;799:1–20.
10. Eloe-Fadrosh EA, Rasko DA. The human microbiome: from symbiosis to pathogenesis. *Annu Rev Med.* 2013;64:145–163.
11. van Rensburg JJ, Lin H, Gao X, et al. The human skin microbiome associates with the outcome of and is influenced by bacterial infection. *MBio.* 2015;6:e01315 –e01315.
12. Gallo RL, Hooper LV. Epithelial antimicrobial defence of the skin and intestine. *Nat Rev Immunol.* 2012;12:503–516.
13. LaRock CN, Todd J, LaRock DL, et al. IL-1β is an innate immune sensor of microbial proteolysis. *Sci Immunol.* 2016;1:(2).

14. Cox DL, Chang P, McDowall AW, et al. The outer membrane, not a coat of host proteins, limits antigenicity of virulent *Treponema pallidum*. *Infect Immun*. 1992;60:1076–1083.
15. Salazar JC, Hazlett KR, Radolf JD. The immune response to infection with *Treponema pallidum*, the stealth pathogen. *Microbes Infect*. 2002;4:1133–1140.
16. Kumar S, Ingle H, Prasad DV, et al. Recognition of bacterial infection by innate immune sensors. *Crit Rev Microbiol*. 2013;39:229–246.
17. Olive C. Pattern recognition receptors: sentinels in innate immunity and targets of new vaccine adjuvants. *Expert Rev Vaccines*. 2012;11:237–256.
18. Salazar JC, Cruz AR, Pope CD, et al. *Treponema pallidum* elicits innate and adaptive cellular immune responses in skin and blood during secondary syphilis: a flow-cytometric analysis. *J Infect Dis*. 2007;195:879–887.
19. LaRock CN, Döhrmann S, Todd J, et al. Group A streptococcal M1 protein sequesters cathelicidin to evade innate immune killing. *Cell Host Microbe*. 2015;18(4):471–477.
20. Ricklin D, Hajishengallis G, Yang K, et al. Complement: a key system for immune surveillance and homeostasis. *Nat Immunol*. 2010;11:785–797.
21. McNeil LK, Zagursky RJ, Lin SL, et al. Role of factor H binding protein in *Neisseria meningitidis* virulence and its potential as a vaccine candidate to broadly protect against meningococcal disease. *Microbiol Mol Biol Rev*. 2013;77:234–252.
22. Tufts DM, Hart TM, Chen GF, et al. Outer surface protein polymorphisms linked to host-spirochete association in Lyme borreliae. *Mol Microbiol*. 2019;111(4):868–882.
23. Madar M, Bencurova E, Mlynarcik P, et al. Exploitation of complement regulatory proteins by *Borrelia* and *Francisella*. *Mol Biosyst*. 2015;11:1684 –1195.
24. Pausa M, Pellis V, Cinco M, et al. Serum-resistant strains of *Borrelia burgdorferi* evade complement-mediated killing by expressing a CD59-like complement inhibitory molecule. *J Immunol*. 2003;170:3214–3222.
25. Blanco DR, Champion CI, Lewinski MA, et al. Immunization with *Treponema pallidum* outer membrane vesicles induces high-titer complement-dependent treponemicidal activity and aggregation of *T. pallidum* rare outer membrane proteins (TROMPs). *J Immunol*. 1999;163(5):2741–2746.
26. Sansonetti PJ, Phalipon A. M cells as ports of entry for enteroinvasive pathogens: mechanisms of interaction, consequences for the disease process. *Semin Immunol*. 1999;11:193–203.
27. Kawai T, Akira S. The role of pattern-recognition receptors in innate immunity: update on Toll-like receptors. *Nat Immunol*. 2010;11:373–384.
28. Brinkmann V, Reichard U, Goosmann C, et al. Neutrophil extracellular traps kill bacteria. *Science.* 2004;303:1532–1535.
29. Cervantes JL, La Vake CJ, Weinerman B, et al. Human TLR8 is activated upon recognition of *Borrelia burgdorferi* RNA in the phagosome of human monocytes. *J Leukoc Biol*. 2013;94(6):1231–1241.
30. Moore MW, Cruz AR, LaVake CJ, et al. Phagocytosis of *Borrelia burgdorferi* and *Treponema pallidum* potentiates innate immune activation and induces gamma interferon production. *Infect Immun*. 2007;75(4):2046–2062.
31. Roilides E, Simitsopoulou M, Katragkou A, et al. How biofilms evade host defenses. *Microbiol Spectr.* 2015;3:3.
32. McBroom RL, Styles AR, Chiu MJ, et al. Secondary syphilis in persons infected with and not infected with HIV-1: a comparative immunohistologic study. *Am J Dermatopathol*. 1999;21:432–441.
33. Van Voorhis WC, Barrett LK, Koelle DM, et al. Primary and secondary syphilis lesions contain mRNA for Th1 cytokines. *J Infect Dis*. 1996;173:491–495.
34. Gross DM, Steere AC, Huber BT. T helper 1 response is dominant and localized to the synovial fluid in patients with Lyme arthritis. *J Immunol.* 1998;160:1022–1028.
35. Cole J, Aberdein J, Jubrail J, et al. The role of macrophages in the innate immune response to *Streptococcus pneumoniae* and *Staphylococcus aureus:* mechanisms and contrasts. *Adv Microb Physiol*. 2014;65:125–202.
36. Peck A, Mellins ED. Precarious balance: Th17 cells in host defense. *Infect Immun*. 2010;78:32–38.
37. Hastey CJ, Elsner RA, Barthold SW, Baumgarth N. Delays and diversions mark the development of B cell responses to *Borrelia burgdorferi* infection. *J Immunol*. 2012;188(11):5612–5622.
38. Zhang JR, Hardham JM, Barbour AG, et al. Antigenic variation in Lyme disease borreliae by promiscuous recombination of VMP-like sequence cassettes. *Cell.* 1997;89:275–285.
39. Giacani L, Molini BJ, Kim EY, et al. Antigenic variation in *Treponema pallidum:* TprK sequence diversity accumulates in response to immune pressure during experimental syphilis. *J Immunol*. 2010;184:3822–3829.
40. Baessler A, Hale JS. Recurrent tonsillitis Tfh cells acquire a killer identity. *Trends Immunol*. 2019;40(5):377–379.
41. Hotchkiss RS, Karl IE. The pathophysiology and treatment of sepsis. *New Engl J Med*. 2002;348:138–150.
42. Needham BD, Trent MS. Fortifying the barrier: the impact of lipid A remodelling on bacterial pathogenesis. *Nat Rev Microbiol*. 2013;11:467–481.

Host Defenses to Fungi

Tobias M. Hohl, Bruce S. Klein, and Michail S. Lionakis

Advances in modern medicine exemplified by immunosuppressive therapies for autoimmunity, precision, and myeloablative therapies for cancer, and hematopoietic stem cell and solid organ transplantation for hematological malignancies and end-organ failure have resulted in a significant expansion of patient populations at risk for developing life-threatening opportunistic fungal infections.[1] The acquired immunodeficiency syndrome (AIDS) pandemic continues to cause significant mortality associated with AIDS-defining mycoses, primarily cryptococcosis and *Pneumocystis* pneumonia in developing regions of the globe. Notably, although there are ~5 million different fungal species, only a few cause human disease (Table 28.1); this is explained by the inability of most fungi to grow at human body temperatures and the robust innate and adaptive antifungal immune responses in immunocompetent individuals.

In recent years, we have witnessed significant progress in understanding (1) the cellular and molecular factors that promote protective antifungal immunity in humans and (2) the genetic and pharmacological factors that heighten human susceptibility to opportunistic mycoses. This immunological knowledge is critical in informing therapeutic and vaccination strategies to combat life-threatening fungal infections in vulnerable patients. Here, we outline recent mechanistic advances on the role of fungal recognition pathways in human antifungal immunity, and we highlight the molecular and cellular basis of host defense against medically important *Candida, Cryptococcus, Aspergillus, Mucorales, Histoplasma, Coccidioides*, and *Blastomyces* with a focus on observations that have clinical and translational implications.

FUNGAL RECOGNITION PATHWAYS AND THEIR CONTRIBUTION TO HUMAN ANTIFUNGAL HOST DEFENSE

Sensing of fungal pathogen-associated molecular patterns (PAMPs) by pattern recognition receptors (PRRs) of innate immune and epithelial cells is the first step in mounting antifungal immune responses. PRR engagement induces signaling cascades that promote the orchestrated recruitment and activation of innate and adaptive immune cells in infected tissues to achieve fungal clearance.[2–5] The C-type lectin receptors (CLRs) and Toll-like receptors (TLRs) are the principal fungal-sensing PRRs, whereas retinoic acid-inducible gene I-like receptors (RLRs) and nucleotide-binding oligomerization domain-like receptors (NLRs) also recognize fungi (Table 28.2); notably, collaborative PRR engagement also occurs and can be exploited clinically as in human chromoblastomycosis.

A major recent breakthrough in the fungal immunology field has been the demonstration of the indispensable contribution of CLR-mediated fungal recognition in human antifungal immunity and the delineation of the function of several members of the CLR signaling pathway (Fig. 28.1).

In brief, binding of Dectin-1 by β-glucan and of Dectin-2, Dectin-3, or Mincle by other fungal polysaccharide PAMPs (e.g., α-mannan) sequentially activates spleen tyrosine kinase (Syk), the CARD9/MALT1/BCL10 complex, and nuclear factor kappa B (NF-κB), which results in inflammasome activation, proinflammatory cytokine and chemokine production, leukocyte recruitment and activation, and Th17 differentiation.[2–4]

CLINICAL RELEVANCE

Immunotherapy Lessons Derived From Improved Basic Understanding of Antifungal Immune Responses

- Collaboration between different fungal pattern recognition receptors may augment downstream protective responses as shown in human chromoblastomycosis where the topical application of the TLR7 agonist imiquimod on skin fungal lesions resulted in clinical remission via synergistic TLR7-Mincle–mediated immunity.
- Adult-onset invasive fungal disease caused by neutralizing autoantibodies against IFN-γ or GM-CSF may respond to B-cell– or plasma cell–depleting therapy with anti-CD20– or anti-CD38–targeted monoclonal antibodies.
- Disseminated infections by intramacrophagic fungi caused by mutations in *IL12RB1* or *STAT4* or by partial IFN-γR1 deficiency may respond to adjunct recombinant IFN-γ treatment.

KEY CONCEPTS

The C-Type Lectin Receptor Pathway

- Dectin-1 binding activates Syk via Src-dependent phosphorylation of its immunoreceptor tyrosine-based activation motif (ITAM), which involves recruitment of the tyrosine phosphatase SHP-2.
- Dectin-2, Dectin-3, and Mincle binding activates Syk via engagement of the ITAM-containing adaptor FcRγ.
- Downstream of Syk activation, the protein kinase C-δ and Vav proteins phosphorylate CARD9, which leads to the formation of the CARD9/MALT1/BCL10 complex and activation of the canonical nuclear factor kappa B (NF-κB) pathway.
- Binding of Dectin-1 can also activate the noncanonical NF-κB pathway (via RAF-1 engagement) or the ERK pathway (via H-Ras and Ras-GRF1 activation).
- CLR/Syk signaling may also negatively regulate antifungal immune responses through JNK1 signaling (via the C-type lectin receptor CD23/FCER2A) or the E3-ubiquitin ligase CBLB. Concordantly, inhibition of JNK1 or CBLB restores antifungal effector functions of phagocytes and improves survival in mouse models of invasive fungal disease.

TABLE 28.1 Common Human Fungal Diseases

Fungal Infection (Most Common Fungal Genera or Species)	Fungal Morphotype	Clinical Syndromes	Common Patient Populations With Acquired Conditions That Place Them At Risk for the Indicated Clinical Syndromes
Candidiasis (*Candida albicans, C. glabrata, C. tropicalis, C. parapsilosis, C. auris*)	Yeast (±pseudohyphae depending on the genera)	Oropharyngeal candidiasis	AIDS
		Esophageal candidiasis	AIDS
		Vulvovaginal candidiasis	Antibiotic use
		Candidemia	Critical illness (ICU)
		Disseminated infection with seeding of various deep-seated organs (kidney, liver, spleen, brain, bone)	Iatrogenic immunosuppression (neutropenia, corticosteroid use)
Cryptococcosis (*Cryptococcus neoformans, C. gattii*)	Yeast	Pneumonia	AIDS, corticosteroid use
		Meningoencephalitis	AIDS
		Disseminated infection	AIDS
Aspergillosis (*Aspergillus fumigatus, A. flavus, A. terreus, A. niger, A. ustus, A. nidulans*)	Mold	Pneumonia	Neutropenia, HSCT
		Disseminated infection	Neutropenia, HSCT
		Keratitis	Direct inoculation
		Allergic bronchopulmonary aspergillosis	Atopy
		Chronic cavitary lung disease	Structural lung disease
Mucormycosis (*Rhizopus, Mucor, Absidia, Rhizomucor, Cunninghamella, Saksenaea, Lichtheimia*)	Mold	Sinopulmonary infection	Neutropenia, HSCT
		Rhinocerebral infection	Diabetic ketoacidosis
		Necrotizing skin infection	Direct inoculation (e.g., victims of natural disasters)
Fusariosis (*Fusarium solani, F. oxysporum, F. proliferatum*)	Mold	Pneumonia	Neutropenia
		Disseminated infection	Neutropenia
		Keratitis	Direct inoculation
Scedosporiosis (*Scedosporium apiospermum, Lomentospora prolificans*)	Mold	Pneumonia	Neutropenia, HSCT
		Disseminated infection	Neutropenia, HSCT
		Skin and subcutaneous tissue infections	Direct inoculation
Histoplasmosis (*Histoplasma capsulatum*)	Dimorphic fungus	Pneumonia	Healthy individuals
		Disseminated infection	AIDS, SOT
		Fibrosing mediastinitis	
Coccidioidomycosis (*Coccidioides immitis, C. posadasii*)	Dimorphic fungus	Pneumonia	Healthy individuals
		Disseminated infection (bone, CNS)	AIDS
Paracoccidioidomycosis (*Paracoccidioides brasiliensis*)	Dimorphic fungus	Pneumonia	Healthy individuals
		Disseminated infection (skin, bone, mucosal surfaces)	AIDS
Blastomycosis (*Blastomyces dermatitidis, B. gilchristii*)	Dimorphic fungus	Pneumonia	Healthy individuals
		Disseminated infection (skin, genitourinary tract, bone, mucosal surfaces)	AIDS
Sporotrichosis (*Sporothrix schenkii*)	Dimorphic fungus	Lymphocutaneous disease (ascending lymphangitis)	Direct inoculation
		Disseminated infection	AIDS
Taralomycosis (*Talaromyces marneffei*)	Dimorphic fungus	Pneumonia	Healthy individuals
		Disseminated infection (skin, bone, mucosal surfaces)	AIDS
Pneumocystosis (*Pneumocystis jirovecii*)	Cysts and trophozoites	Pneumonia	AIDS, cancer
		Disseminated infection	AIDS
Dermatophytosis (*Epidermophyton, Trichophyton, Microsporum*)	Mold	Skin and nail infections	Healthy individuals
Chromoblastomycosis (*Fonsecaea pedrosoi, F. monophora*)	Yeast	Skin and subcutaneous tissue infections	Healthy individuals
Eumycetoma (*Maadurella* spp.)	Mold	Skin and subcutaneous tissue infections	Healthy individuals
Phaeohyphomycosis (*Exophiala* spp., *Cladophialophora* spp., *Alternaria* spp., *Phialophora* spp., *Rhinocladiella* spp.)	Yeast or mold	Pneumonia	Healthy individuals
		CNS infection	Healthy individuals, iatrogenic immunosuppression
		Skin infection	Direct inoculation
		Disseminated infection	HSCT, iatrogenic immunosuppression

AIDS, Acquired immunodeficiency syndrome; *CNS*, central nervous system; *HSCT*, hematopoietic stem cell transplantation; *ICU*, intensive care unit; *SOT*, solid organ transplantation.

TABLE 28.2 Fungal Pathogen–Associated Molecular Patterns and Their Associated Pattern Recognition Receptors

PRR (Gene)	PRR Class	Fungal PAMPs	Fungal Genera
Dectin-1 (CLEC7A)	CLR	β-Glucan	*Candida, Cryptococcus, Aspergillus, Histoplasma, Coccidioides, Paracoccidioides, Pneumocystis, Exserohilum*
Dectin-2 (CLEC6A)	CLR	α-Mannans *O*-linked mannoproteins	*Candida, Aspergillus, Coccidioides, Paracoccidioides, Blastomyces*
Dectin-3 (CLEC4D)	CLR	α-Mannans	*Candida*
MelLec (CLEC1A)	CLR	DHN-melanin	*Aspergillus, Fonsecaea, Cladosporium*
Mincle (CLEC4E)	CLR	Glyceroglycolipids α-Mannose	*Candida, Fonsecaea, Pneumocystis, Malassezia, Saccharomyces*
CD23 (FCER2A)	CLR	β-Glucan α-Mannans	*Candida*
DC-SIGN (CD209)	CLR	Mannans Galactomannan	*Candida, Aspergillus, Chrysosporium*
Mannose receptor (CD206)	CLR	Mannans *N*-linked mannans	*Candida, Pneumocystis*
TLR1	TLR	Glucuronoxylomannans	*Cryptococcus*
TLR2	TLR	α-Glucans Mannan Phospholipomannan Glucuronoxylomannan Chitin oligomers	*Candida, Cryptococcus*
TLR3	TLR	Fungal RNA	*Aspergillus*
TLR4	TLR	*O*-Linked mannans Glucuronoxylomannan	*Candida, Cryptococcus, Aspergillus, Scedosporium*
TLR6	TLR	Phospholipomannans Glucuronoxylomannans	*Candida*
TLR7	TLR	Fungal RNA	*Candida*
TLR9	TLR	Fungal DNA Chitin	*Candida, Cryptococcus, Aspergillus*
NOD1	NLR	Unknown	*Aspergillus*
NOD2	NLR	Chitin	*Candida*
NLRP3	NLR	Unknown	*Candida, Aspergillus*
NLRP10	NLR	Unknown	*Candida*
NLRC4	NLR	Unknown	*Candida*
MDA5	RLR	Unknown	*Candida, Aspergillus*
Galectin 3 (GAL3)	CLR	β-Mannosides	*Candida, Aspergillus, Paracoccidioides*
Pentraxin 3 (PTX3)		Galactomannan	*Aspergillus*

CLR, C-type lectin receptor; *DC-SIGN*, dendritic cell-specific intercellular adhesion molecule-3-grabbing nonintegrin; *DHN*, 1, 8-dihydroxynaphthalene; *MDA5*, melanoma differentiation-associated protein 5; *NLR*, nucleotide-binding oligomerization domain (NOD)–like receptor; *PAMP*, pathogen-associated molecular pattern; *PRR*, pattern recognition receptor; *RLR*, retinoic acid-inducible gene I (RIG-I)–like receptor; *TLR*, Toll-like receptor.

The importance of the CLR/Syk/CARD9 pathway in human antifungal immunity is clearly demonstrated by the identification and characterization of patients with autosomal-recessive CARD9 deficiency who develop severe refractory mucocutaneous and invasive fungal diseases. Remarkably, among the ~450 known human primary immunodeficiencies (PIDs), CARD9 deficiency is the only one that manifests specifically with fungal susceptibility without accompanying bacterial, viral. or parasitic infections, or allergic, autoimmune, or neoplastic manifestations. Notably, CARD9-deficient humans develop tissue-specific infections by a select group of fungi that include *Candida* (mucosal, central nervous system [CNS], bone), molds (CNS, intraabdominal), phaeohyphomycetes (CNS, lung, skin) and dermatophytes (skin) but not *Cryptococcus*, endemic dimorphic fungi, or *Pneumocystis*.[4,6]

Instead, inherited deficiencies in TLR (i.e., MYD88, IRAK4) or other non-CLR (i.e., MDA5) PRRs predispose to bacterial and/or viral, not fungal, infections underscoring the crucial and specific role of CLR/Syk/CARD9 signaling in human antifungal immunity and its ability to compensate for other PRR deficiencies.[7] However, single nucleotide polymorphisms (SNPs) in TLR pathway constituents are linked to increased risk of invasive candidiasis and aspergillosis in patients with critical illness and/or iatrogenic immunosuppression, indicating that TLR-dependent fungal recognition contributes to optimal human antifungal immune responses in certain settings.

IMMUNE RESPONSES TO YEAST FUNGI

Two major medically important yeast fungi are *Candida* and *Cryptococcus* spp. *Candida* is commensal in the gastrointestinal and female reproductive tract of most humans. When perturbations in host immune responses or microbiota occur, *Candida* can cause opportunistic infection in the form of either mucocutaneous or deep-seated invasive disease (see Table 28.1).[8] The major etiological agents of human *Candida* infections, which form interchanging yeast and pseudohyphal structures that vary depending on the species, are *C. albicans, C. glabrata, C. parapsilosis*, *C. tropicalis*, and the emerging drug-resistant *C. auris*. Importantly, the immune requirements for protection against mucocutaneous versus invasive *Candida* infections are clearly segregated, which is reflected in the distinct patient populations who are at risk for mucocutaneous versus invasive candidiasis. Indeed, whereas lymphocytes are critical for protection against mucocutaneous candidiasis, myeloid phagocytes are indispensable for protection against invasive candidiasis, but not vice versa.[4]

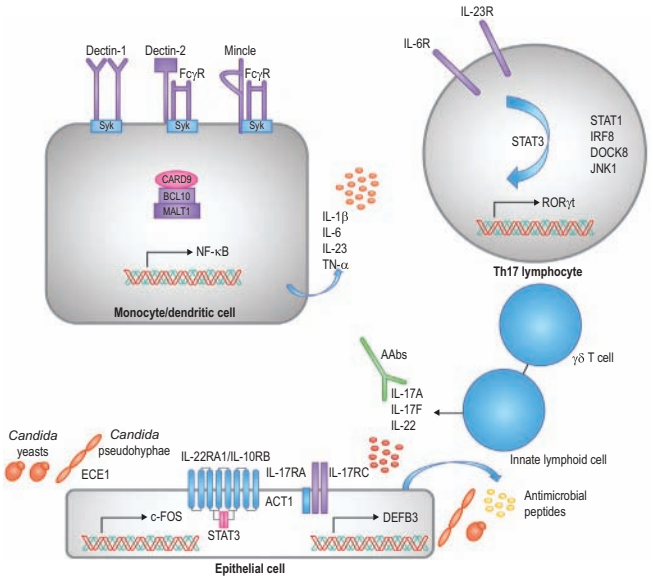


FIG. 28.1 Antifungal Immune Responses at the Mucosal Interface. *Candida albicans* sensing by C-type lectin receptors leads to sequential activation of Syk, the CARD9/MALT1/BCL10 complex and nuclear factor kappa B (*NF-κB*) signaling, which promotes the production of proinflammatory mediators that instruct Th17 differentiation. STAT3 promotes RORγt-mediated Th17 differentiation, which is defective in patients with Job's syndrome caused by dominant-negative *STAT3* mutations. CARD9, DOCK8, JNK1, and IRF8 also contribute to Th17 differentiation, and human inborn errors of immunity affecting each of these genes causes CMC. CMC also occurs in patients with *RORC* mutations who feature impaired Th17 differentiation. Heterozygous gain-of-function mutations in *STAT1* lead to a cytokine milieu that inhibits Th17 differentiation. Th17 cells produce IL-17A, IL-17F, and IL-22. IL-17A and IL-17F, via their binding to IL-17RA and IL-17RC receptors on epithelial cells of the suprabasal epithelial layer and the downstream adaptor protein ACT1, induce the epithelial cell production of antimicrobial peptides (β-defensins, S100A8/S100A9) that inhibit *Candida* growth at the mucosa. IL-17 cytokines are also produced by Tc17 cells, γδ T cells, and innate lymphoid cells, in frequencies that vary depending on the cell type and the mucosal surface. IL-22 is also produced by type-17 lymphocytes and binds to its receptor IL-22R1/IL-10RB on epithelial cells of the basal epithelial layer where it activates STAT3 to promote epithelial cell survival, proliferation, and repair. CMC develops in patients with mutations in *IL17F*, *IL17RA*, *IL17RC*, and *TRAF3IP2* (ACT1), which impair IL-17 receptor-dependent signaling, and in patients with thymoma or autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy caused by *AIRE* mutations associated with autoantibodies against Th17 cell-derived cytokines. The *Candida albicans* hyphal-associated peptide toxin candidalysin (encoded by Ece1) promotes epithelial cell c-FOS activation to produce IL-1α, IL-6, G-CSF, and GM-CSF after *Candida* infection. In humans, inherited mutations in the genes in *red font* have been described to feature CMC. Abbreviations: AAbs, autoantibodies; AIRE, autoimmune regulator; BCL10, B-cell lymphoma/leukemia 10; CARD9, caspase recruitment domain-containing protein 9; CMC, chronic mucocutaneous candidiasis; DOCK8, dedicator of cytokinesis 8; Ece1, extent of cell elongation protein 1; G-CSF; granulocyte colony-stimulating factor; GM-CSF, granulocyte–macrophage colony-stimulating factor; IRF8, interferon regulatory factor 8; JNK1, c-Jun N-terminal kinase 1; MALT1, mucosa-associated lymphoid tissue lymphoma translocation protein 1; RORγt, RAR-related orphan receptor γ; STAT, signal transducer and activator of transcription; Syk, spleen tyrosine kinase. (Modified from Lionakis MS, Netea MG, Holland SM. Mendelian genetics of human susceptibility to fungal infection. *Cold Spring Harb Perspect Med*, 2014;4[6]:a019638.)

KEY CONCEPTS

Fungus-Specific Immune Cell Requirements for Optimal Host Defense

- Different fungal pathogens depend on different arms of the immune system for protective host defense.
- Inhaled mold fungi (*Aspergillus, Mucorales, Fusarium, Scedosporium*) depend on professional phagocytes (neutrophils, monocytes/macrophages, dendritic cells) for effective host defense.
- Commensal *Candida* depends on IL-17-producing lymphoid cells and epithelial cells for host defense at the mucosal interface.
- Instead, *Candida* depends on professional phagocytes (neutrophils, monocytes/macrophages), not lymphocytes, for host defense during deep-seated invasive infection.
- Intramacrophagic fungi (*Cryptococcus, Histoplasma, Coccidioides, Paracoccidioides, Blastomyces, Talaromyces, Sporothrix*) depend on the cross-talk between IFN-γ-producing lymphocytes and IL-12-producing macrophages for effective host defense.

As such, AIDS patients are at risk for mucosal candidiasis whereas neutropenia, receipt of cytotoxic chemotherapy, and critical illness in intensive care unit (ICU) patients represent major risk factors for invasive candidiasis. Worldwide, *Candida* causes an estimated 400,000 life-threatening invasive infections annually with mortality that exceeds 40%. Moreover, although not life-threatening, the global burden of mucosal candidiasis is substantial as ~75% of all women develop at least one episode of vulvovaginal candidiasis (VVC) during their lifetime.

The two major etiological agents of cryptococcosis are *Cryptococcus neoformans* and *C. gattii*, which are ubiquitous spherical yeasts that are inhaled into the human lung where they are efficiently contained by tissue-resident phagocytes in immunocompetent individuals. However, in patients with immune compromise, particularly those with AIDS, *Cryptococcus* escapes the lungs and causes meningoencephalitis (see Table 28.1); in fact, cryptococcosis is the most common cause

of fungal meningitis in humans. Worldwide, cryptococcosis causes ~200,000 deaths annually, most in sub-Saharan Africa where access to antiretroviral therapy is limited, and accounts for ~15% of AIDS-related deaths.

Mucosal Candidiasis

Mucosal candidiasis manifests as oropharyngeal, esophageal, or vulvovaginal infection. Over the past decade, it has become evident that effective defense against oropharyngeal candidiasis (OPC) is critically dependent on functional IL-17 receptor (IL-17R) signaling (see Fig. 28.1).[9] Mechanistically, IL-17A and IL-17F are produced by mucosal αβ T cells (primarily natural Th17 cells and less so Tc17 cells), γδ T cells, and, to a lesser extent, innate lymphoid cells during OPC. Recent work has uncovered the contribution of the *C. albicans* hyphal-associated peptide toxin candidalysin in priming natural Th17 cell responses during OPC. Engagement of IL-17RA and IL-17RC by IL-17A/IL-17F on epithelial cells of the suprabasal epithelial layer promotes the production of antimicrobial molecules (β-defensins, S100a8/S100a9) that inhibit *Candida* growth and curtail its mucosal invasion. Recent work has shed light on the mechanisms by which IL-22, another type-17 lymphocyte-derived cytokine, protects during OPC. Specifically, engagement of IL-22RA1 by IL-22 on epithelial cells of the basal epithelial layer activates STAT3, which promotes epithelial cell survival and proliferation and replenishes the IL-17RA-expressing basal epithelial layer, thereby enabling its response to IL-17A/IL-17F. Thus, IL-17 and IL-22 exert spatially distinct, yet cooperative, protective functions during OPC. Interestingly, as opposed to OPC, IL-17R signaling appears dispensable for host defense against VVC, underscoring the differential host factors that control *Candida* in various mucosal surfaces. Indeed, AIDS patients are at risk for OPC and esophageal candidiasis but not VVC. Instead, VVC is typically seen following antibiotic treatment, highlighting the role of vaginal microbiota in restricting local *Candida* growth.

Invasive Candidiasis

C. albicans causes the majority of invasive *Candida* infections, although *C. glabrata, C. tropicalis*, and *C. parapsilosis* are increasingly observed in recent years. Drug-resistant *C. auris*, which features a propensity for prolonged persistence on the human skin, is an emerging public health concern associated with large outbreaks in healthcare facilities. Herein, we focus on immunity to *C. albicans*, which has been studied in-depth relative to non-*albicans Candida*.

Three major conditions predispose humans to invasive candidiasis.[8] First, broad-spectrum antibiotics deplete certain commensal gut microbiota and the antifungal protective factors they secrete, enabling increased *Candida* gut colonization.[10] Second, breach of gastrointestinal (i.e., gastrointestinal surgery, chemotherapy-induced mucositis) and cutaneous (i.e., central venous catheters) barriers permit *Candida* translocation into the bloodstream. Third, iatrogenic immunosuppression (i.e., neutropenia, corticosteroid use) impairs innate anti-*Candida* responses and promotes *Candida* dissemination and tissue invasion.

Myeloid phagocytes, but not lymphocytes, mediate protective immunity during invasive candidiasis.[11] Prompt neutrophil recruitment and activation in *Candida*-infected tissues is critical for effective host defense, in keeping with profound susceptibility of neutropenic mice and humans to invasive candidiasis. Recent work has uncovered the fungal and host factors that promote protective neutrophil recruitment in the *Candida*-infected CNS. Specifically, CNS-resident microglia respond to Candidalysin in a p38- and c-Fos-dependent manner to produce IL-1β, which drives CXCL1 production by microglia to recruit CXCR2-expressing neutrophils in the infected CNS. This intricate network relies on functional CARD9 on microglia for generation of both IL-1β (at the levels of transcriptional pro-IL-1β induction and inflammasome-dependent IL-1β maturation) and CXCL1 and is impaired in CARD9-deficient patients who manifest CNS-specific neutropenia that underlies their CNS-targeted candidiasis phenotype.[12]

Upon recruitment into tissues, neutrophils effectively inhibit *Candida* growth via oxidative and non-oxidative mechanisms that vary depending on fungal opsonization; specifically, NADPH oxidase, Syk, FcγR, and protein kinase C operate during opsonized *Candida* killing, whereas Syk, CARD9, complement receptor 3 (CR3), and phosphoinositide 3-kinase enable unopsonized *Candida* killing. A neutrophil-natural killer (NK) cell cross-talk was recently described to drive NK cell-induced GM-CSF-dependent priming of neutrophil candidacidal activity. Inflammatory monocytes, via IL-15, and monocyte-derived dendritic cells (Mo-DCs), via IL-23, activate NK cells to produce GM-CSF.[4] In certain settings, such as upon neutrophil recovery, neutrophils may exert immunopathogenic effects causing tissue injury. Factors associated with neutrophil-driven immunopathology have recently been defined in mice and humans (i.e., CCR1, TEC, MCPIP1, IFN-γ) and may represent therapeutic targets in selected patients.[4]

Monocytes/macrophages are also critical for host defense during invasive candidiasis via fungal uptake and killing. CCR2 recruits inflammatory monocytes and promotes fungal control in the kidney and CNS, whereas tissue-resident macrophages depend upon CX3CR1 for survival and effector function in the kidney, liver, and CNS. Monocytes/macrophages also protect during *C. albicans* re-infection by exhibiting innate immunological memory, termed trained immunity. Mechanistically, β-glucan engagement of Dectin-1 on mouse and human monocytes activates a RAF-1/AKT/mTOR/HIF-1α pathway that leads to epigenetic reprogramming with histone trimethylation and acetylation and metabolic reprogramming with switching from oxidative phosphorylation to aerobic glycolysis.[4]

Cryptococcosis

An effective cross-talk between tissue-resident and recruited mononuclear phagocytes (monocytes/macrophages, microglia, DCs) and T lymphocytes is critical for *Cryptococcus* control within the lung and CNS.[13,14] In agreement, patients with defects in monocyte/macrophage function and lymphocytopenia are at risk for cryptococcosis. Instead, neutropenia is not a risk factor for human cryptococcosis and, in fact, neutrophils contribute to immunopathology in *Cryptococcus*-infected mice.

Cryptococcus PAMP recognition is redundantly mediated by several TLRs, CLRs, and NLRs and is evaded by cryptococcal virulence factors, principally the polysaccharide capsule. A protective cryptococcal response that leads to fungal eradication is associated with the induction of proinflammatory type-1 immunity with generation of an IFN-γ/IL-12/IL-18/TNF–dominant cytokine milieu, M1 (classical) macrophage activation, and Th1 differentiation.[14,15] Indeed, IFN-γ– and TNF–predominant *Cryptococcus*-specific $CD4^+$ T-cell responses and increased IFN-γ and TNF levels in the *Cryptococcus*-infected cerebrospinal fluid are associated with improved prognosis of cryptococcosis in AIDS patients. In the setting of protective type 1 immunity,

upon cryptococcal yeast internalization within the phagosome, phagosome–lysosome fusion and acidification ensue, resulting in yeast degradation that is associated with the production of reactive oxygen and nitrogen species. Phagocytes, primarily DCs, present cryptococcal antigens to T lymphocytes to initiate and direct protective adaptive immune responses. Instead, a nonprotective cryptococcal response that results in fungal persistence is associated with the induction of a type 2 immune responses, the generation of an IL-4/IL-5/IL-13–dominant cytokine milieu, M2 (alternative) macrophage activation with pathogen extracellular escape, eosinophil recruitment, and Th2 differentiation.[14,15]

Among lymphocytes, both CD4 and CD8 T cells can inhibit *Cryptococcus* growth via direct killing and via proinflammatory cytokine production that recruits and activates phagocytes for fungal killing. B lymphocytes and antibodies also contribute to protection via a variety of mechanisms that include fungal opsonization for efficient phagocyte uptake. However, T lymphocytes may also exert pathogenic effects during cryptococcosis, as exemplified by the development of immune reconstitution inflammatory syndrome in AIDS patients following initiation of antiretroviral therapy and immune restoration, during which enriched frequencies of Ki-67$^+$ PD-1$^+$ Th1- and Th17-biased CD4 T cells accumulate, associated with increased levels of proinflammatory cytokines (IL-6, IL-7, IFN-γ, TNF) and worsening clinical manifestations that require immunomodulatory treatment for amelioration.[4]

PATIENTS AT RISK FOR YEAST INFECTIONS DUE TO GENETIC OR PHARMACOLOGICAL FACTORS

Mucosal Candidiasis

Concordant with the critical contribution of IL-17R signaling to mucosal antifungal defense, patients with genetic deficiency in IL-17RA, IL-17RC, IL-17F, or the IL-17R adaptor TRAF3IP2 (ACT1) develop chronic mucocutaneous candidiasis (CMC) without invasive candidiasis (Table 28.3).[16] IL-17RA and ACT1 (but not IL-17RC or IL-17F) deficiencies also underlie skin staphylococcal and pulmonary bacterial infection susceptibility. Several other CMC-manifesting monogenic disorders are associated with direct or indirect impairment of IL-17R-dependent responses including mutations in *RORC, STAT1, STAT3, ZNF341, DOCK8, IRF8, MALT1, BCL10, IL12B, IL12RB1, CARD9*, and *JNK1* (see Table 28.3).[4,17] CMC is also seen in diseases associated with autoantibodies against Th17 cell-derived cytokines such as thymoma and autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy caused by *AIRE* mutations. Moreover, the use of biological agents targeting various constituents of IL-17R signaling (i.e., IL-12p40, IL-23p19, IL-17A, IL-17A/IL-17F, IL-17RA) for psoriasis and inflammatory bowel disease has provided another layer of evidence for the importance of IL-17 in human mucosal antifungal defense, as these patients occasionally develop refractory mucosal candidiasis (frequency, ~2% to 4%).

TABLE 28.3 Inborn Errors of Immunity That Manifest With Fungal Infection Susceptibility in Humans

Gene Name	Mode of Inheritance (Clinical Syndrome, If Applicable)	Fungal Infection Susceptibility	Nonfungal Infection Manifestations	Immunological Mechanisms Accounting for Fungal Infection Susceptibility
IL17F	AD	CMC	Atopy	Impaired IL-17R immune responses
IL17RA	AR	CMC	Skin staphylococcal infections Bacterial pneumonias Atopy	Impaired IL-17R immune responses
IL17RC	AR	CMC	None	Impaired IL-17R immune responses
TRAF3IP2	AR	CMC	Skin staphylococcal infections Bacterial pneumonias Atopy	Impaired IL-17R immune responses
RORC	AR (MSMD)	CMC	Mycobacterial infections	Impaired Th17 differentiation
DOCK8	AR (hyper-IgE syndrome)	CMC	Viral skin infections, eczema, malignancies	
STAT1	AD	CMC, aspergillosis, mucormycosis, cutaneous fusariosis, histoplasmosis, coccidioidomycosis	Bacterial and viral infections, multisystem autoimmunity, aneurysms, hypothyroidism, malignancies	Impaired Th17 differentiation
STAT3	AD (hyper-IgE syndrome; Job syndrome)	CMC, dermatophytosis, cryptococcosis, aspergillosis, scedosporiosis, histoplasmosis, coccidioidomycosis	Eczema, bacterial skin and lung infections, aneurysms, skeletal abnormalities	Impaired Th17 differentiation
ZNF341	AR	CMC	Eczema, bacterial infections, skeletal abnormalities	Impaired Th17 differentiation
MAPK8	AD	CMC	Skin staphylococcal infections, bacterial urinary tract infections, Ehlers-Danlos syndrome-like connective tissue disorder	Impaired Th17 differentiation, impaired IL-17R immune responses
AIRE	AR or AD (APECED)	CMC	Multisystem endocrine and nonendocrine autoimmunity, ectodermal dystrophy	Autoantibodies to Th17 cell-derived cytokines
MALT1	AR	CMC	Viral and bacterial infections, bronchiectasis, hypogammaglobulinemia	Impaired T-lymphocyte activation
BCL10	AR	CMC	Viral and mycobacterial infections, hypogammaglobulinemia	Lymphocytopenia

Continued

TABLE 28.3 Inborn Errors of Immunity That Manifest With Fungal Infection Susceptibility in Humans—cont'd

Gene Name	Mode of Inheritance (Clinical Syndrome, If Applicable)	Fungal Infection Susceptibility	Nonfungal Infection Manifestations	Immunological Mechanisms Accounting for Fungal Infection Susceptibility
IRF8	AR	CMC	Mycobacterial infections	Decreased number of Th17 cells
CLEC7A	AR	Vaginal candidiasis, onychomycosis	None	Decreased IL-17 production
CARD9	AR	CMC, CNS candidiasis, extrapulmonary aspergillosis, cutaneous mucormycosis, phaeohyphomycosis, deep dermatophytosis	None	Decreased number of Th17 cells (CMC), impaired microglial-neutrophil cross-talk resulting in CNS neutropenia (CNS candidiasis), impaired neutrophil *Candida* killing (invasive candidiasis)
CYBA	AR (CGD)	Invasive candidiasis (<5%), aspergillosis (~40%)	Infections by *Staphylococcus, Nocardia*, Serratia, and *Burkholderia*, colitis	Lack of superoxide generation
CYBB	X-linked (CGD)	Invasive candidiasis (<5%), aspergillosis (~40%)	Infections by *Staphylococcus, Nocardia*, Serratia, and *Burkholderia*, colitis	Lack of superoxide generation
NCF1	AR (CGD)	Invasive candidiasis (<5%), aspergillosis (~40%)	Infections by *Staphylococcus, Nocardia*, Serratia, and *Burkholderia*, colitis	Lack of superoxide generation
NCF2	AR (CGD)	Invasive candidiasis (<5%), aspergillosis (~40%)	Infections by *Staphylococcus*, *Nocardia, Serratia,* and *Burkholderia*, colitis	Lack of superoxide generation
NCF4	AR (CGD)	Histoplasmosis	Colitis	Lack of superoxide generation
MPO	AR	Invasive candidiasis (<5%)	None	Lack of hypochlorous acid generation
ELA2	AR (SCN)	Invasive candidiasis, aspergillosis	Pyogenic infections, periodontitis	Neutropenia
HAX1	AR (SCN)	Invasive candidiasis, aspergillosis	Pyogenic infections, periodontitis	Neutropenia
CD18	AR (LAD)	Invasive candidiasis, aspergillosis	Impaired wound healing, severe periodontitis, colitis	Impaired trafficking of neutrophils to infected tissues
CTSC	AR (Papillon-Lefevre syndrome)	Mucormycosis	Pyogenic infections, periodontitis, palmoplantar keratoderma	Impaired activation of granule serine proteases
GATA2	AD (MonoMAC syndrome)	Cryptococcosis, aspergillosis, histoplasmosis, blastomycosis	Viral and mycobacterial infections, leukemia, lymphedema	Monocytopenia, decreased DCs, neutrophil granule abnormalities
IFNGR1	AD (MSMD)	Histoplasmosis, coccidioidomycosis	Intramacrophagic pathogen infections	Impaired IFN-γ cellular responses
IL12B	AR (MSMD)	CMC	Mycobacterial infections	Impaired Th17 differentiation
IL12RB1	AR or AD (MSMD)	CMC, cryptococcosis, histoplasmosis, coccidioidomycosis, paracoccidioidomycosis	Intramacrophagic pathogen infections	Impaired Th17 differentiation (CMC), impaired IL-12/IL-23-dependent IFN-γ production (systemic mycoses)
STAT4	AD	Paracoccidioidomycosis	None	Impaired IFN-γ production
CD40L	X-linked	PCP	Bacterial, mycobacterial, and parasitic infections, colitis	Impaired T-lymphocyte responses
NEMO/ IKBKG	X-linked	CMC, PCP	Bacterial, mycobacterial, and viral infections, anhidrotic ectodermal dysplasia	Lymphocytopenia
IKBA	AD	CMC, PCP	Bacterial, mycobacterial, and viral infections, anhidrotic ectodermal dysplasia, colitis	Lymphocytopenia
IL21R	AR	CMC, PCP	Cryptosporidiosis with cholangitis and liver fibrosis	Impaired T-lymphocyte activation
BTK	X-linked (XLA)	PCP, cryptococcosis	Bacterial infections, hypogammaglobulinemia, colitis	Impaired B-lymphocyte signaling

AD, Autosomal dominant; *AIRE*, autoimmune regulator; *APECED*, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy; *AR*, autosomal recessive; *BCL10*, B-cell lymphoma/leukemia 10; *BTK*, Bruton tyrosine kinase; *CARD9*, caspase recruitment domain-containing protein 9; *CMC*, chronic mucocutaneous candidiasis; *CTSC*, cathepsin C; *CYB*, cytochrome B-245; *DCs*, dendritic cells; *DOCK8*, dedicator of cytokinesis 8; *ELA2*, elastase 2; *GATA2*, GATA binding protein 2; *HAX1*, HCLS1-associated protein X-1; *IKBA*, NF-κB inhibitor alpha; *IKBKG*, Inhibitor of NF-κB kinase regulatory subunit γ; *IL-17R*, interleukin receptor; *IRF8*, interferon regulatory factor 8; *JNK1*, c-Jun N-terminal kinase 1; *LAD*, leukocyte adhesion disorder; *MALT1*, mucosa-associated lymphoid tissue lymphoma translocation protein 1; *MAPK8*, mitogen-activated protein kinase 8; *MPO*, myeloperoxidase; *MSMD*, mendelian susceptibility to mycobacterial disease; *NCF*, neutrophil cytosol factor; *NEMO*, NF-κB essential modulator; *PCP*, *Pneumocystis* pneumonia; *RORC*, RAR orphan receptor C; *SCN*, severe congenital neutropenia; *STAT*, signal transducer and activator of transcription; *TRAF3IP2*, TRAF interacting protein 2; *XLA*, X-linked agammaglobulinemia; *ZNF341*, zinc finger protein 341.

Invasive Candidiasis

As mentioned above, CARD9 deficiency predisposes to mucosal candidiasis, associated with impaired Th17 differentiation, and CNS-targeted invasive candidiasis, caused by impaired microglial-neutrophil cross-talk (see Table 28.3).[6] In fact, CARD9 deficiency is the only known inherited condition that predisposes to both mucosal and invasive candidiasis, in keeping with the central positioning of CARD9 in antifungal immune responses. The advent of Syk inhibitors for the treatment of autoimmune and neoplastic conditions is likely to promote iatrogenic fungal susceptibility, as suggested by early reports of mucocutaneous fungal disease in Syk inhibitor-treated individuals.

Furthermore, a subset of patients with inherited disorders in the phagocyte oxidative burst machinery featuring complete myeloperoxidase deficiency or chronic granulomatous disease (CGD), caused by mutations in the NADPH oxidase complex, develop invasive candidiasis without associated iatrogenic risk factors (frequency, <5%) (see Table 28.3). The observation that the majority of myeloperoxidase-deficient and CGD patients do not develop invasive candidiasis indicates that intact mucocutaneous barriers and nonoxidative-dependent fungal killing mechanisms compensate for absent oxidative-dependent phagocyte functions in these patients.[7,18]

In ICU patients, SNPs in PRRs (*TLR1*), cytokines (*IL10, IL12B, TNF*), chemokines and their receptors (*CCL8, CXCR1, CX3CR1*), interferon signaling molecules (*STAT1, SP110, PSMB8*), and other immune factors (*TAGAP, CD58, LCE4A-C1orf68*) have been linked to heightened risk of invasive candidiasis and/or worse outcome following invasive candidiasis; some of them (i.e., *CXCR1-T276, CX3CR1-M280*, others) were shown to be dysfunctional as they confer phagocyte defects similar to those observed in the corresponding gene-deficient mouse phagocytes. Collectively, these SNPs show promise for the development of precision-medicine risk stratification, prophylaxis, and prognostication strategies in ICU patients.[8]

Cryptococcosis

Cryptococcosis develops in patients with mutations in the IL-12/IFN-γ signaling pathway, which underlies Mendelian susceptibility to mycobacterial disease and also predisposes to disseminated infections by other intracellular fungal (endemic dimorphic fungi, see below) and bacterial (*Salmonella*) pathogens (see Table 28.3, Fig. 28.2). These patients exhibit defective IL-12/IFN-γ-dependent lymphocyte-macrophage cross-talk that impairs intramacrophagic cryptococcal clearance. Mirroring this inherited susceptibility, cryptococcosis is also seen in adults with autoantibodies against IFN-γ. Adult-onset acquired immunodeficiency featuring cryptococcosis, primarily caused by *C. gattii*, also develops in patients with autoantibodies against GM-CSF; these autoantibody-associated conditions may respond to anti-CD20- or anti-CD38-targeted therapies. Last, cryptococcosis develops in patients with autosomal-dominant haploinsufficiency of the transcription factor GATA2, which manifests myelodysplasia, lymphedema, and broad infection susceptibility.[4]

Cryptococcosis is also observed in patients who receive biological agents that deplete T lymphocytes (i.e., alemtuzumab), inhibit Janus kinase (JAK) signaling—which abrogates intramacrophagic fungal control—or inhibit Bruton tyrosine kinase (BTK)—which impairs B-lymphocyte and macrophage responses—in agreement with reports of cryptococcosis in patients with inherited BTK deficiency (i.e., X-linked agammaglobulinemia [XLA]).

Immune Responses to Molds

Molds primarily cause invasive disease in the human respiratory tract (see Table 28.1), particularly in immunocompromised individuals. The major etiological agents of life-threatening mold pneumonia include *Aspergillus*, *Scedosporium*, *Fusarium* spp., and the *Mucorales* order, all of which form branching tubular filaments, termed hyphae. Tissue-invasive hyphae are highly destructive in human tissues and have the capacity to disseminate to extrapulmonary sites. Beyond the respiratory tree, molds are associated with severe skin and soft-tissue infections when fungal cells are introduced by traumatic inoculation (e.g. in victims of natural disasters). *Aspergillus* spp. are further linked to allergenic disease in patients with underlying atopy and asthma and to chronic cavitary disease with varying degrees of hyphal tissue invasion in patients with cystic fibrosis or structural lung disease.[19]

Molds are saprophytic organisms that thrive on decaying organic matter in the environment and form a mycelium, a collection of branching, interconnected hyphae. Coniodiophores are specialized hyphal cells that are located at the mycelial interface with air and generate conidia (i.e., vegetative spores), the infectious propagules. Humans inhale airborne mold conidia on a daily basis and exposure to these pathogens is ubiquitous. The respiratory innate immune system rapidly inactivates conidia that bypass mucociliary clearance in individuals without underlying injury to the immune system and to the lung architecture.

Worldwide, *Aspergillus* and agents of mucormycosis cause an estimated 200,000 and 10,000 life-threatening invasive infections annually, respectively. Vulnerable patient groups include individuals with disease- or therapy-related injury to the myeloid cell compartment of the immune system (e.g., patients with leukemia, lymphoma, hematopoietic stem cell, and lung transplantation), and individuals on long-term immune suppression for autoimmune diseases. More recently, critically ill patients with severe influenza or COVID-19 and patients treated with small-molecule drugs (i.e., ibrutinib) that target host signaling molecules involved in fungal immune surveillance have been identified as vulnerable to invasive mold infections, in particular invasive aspergillosis. Despite the widespread availability of two major classes of mold-active drugs, the polyenes and mold-active triazoles, fungus-attributable mortality rates for aspergillosis, fusariosis, and mucormycosis range from 25% to 90%, with the highest mortality observed in patients with extrapulmonary dissemination in the absence of myeloid cell recovery or reconstitution.[19]

Aspergillosis

There are more than 200 *Aspergillus* species, though only a handful cause human disease. *A. fumigatus* is responsible for approximately two-thirds of invasive infections, with *A. flavus*, *A. niger*, *A. terreus*, *A. ustus, and A. nidulans* observed in a minority of cases. The ensuing discussion will focus primarily on immune responses to *A. fumigatus*, since this species has been characterized and studied in-depth compared to the non-*fumigatus Aspergillus* spp.

The innate immune system plays a dominant role in defense against all filamentous molds—and against *Aspergillus* in particular. This concept arose out of clinical observations in the 1970s and 1980s in which the susceptibility of patients with acute leukemia to aspergillosis correlated with the duration of neutropenia. In addition, qualitative neutrophil defects predispose patients to aspergillosis, exemplified by a ~40% to 50% lifetime

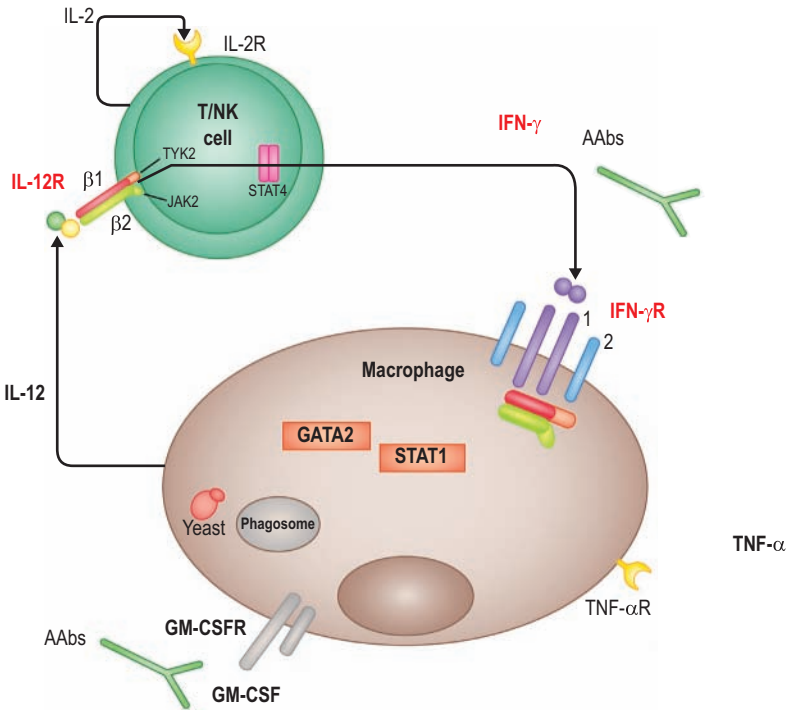


FIG. 28.2 An orchestrated lymphocyte-Macrophage Cross-Talk promotes Sterilizing Immunity Against Intramacrophagic fungi. The interplay between T lymphocytes and NK cells with monocytes/macrophages drives clearance of intramacrophagic fungi (*Cryptococcus*, endemic dimorphic fungi). IL-12 is released by monocytes/macrophages in response to fungal engulfment and binds to its receptor, which consists of IL-12Rβ1 and IL-12Rβ2 subunits on the surface of T lymphocytes and NK cells. IL-12 binding on the IL-12 receptor activates STAT4 via TYK2 and JAK2 and leads to the transcription and secretion of IFN-γ. IFN-γ then binds to its receptor, which consists of IFN-γR1 and IFN-γR2 subunits on the surface of monocytes/macrophages. IFN-γ binding on the IFN-γ receptor activates STAT1 via JAK1 and JAK2, which enables intracellular fungal killing via effector mechanisms that are not well-characterized. GATA2 is critical for monocyte, DC, and NK-cell development and effector function, and human autosomal-dominant GATA2-haploinsufficiency leads to infection susceptibility by intramacrophagic fungi. Mutations in the β1 subunit of the IL-12 receptor (*IL12RB1*), in either subunit of the IFN-γ receptor, and gain-of-function *STAT1* mutations also lead to infection susceptibility by intramacrophagic fungi (and mycobacteria). TNF and GM-CSF bind to their receptors on the surface of monocytes/macrophages, resulting in monocyte/macrophage activation. Pharmacological inhibition of TNF and autoantibodies against IFN-γ or GM-CSF are associated with infection susceptibility by intramacrophagic fungi. Inherited mutations in genes encoding the proteins in IFN-γR and IL-12R (*red font*) have been described to feature infections by intramacrophagic fungi in humans. The molecular mechanisms by which GATA2, STAT1, and TNF promote intramacrophagic fungal killing remain poorly understood. IL-2 promotes T-lymphocyte proliferation and activation by binding to its IL-2 receptor. Abbreviations: AAbs, autoantibodies; DC, dendritic cell; GATA2, GATA binding protein 2; GM-CSF, granulocyte–macrophage colony-stimulating factor; JAK, Janus kinase; NK, natural killer; STAT, signal transducer and activator of transcription; TYK2, tyrosine kinase 2. (Modified from Lionakis MS, Levitz SM. Host control of fungal infections: lessons from basic studies and human cohorts. *Annu Rev Immunol*, 2018;36:157–191.)

incidence of aspergillosis in patients with CGD, the hallmark of which is the loss of the oxidative burst, primarily in neutrophils. In contrast to the central role of myeloid cells, in particular neutrophils, in host defense against molds, humans with defects in lymphoid cell number or function, including CD4 and CD8 T, NK, and B cells, are typically not prone to these infections. These observations have been recapitulated in mouse models of mold infections, highlighting conserved principles of host defense across mammalian species.[3]

The central role of the innate immune response is to prevent the germination of inhaled mold conidia. Unchecked, inhaled mold conidia swell, form a germ tube, and extend hyphae that can pierce phagocytes and cross tissue barriers. To prevent this process, tissue-resident alveolar macrophages and DCs engulf conidia, restrict germination within the phagolysosome, and initiate the recruitment of neutrophils, circulating monocytes, and plasmacytoid DCs (pDCs) in the fungus-infected lung (Fig. 28.3). These recruited cell subsets cooperate to condition the lung inflammatory milieu to promote fungal cell killing, primarily via the action of NADPH oxidase.

Innate immune recognition of germinating conidia relies on the stage-specific exposure of fungal polysaccharides that trigger host CLR signaling pathways. In the case of molds, this process is best understood for *A. fumigatus* in which conidial swelling leads to the surface exposure of β-(1, 3) glucan polysaccharides that in turn activates the Dectin-1, Syk, and CARD9 signal transduction pathway, culminating in NF-κB nuclear translocation and in induction of proinflammatory cytokines—including IL-1β and mediators of neutrophil chemotaxis (e.g., CXCL1, CXCL2). *Aspergillus* spp. also express polysaccharide ligands that activate Dectin-2-mediated signaling via FcRγ, Syk, and CARD9. Unlike hyphae, *Aspergillus* conidia contain a pigment, termed melanin, that can activate the CLR MelLec (*clec1a*) though the downstream signaling events have not yet been elucidated. While the

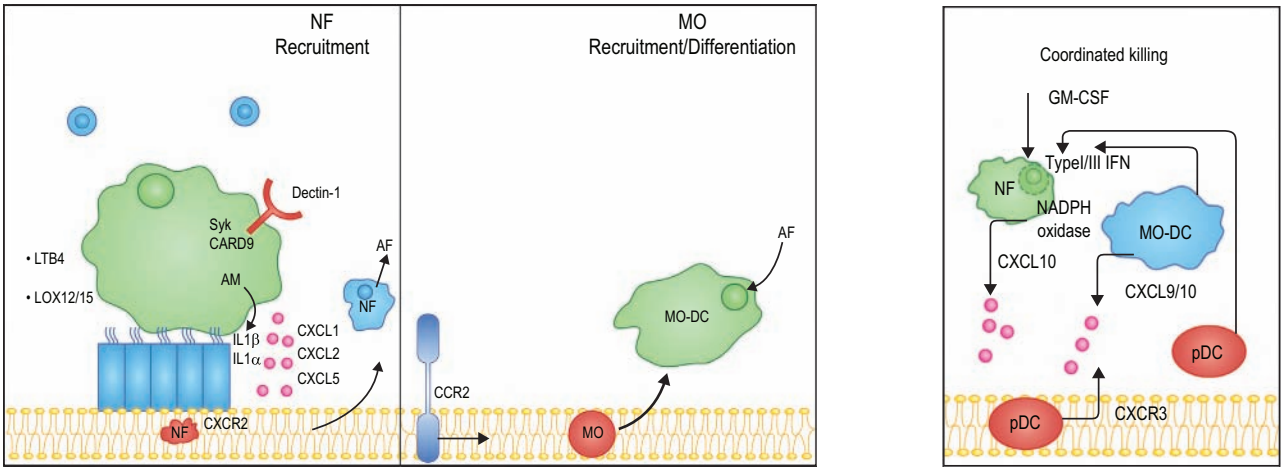


FIG. 28.3 Induction of the Innate Immune Response to Pulmonary Infection with *Aspergillus fumigatus*. *Aspergillus fumigatus* conidia are rapidly engulfed by lung-resident alveolar macrophages and trigger the induction of Dectin-1/Syk/CARD9 and eicosanoid signaling that result in the release of neutrophil- and monocyte-targeted chemoattractants (*left panel*). Neutrophils and monocytes enter the circulation and traffic into the *Aspergillus*-infected lung via CXCR2- and CCR2-mediated signals, respectively (*left and middle panels*). In the lung, fungus-engaged neutrophils and monocyte-derived dendritic cells kill conidia directly and condition the inflammatory milieu, in part through the release of type I interferons and CXCL9 and CXCL10 (*right panel*). This triggers the recruitment of $CXCR3^+$ plasmacytoid dendritic cells that further enhance the killing capacity of neutrophils by increasing the oxidative burst (right panel). These processes trigger a regulated cell death program in *Aspergillus* conidia, resulting in sterilizing immunity in the respiratory tree. *NF*, neutrophil; *MO*, monocyte; *AF*, *Aspergillus fumigatus*; *AM*, alveolar macrophages; *pDC*, plasmacytoid dendritic cells.

CLR-Syk-CARD9 signaling pathway operates in hematopoietic cells to mediate the recruitment of myeloid effector cells, *Aspergillus* infection also elicits the rapid release of IL-1α, which together with IL-1β, can activate IL-1 receptor signaling on lung stromal cells; this pathway provides an alternate source of neutrophil-targeted chemoattractants. Beyond CXCL1, CXCL2, and CXCL5, leukotriene B4 is another potent neutrophil-recruiting mediator released early in *Aspergillus* infection.[20]

As mentioned above, the primary neutrophil killing mechanism involves products of NADPH oxidase that act on conidia and on hyphae. Lung-infiltrating monocytes differentiate into Mo-DCs in the lung and contribute to direct conidial killing as well. Infected neutrophils and Mo-DCs release CXCL9 and CXCL10, both of which serve to mediate the entry of $CXCR3^+$ pDCs into the lung. Although pDCs do not appear to kill fungal cells directly, pDCs enhance the killing activity of neutrophils by boosting the oxidative burst.[21] Several additional mediators play key roles in clearing *Aspergillus* conidia and in mediating sterilizing immunity. Chief among these are type I and type III interferons, produced within hours post-infection, that, together with GM-CSF, condition the lung inflammatory environment to increase oxidative killing mechanisms.[22] Neutrophil NADPH oxidase activity kills fungal cells by inducing a regulated cell death process within conidia.

Nonoxidative killing mechanisms also contribute to sterilizing immunity. This concept has been detailed in a mouse model of *Aspergillus* keratitis, a devastating sight-threatening disease found primarily in agricultural workers in resource-limited countries. In the eye, fungal hyphae form rapidly because the resident ocular innate immune system is insufficient to control early germination, unlike the situation in the respiratory tree. When recruited neutrophils encounter *Aspergillus* hyphae, they can undergo a form of regulated cell death, termed NETosis, a process that requires NADPH oxidase activity and results in the extrusion of actin and chromatin fibers, histones, and the proteins calprotectin (S100A8/S100A9) and pentraxin-3. Calprotectin is an abundant protein in the neutrophil cytosol that acts to sequester the divalent cations Zn^{2+} and Mn^{2+}. Mice that lack calprotectin are susceptible to ocular, but not respiratory, challenge with *A. fumigatus* conidia. Host iron sequestration—for example via lactoferrin—contributes to nutritional immunity, the process by which host cells prevent *Aspergillus* nutrient acquisition. The precise role of NETosis in *Aspergillus* killing remains unclear because genetic tools that interfere with formation of NETs without affecting other antimicrobial functions are lacking. In macrophages, recent studies indicate that *Aspergillus* uptake induces LC3-associated phagocytosis (LAP), a process in which components of the autophagy machinery conjugate LC3 to fungus-containing phagosomes. The melanin pigment found in conidia can block this process. Loss of the conidial pigment during germination enhances this noncanonical form of autophagy to promote fungal killing, partly by metabolic reprogramming of macrophages.[3]

Although lymphoid cells are not essential for protection against *Aspergillus* and other filamentous molds, fungus-specific CD4 T cells, specifically Th2 and Th17 cells, contribute to *Aspergillus*-associated allergenic disease. Recent work demonstrated that CD4 T cells primed in the intestine to respond to commensal *C. albicans* can migrate to the lung and cross-react with antigens found in inhaled *Aspergillus* and other molds to exacerbate fungal asthma. These cross-reactive CD4 T cells produce IL-17, which is protective against mucosal fungal infections, but aggravates allergenic immune pathology in the lung. Moreover, *Aspergillus* encodes a number of proteases (e.g., Alp1p) that act as allergens by causing tissue damage to lung epithelial cells; this injury can activate mechanosensitive channels that promote allergen sensitization via the calcineurin signaling pathway. In another model of fungal asthma, *Aspergillus* protease antigens promote allergenic disease by cleaving the clotting protein fibrinogen; in this setting, fibrinogen cleavage products promote allergic inflammation by activating TLR4 signaling[23] (Chapter 43).

Agents of Mucormycosis

The order *Mucorales* contains the agents of mucormycosis, the most common of which are *Mucor*, *Rhizopus*, *Rhizomucor*, *Absidia*, *Cunninghamella, Saksenaea*, and *Lichtheimia* species. This class of organisms forms branching, ribbon-like aseptate hyphae in human tissues. Similar to aspergillosis, the innate immune system, and neutrophils in particular, plays a dominant role in protection against mucormycosis, with no essential requirement for adaptive immune responses in host defense, based on mouse models of disease and observations in humans with PIDs. Nutritional immunity contributes to alveolar macrophage-mediated defense against inhaled *Mucorales* conidia.

A unique characteristic of mucormycosis is the occurrence of rhinocerebral disease in patients with uncontrolled diabetic ketoacidosis. Recent laboratory studies indicate that *Rhizopus* can exploit the upregulation of glucose-regulated protein (GRP78) expressed in nasal epithelial and in endothelial cells during hyperglycemia. The *Rhizopus* spore coat protein CotH3 binds GRP78, resulting in GRP78-mediated fungal cell internalization and tissue invasion. Additional *Mucorales* CotH family members contribute to fungal pathogenesis, including CotH7 that binds to β_1 integrin. CotH family members are widely expressed among the *Mucorales* order and absent from non-*Mucorales* fungal pathogens. This finding may provide a potential explanation for the susceptibility of patients with diabetic ketoacidosis to mucormycosis. In mouse models, antibodies directed against GRP78 and CotH3 ameliorate the outcome of mucormycosis.[24]

PATIENTS AT RISK FOR INVASIVE MOLD INFECTIONS DUE TO GENETIC OR PHARMACOLOGICAL FACTORS

The link between a PID and mold infections is best established for CGD in which defects in the CYBB, CYBA, NCF1, NCF2, or NCF4 subunits of phagocyte oxidase prevent the assembly and function of the complex. Invasive aspergillosis is a hallmark infection in CGD, particularly if cultures reveal the presence of *Aspergillus nidulans* (see Table 28.3); the exception is CGD caused by NCF4 deficiency that features predominantly colitis without profound mold susceptibility. Mucormycosis is uncommon in CGD and is typically observed when patients are treated with corticosteroids or other immunosuppressants for the inflammatory sequelae of this disease. CARD9 deficiency is associated with extrapulmonary aspergillosis and cutaneous mucormycosis, likely due to the strict CARD9 dependency of neutrophil recruitment to sites of mold infections outside of the respiratory tree.[4,6]

Aspergillosis occurs in severe congenital neutropenia, caused by *ELA2* and *HAX1* mutations, and in leukocyte adhesion deficiency, caused by mutations in β_2 integrins (*CD18*, *CD11a, -b*, or *-c*) that impairs neutrophil trafficking into infected tissues. Mucormycosis has been reported in patients with Papillon-Lefèvre syndrome in which the activation of neutrophil granule serine proteases, specifically cathepsin C, is defective. Patients with GATA2 haploinsufficiency develop monocytopenia and NK-cell lymphopenia and can develop aspergillosis, among other fungal and viral infections. In Job's syndrome, termed autosomal-dominant hyper-IgE syndrome, mutations in *STAT3* lead to primary staphylococcal lung infections and ensuing secondary *Aspergillus* infections at sites of cavitary lung lesions. Autosomal-dominant mutations in *STAT1*, a transducer of interferon responses, are associated with aspergillosis, mucormycosis, and cutaneous fusariosis. A common theme to PIDs that give rise to invasive mold infections is that all result in qualitative or quantitative defects in myeloid cells.[7]

Patients treated with the BTK inhibitor ibrutinib can develop invasive aspergillosis with a predilection for CNS disease, particularly when ibrutinib was combined with other therapies that target lymphoma. However, XLA humans that carry loss-of-function *BTK* mutations were previously not observed to be susceptible to invasive mold infections. Recent laboratory studies indicate that murine BTK-deficiency enhances *A. fumigatus* susceptibility, likely due a requirement for BTK signaling in myeloid cells to achieve optimal anti-*Aspergillus* effector activity.

In allogeneic hematopoietic stem cell transplantation recipients, SNPs in membrane-bound and soluble receptors (i.e., *CLEC7A*, *CLEC1A*, *PTX3*, *TLR4, TLR6, CD209, NOD2, PLG, TNFR1*), ion chelators implicated in nutritional immunity (*S100B*), and cytokines (*IFNG, CXCL10*) have been linked to heightened risk of transplant-associated aspergillosis.[25] However, genetic defects in these molecules are not associated with mold infections in otherwise immunocompetent individuals, highlighting molecular redundancy in innate recognition of mold conidia and hyphae.

IMMUNE RESPONSES TO DIMORPHIC FUNGI

In North America, the endemic dimorphic fungi represent a significant cause of respiratory and systemic diseases (see Table 28.1). The etiological agents include *Blastomyces dermatitidis, B. gilchristii* and related species, *Histoplasma capsulatum, Coccidioides immitis*, and *C. posadasii*. These fungi are geographically distributed in the United States; *Blastomyces* and *Histoplasma* are found in the Midwestern, Southcentral, and Southeastern regions, and *Coccidioides* in the Southwest. Collectively, these pathogens produce greater than 1 million new infections annually. The fungi grow as mold in the soil and, after spores are inhaled, they convert to pathogenic yeast or spherules (*Coccidioides*). The dimorphic fungi cause infection in healthy persons, but disease may be intensified in immunocompromised hosts. Infected persons usually have no prior underlying immune impairment; those that do may manifest disease reactivation and dissemination. A variable percentage of previously healthy persons exhibit progressive pulmonary or extrapulmonary disease. In histoplasmosis, the frequency is ~10% in those with clinically apparent infection; in coccidioidomycosis, it is ~30%. In blastomycosis, the rates of progressive disease may be even higher. A hallmark of these disorders in healthy hosts is the development of lasting immunity after primary infection.

Host defense requires the development of an adaptive immune response chiefly involving antigen-specific CD4 T cells. These cells promote immunity by releasing cytokines and providing help to B cells and CD8 T cells. Loss of these functions underpins susceptibility of AIDS patients to these infections. In mice and humans, Th1- and Th17-biased responses confer protection. Th2 cells exacerbate infection by dampening Th1 responses and inducing alternatively activated macrophages. Regulatory T cells fine tune the immune response and limit collateral damage but can also promote fungal persistence.[15,26]

Histoplasmosis

There are an estimated 500,000 new *Histoplasma* infections annually in the United States. While the vast majority of infections

are asymptomatic, 10% become progressive. Infection especially poses a threat to solid organ transplant recipients, AIDS patients, and other immunocompromised hosts. Upon inhalation, *Histoplasma* spores settle into bronchioles and alveoli and convert into pathogenic yeasts within resident phagocytes and DCs. These cells transport *Histoplasma* to visceral and lymphoid organs where the infection expands. Within weeks, Th1 immunity is activated, arresting *Histoplasma* replication in monocytes and macrophages, although tissues may not be sterilized. Conversely, heightened Th2 responses exacerbate disease by producing type-2 cytokines that dampen the activity of macrophages. DCs present *Histoplasma* antigen to CD4 T cells and dictate their differentiation into Th1 cells. DCs initiate Th1 differentiation by secreting IL-12; they foster Th2 polarization by releasing soluble mediators or by contact with surface molecules such as Jagged, CD80/86, and OX40L. DCs also modulate Th1 or Th2 responses by releasing chemoattractants, including CXCL10 (Th1 cell migration), CCL17, and CCL22 (Th2 cells).[26,27]

Intracellular residence of *Histoplasma* poses a stress to the fungus as it transitions from soil, which contains more zinc than macrophages. GM-CSF, a Th1 cytokine, activates the anti-*Histoplasma* activity of macrophages, stimulates movement of zinc into the cytosol of macrophages, and triggers production of zinc-binding proteins. Metallothionein 1 and 2 sequester zinc, reduce free metal in host cells and starve the yeast. Sequestration of zinc by metallothionein also increases reactive oxygen species. Enhanced reactive oxygen species and zinc sequestration from yeast enhance fungal killing. Conversely, IL-4 production induces zinc accumulation by *H. capsulatum*, boosting its intracellular survival.[28]

Coccidioidomycosis

There are an estimated 25,000 cases of coccidioidomycosis that require medical treatment annually. Although a minority of those exposed to *Coccidioides* develop acute illness, up to 29% of community-acquired pneumonia may be due to *Coccidioides* in endemic areas. Chronic infection is common, with many lung nodules attributable to *Coccidioides* in endemic areas. Disseminated coccidioidomycosis is rare, appearing in less than 1% of infections, but is a devastating and often fatal complication.

In murine models, PRRs and signaling adaptors foster innate control of *Coccidioides* infection. MyD88 and CARD9 adaptors, Dectin-1, and IL-1R1 orchestrate the recognition of spherule wall components. CD4 T helper cells and CD8 cytotoxic cells contribute to adaptive immunity. Th1 and Th17 responses confer protection, while Th2 and IL-10-producing cells may foster pathological responses.[29]

Live attenuated vaccines confer protection in experimental animals. These include a strain with two disrupted chitinase genes (CTS2 and CTS3; *cts2*/*ard1*/*cts3*Δ mutant) and another with a *CPS1* gene deletion (Δ*CPS1* mutant). These strains are highly attenuated and well tolerated even as primary pulmonary infection, including in NSG mice that lack cellular and humoral immunity (Δ*CPS1* strain). The vaccines underscore a role for IFN-γ as a correlate of protection. The Δ*CPS1* vaccine is under study in experimental dogs. Despite the capacity of live, attenuated vaccines to elicit protective immunity in mice, they may not be safe for individuals with compromised immune systems.

A dozen *Coccidioides* antigens have been identified and characterized as vaccine candidates. Multivalent vaccines that contain two or more antigens are more potent against experimental pulmonary *Coccidioides* infection than vaccines containing a single antigen. A *Coccidioides* polypeptide vaccine consists of three antigens (cell wall antigen 2 [Ag2/Pra], *Coccidioides*-specific antigen [Cs-Ag], proximal matrix protein 1 [Pmp1]), and five pathogen-derived peptides with affinity for human major histocompatibility complex (MHC) class II molecules. When formulated in glucan particles as an adjuvant the vaccine protects both C57BL/6 and HLA-DR4 "humanized" mice. Th1 and Th17 cells participate in mediating this *Coccidioides* vaccine effect.[4]

Blastomycosis

The majority of blastomycosis cases present as fungal pneumonia. Although blastomycosis is comparatively uncommon, with perhaps ~3000 cases reported annually in highly endemic areas, the incidence may reach 40 cases per year per 100,000 population. Blastomycosis is more common in people of African American, Native American, and Asian, especially Hmong, ancestry. Some of these associations may be due to a combination of genetic and nongenetic factors.

Once in the yeast phase, *B. dermatitidis* and *B. gilchristii* upregulate *Blastomyces*-adhesin-1 (BAD1), an essential virulence factor that is specific to the yeast phase. BAD1 is an adhesin and immune evasin. BAD1 binds the yeast to host tissue and immune cells via interactions with heparin sulfate and CR3 and CD14. Surface-bound and soluble BAD1 inhibit TNF production by macrophages and neutrophils. BAD1 also inhibits CD4 T-lymphocyte activation via interaction with heparan sulfate modifications on CD47, resulting in reduced IL-17 and IFN-γ production. Other immune evasive strategies of the yeast include cleavage of chemokines that recruit $CCR2^+$ inflammatory monocytes to the lung and cleavage of type-1 cytokines such as GM-CSF by fungal dipeptidyl peptidase-4 (Dpp4).[30,31]

Although *Blastomyces* yeast subvert and evade the immune system, host macrophages and neutrophils kill a large portion of conidia inhaled into the lung. CD4 T cells coordinate adaptive immunity by activating macrophages that enhance yeast killing. This requires Th1 cytokines such as TNF and IFN-γ and production of IL-17 by Th17 cells. Cell-mediated immunity following recovery from blastomycosis can last for at least 2 years.

Deletion of BAD1 severely attenuates the fungus. *BAD1*Δ yeast can vaccinate mice against experimental infection. Upon subcutaneous injection, *BAD1*Δ yeasts are transported by inflammatory monocytes to the draining lymph nodes where resident DCs present antigen to T cells. Naïve CD4 T lymphocytes differentiate into Th1 and Th17 cells in a manner that requires Dectin-2/FcRγ/Syk/Card9, Dectin-3, and mannan receptor signaling pathways. Once differentiated, Th1 and Th17 cells migrate to the lungs in response to infection to mediate vaccine immunity by secreting IFN-γ and IL-17 to recruit and activate neutrophils and macrophages. In the absence of CD4 T cells, *BAD1*Δ vaccinated mice use IFN-γ and IL-17 CD8 T lymphocytes (Tc1 and Tc17 cells) to mediate protective immunity. Proliferation and activation of IL-17-producing CD8 T cells occurs through MyD88-Akt1-mTOR signaling. *BAD1*Δ vaccine antigens that induce protective CD4 T cells include calnexin and endoglucanase 2, which are conserved among ascomycete fungi, including *H. capsulatum, C. posadasii, A. fumigatus, Fonsecaea pedrosoi* (the agent of chromoblastomycosis), *and Pseudogymnoascus destructans* (the fungus responsible for white-nose syndrome in bats). Calnexin- and endoglucanase-based vaccines protect against pneumonia caused by the three major endemic dimorphic fungi.[30,31]

In addition, in a mouse model of blastomycosis, lung epithelial cells promote innate immunity in a manner that requires NF-κB. Signaling through IL-1R1 and MyD88 drives IL-17–producing innate lymphocytes that engage additional myeloid cells.

PATIENTS AT RISK FOR ENDEMIC MYCOSES DUE TO GENETIC OR PHARMACOLOGICAL FACTORS

Disseminated histoplasmosis is most commonly associated with the immunocompromised patient; dissemination is rare in the immunocompetent host. Common sites of dissemination include the CNS and gastrointestinal tract. Patients receiving TNF inhibitors are at risk of disseminated infection as are those receiving the anti-IFN-γ monoclonal antibody emapalumab. Progression to disseminated coccidioidomycosis is likewise rare, occurring in less than 1% of infections, but immunosuppression, including use of TNF inhibitors, increases risk of dissemination.

Many of the same genetic deficits in immunity predispose patients to disseminated histoplasmosis or coccidioidomycosis. These include loss-of-function mutations impacting IFN-γR1, IL-12Rβ1, and STAT3 and gain-of-function mutations in STAT1 (see Table 28.3). Based on a small number of reported cases of coccidioidomycosis, the fungus may disseminate preferentially to distinct sites depending on which component of immunity is impaired. For example, *STAT3* mutations are associated with dissemination to the CNS and gastrointestinal tract, while mutations that directly impact type-1 responses are associated with dissemination to bone and lymph node.[7]

Race and ethnicity influence the rates and patterns of endemic mycoses. African ancestry has been associated with increased risk of severe histoplasmosis in AIDS patients as well as increased overall risk of infection. The genetic determinants of racial disparities in *Histoplasma* pneumonia or disseminated histoplasmosis remain unknown.

The risk of disseminated coccidioidomycosis is unequal across people of different ethnicities and genetic backgrounds. Native American, African, and Pacific Islander ancestry all confer increased risk of dissemination. A few reports have addressed population-level patterns of susceptibility to disseminated coccidioidomycosis. Some of the risk may be attributable to differences in antigen presentation, as certain HLA alleles are associated with dissemination in specific populations. A preliminary genomic study of 58 patients with disseminated coccidioidomycosis identified 103 rare variants in 21 genes associated with antifungal immunity. Affected pathways included IL-17 signaling, IL-12/IFN-γ signaling, and NF-κB signaling. Two-thirds of patients had functionally relevant variants with a population frequency of less than 0.1%, including several patients with biallelic deleterious variants in key immune-related genes.

Recent work has addressed the genetic underpinning of disparate rates of blastomycosis in healthy individuals. A homozygosity mapping approach was used to study nine Hmong blastomycosis patient genomes, and candidate susceptibility variants were identified on the basis of their rarity in European populations and other features that may indicate that a given variant influences susceptibility. A block of variants near IL-6 is nearly fixed in Wisconsin Hmong but rare in European populations. Healthy Wisconsin Hmong donors had hypoactive IL-6 and antifungal Th17 responses compared to healthy European donors.[32] The latter finding is consistent with differences in IL-6 responses, which could also explain several as-yet unstudied population differences that impact T-cell development. To date, the only reported monogenic condition associated with blastomycosis is GATA2 haploinsufficiency.

SUMMARY

Fungi have emerged as significant causes of mortality in the rapidly expanding populations of immunosuppressed patients who are at risk for opportunistic fungal disease. A recent remarkable surge in immunological discoveries stemming from basic studies in animal models of fungal infections and observations in human cohorts with inherited or acquired susceptibility to mucosal and invasive mycoses has significantly advanced our understanding of the cellular and molecular factors that promote protective immune responses during fungal exposures of humans. This knowledge shows promise for devising precision-medicine strategies for risk assessment, vaccination, prophylaxis, immunotherapy, and prognostication of fungus-infected patients.

ON THE HORIZON

- The identification of CARD9 deficiency as, to date, the only known human monogenic disorder that features fungal-specific infection susceptibility underscores the critical contribution of C-type lectin receptor (CLR) signaling in human antifungal host defense and paves the way for the discovery of other novel inborn errors of immunity in CLR-related pathway genes that may underlie inherited susceptibility to tissue-specific fungal disease.
- In patients with iatrogenic immunosuppression or critical illness who develop life-threatening invasive fungal infections, the identification of dysfunctional single nucleotide polymorphisms in immune-related genes that heighten fungal infection susceptibility may lead to the development of individualized strategies for risk assessment and antifungal prophylaxis.
- A deeper understanding of the cellular and molecular immune factors that protect against fungal disease may inform immunotherapy and vaccination strategies in vulnerable patients.
- The advent of precision-medicine biological therapies that target key fungal surveillance immune pathways for the treatment of autoimmune and neoplastic disorders has resulted in new patient populations at risk for opportunistic fungal disease, and understanding the epidemiology and immunology of such infections can help devising targeted antifungal prophylaxis strategies for at-risk individuals.

ACKNOWLEDGMENTS

T.M.H. is supported by a Burroughs Wellcome Fund Investigator in the Pathogenesis of Infectious Diseases Award and by NIH grants P30 CA 008748 (to MSKCC), R01 AI093808, R01 AI139632, and R21 AI142639. B.S.K. is supported by NIH grants R01 AI130411, R01 AI035681, R01 AI040996, R01 AI093553, U19 AI142720, and T32 AI055397. This work was support by the Division of Intramural Research of the National Institute of Allergy and Infectious Diseases, National Institutes of Health (M.S.L.).

REFERENCES

1. Brown GD, Denning DW, Gow NA, et al. Hidden killers: human fungal infections. *Sci Transl Med.* 2012;4(165):165rv13.
2. Brown GD, Willment JA, Whitehead L. C-type lectins in immunity and homeostasis. *Nat Rev Immunol.* 2018;18(6):374–389.

3. Lionakis MS, Iliev ID, Hohl TM. Immunity against fungi. *JCI Insight.* 2017;2:(11).
4. Lionakis MS, Levitz SM. Host control of fungal infections: lessons from basic studies and human cohorts. *Annu Rev Immunol*. 2018;36:157–191.
5. Gow NA, van de Veerdonk FL, Brown AJ, et al. *Candida albicans* morphogenesis and host defence: discriminating invasion from colonization. *Nat Rev Microbiol*. 2011;10(2):112–122.
6. Corvilain E, Casanova JL, Puel A. Inherited CARD9 deficiency: invasive disease caused by ascomycete fungi in previously healthy children and adults. *J Clin Immunol*. 2018;38(6):656–693.
7. Lionakis MS, Netea MG, Holland SM. Mendelian genetics of human susceptibility to fungal infection. *Cold Spring Harb Perspect Med*. 2014;4:(6).
8. Pappas PG, Lionakis MS, Arendrup MC, et al. Invasive candidiasis. *Nat Rev Dis Primers*. 2018;4:18026.
9. Gaffen SL, Moutsopoulos NM. Regulation of host-microbe interactions at oral mucosal barriers by type 17 immunity. *Sci Immunol.* 2020;5:(43).
10. Fan D, Coughlin LA, Neubauer MM, et al. Activation of HIF-1alpha and LL-37 by commensal bacteria inhibits *Candida albicans* colonization. *Nat Med.* 2015;21(7):808–814.
11. Netea MG, Joosten LA, van der Meer JW, et al. Immune defence against *Candida* fungal infections. *Nat Rev Immunol*. 2015;15(10):630–642.
12. Drummond RA, Swamydas M, Oikonomou V, et al. CARD9(+) microglia promote antifungal immunity via IL-1beta- and CXCL1-mediated neutrophil recruitment. *Nat Immunol*. 2019;20(5):559–570.
13. Campuzano A, Wormley FL. Innate immunity against *Cryptococcus*, from recognition to elimination. *J Fungi (Basel)*. 2018;4:(1).
14. Mukaremera L, Nielsen K. Adaptive immunity to *Cryptococcus neoformans* infections. *J Fungi (Basel)*. 2017;3:(4).
15. Wuthrich M, Deepe Jr GS, Klein B. Adaptive immunity to fungi. *Annu Rev Immunol*. 2012;30:115–148.
16. Li J, Casanova JL, Puel A. Mucocutaneous IL-17 immunity in mice and humans: host defense vs. excessive inflammation. *Mucosal Immunol.* 2018;11(3):581–589.
17. Notarangelo LD, Bacchetta R, Casanova JL, et al. Human inborn errors of immunity: an expanding universe. *Sci Immunol.* 2020;5:(49).
18. Brown GD. Innate antifungal immunity: the key role of phagocytes. *Annu Rev Immunol*. 2011;29:1–21.
19. Latge JP, Chamilos G. *Aspergillus fumigatus* and aspergillosis in 2019. *Clin Microbiol Rev*. 2019;33:(1).
20. Tischler BY, Hohl TM. Menacing mold: recent advances in *Aspergillus* pathogenesis and host defense. *J Mol Biol*. 2019;431(21):4229–4246.
21. Guo Y, Kasahara S, Jhingran A, et al. During *Aspergillus* infection, monocyte-derived DCs, neutrophils, and plasmacytoid DCs enhance innate immune defense through CXCR3-dependent crosstalk. *Cell Host Microbe*. 2020;28(1):104–116.e4.
22. Espinosa V, Dutta O, McElrath C, et al. Type III interferon is a critical regulator of innate antifungal immunity. *Sci Immunol.* 2017;2:(16).
23. Tung HY, Li E, Landers C, et al. Advances and evolving concepts in allergic asthma. *Semin Respir Crit Care Med*. 2018;39(1):64–81.
24. Baldin C, Ibrahim AS. Molecular mechanisms of mucormycosis—the bitter and the sweet. *PLoS Pathog*. 2017;13(8):e1006408.
25. Fisher CE, Hohl TM, Fan W, et al. Validation of single nucleotide polymorphisms in invasive aspergillosis following hematopoietic cell transplantation. *Blood.* 2017;129(19):2693–2701.
26. Verma A, Wuthrich M, Deepe G, et al. Adaptive immunity to fungi. *Cold Spring Harb Perspect Med*. 2014;5(3):a019612.
27. Kauffman CA. Histoplasmosis: a clinical and laboratory update. *Clin Microbiol Rev*. 2007;20(1):115–132.
28. Subramanian Vignesh K, Landero Figueroa JA, Porollo A, et al. Granulocyte macrophage-colony stimulating factor induced Zn sequestration enhances macrophage superoxide and limits intracellular pathogen survival. *Immunity.* 2013;39(4):697–710.
29. Hung CY, Hsu AP, Holland SM, et al. A review of innate and adaptive immunity to coccidioidomycosis. *Med Mycol*. 2019;57(suppl_1):S85–S92.
30. McDermott AJ, Klein BS. Helper T-cell responses and pulmonary fungal infections. *Immunology.* 2018;155(2):155–163.
31. Wuthrich M, Ersland K, Sullivan T, et al. Fungi subvert vaccine T cell priming at the respiratory mucosa by preventing chemokine-induced influx of inflammatory monocytes. *Immunity.* 2012;36(4):680–692.
32. Merkhofer Jr RM, O'Neill MB, Xiong D, et al. Investigation of genetic susceptibility to blastomycosis reveals interleukin-6 as a potential susceptibility locus. *MBio.* 2019;10:(3).

Host Defenses to Protozoa

Robin Stephens, Jude E. Uzonna, and Sara M. Dann

Protozoal infections are an important cause of morbidity and mortality worldwide (Table 29.1). Protozoan pathogens exact their major toll in the tropics, but infection by these parasites remains a significant problem everywhere because of travel, military action, and immigration, the susceptibility of patients with acquired immunodeficiency syndrome (AIDS) to opportunistic protozoans, and episodic transmission within communities.

KEY CONCEPTS

Host Defense Against Protozoa

- Interaction of the parasite with host cells induces various cytokines that stimulate the innate and adaptive immune responses to eliminate the pathogen. Immunoregulatory cytokines can inhibit or downregulate both antiparasitic and host-damaging responses to enable the persistence of tissue parasitism.
- The outcome of infection is determined by the balance between the infection-promoting and the host-protective cytokines and effector cells. Often there is a mixed response, resulting in a persistent infection, but with less tissue damage than a hyper-inflammatory response would do.
- A persistently infected host may develop clinical disease if there is a waning of the immune mechanisms (e.g., in acquired immunodeficiency syndrome [AIDS]) that are critical to the control of infection, or a build-up of tissue damage over time. However, persistence can also protect from dangerous re-infection.

Protozoan pathogens make up a group of highly diverse organisms that represent a wide array of mechanisms of pathogenesis and immune evasion. There are numerous host targets for the intracellular protozoan parasites, including erythrocytes (*Plasmodium* and *Babesia*), macrophages (*Leishmania* and *Toxoplasma gondii*), or multiple cell types (*Trypanosoma cruzi, Plasmodium, Leishmania*). The luminal parasitic protozoan may be extracellular, such as amebae and the flagellates (*Giardia* and *Trichomonas*), or primarily intracellular, such as the coccidian parasite *Cryptosporidium*.

The innate and adaptive immune systems respond in diverse ways to the blood and tissue and intestinal protozoan pathogens. Neutrophils, macrophages, γδ T cells, and natural killer (NK) cells are the effector cells that mediate the innate response against the extracellular protozoan parasites. The NK cell–activated macrophage system is central to the innate response to intracellular parasites (Fig. 29.1) (Chapters 3 and 26). The innate cytokine response activates phagocytes and is critical to the induction of the adaptive immune response via antigen presentation by dendritic cells (DCs). For the intracellular pathogens (e.g., *Leishmania* spp., *T. cruzi*, *T. gondii*), the early production of interleukin-12 (IL-12) and interferon-γ (IFN-γ) drives the differentiation of T cells to a protective T-helper 1 (Th1) phenotype.[1] In most cases, CD4 T cells play a primary role in adaptive cellular immunity, but CD8 T cells can also be critically important through cytokine production (e.g., *T. cruzi*, *T. gondii*) or direct cytotoxic activity (e.g., *Cryptosporidium*). For the parasites with an extracellular stage (e.g., *Plasmodium* spp., *Trypanosoma* spp., *Giardia*, and *Trichomonas*), specific antibodies are critical to acquired immunity.

Intensive effort has been dedicated to the development of effective vaccines for protozoal diseases, but as of 2020, only one malaria vaccine (RTS,S) has reached the stage of clinical use, and efficacy and longevity are still limited. The reader is referred to the following reviews of potential vaccine candidates.[2–6] A discussion of the immune responses to some of the individual protozoal pathogens follows.

PLASMODIUM SPP.

Pathogenesis

Soon after *Plasmodium* spp. sporozoites are injected into the bloodstream by female *Anopheles* spp. mosquitos, they invade hepatocytes and undergo schizogony (asexual reproduction), multiplying asexually up to 10,000 times. Hypnozoites, a dormant form of *P. vivax* and *P. ovale*, can reside within hepatocytes for years and then undergo schizogony. *P. falciparum*, the most lethal species does not have a dormant form. Merozoites are then released from hepatocytes within a membranous structure, or merosome. Merozoites are released in the blood to invade red blood cells (RBCs). While the liver stage elicits primarily intracellular immunity, infected RBCs are the disease-causing stage. Parasite maturation within the RBC results in schizogony and release of new invasive merozoites through egress. The clinicopathological features of malaria are largely caused by the immune response to infected RBCs and egress. The cyclical release of parasites is regulated by food intake in mice and is associated with cytokine production and fever. This proinflammatory cytokine cascade plays a central role in the pathogenesis of severe malaria. Recognition of several pathogen-associated molecular patterns (PAMPs) drive this activation, including parasite proteins with glycophosphatidyl inositol (GPI) anchors, hemozoin, and uric acid in the extracellular milieu, and phagocytosed parasite nucleic acids.[7] The early production of proinflammatory cytokines (tumor necrosis factor [TNF], lymphotoxin, and IFN-γ) by NK cells and memory T cells leads to endothelial activation. This increases expression of endothelial adhesion molecules and hyper-coagulation, leading to poor perfusion due to unregulated clotting, immune cell

TABLE 29.1 Worldwide Significance of the Major Protozoal Infections

Parasite	Estimated Worldwide Cases (Annual Mortality)	Clinical Manifestations
Plasmodium spp.	200–250 million (*P. falciparum*: <1 million deaths/year, primarily children)	Fever, aches with potential complications in pregnancy and children: cerebral involvement, renal failure, pulmonary edema
Leishmania spp.	10–50 million people infected, 1.2 million new cases per year	Asymptomatic infection; skin ulcers or nodules; destructive oropharyngeal lesions; visceral disease with fever, hepatosplenomegaly, cachexia, pancytopenia
Trypanosoma cruzi	24 million (60,000 deaths)	Asymptomatic infection; dysrhythmias or chronic heart failure; hypertrophy and dilation of the esophagus, colon
Toxoplasma gondii	Several hundred million people infected worldwide. 5%–11% of healthy US adults are seropositive	Self-limited fever, hepatosplenomegaly; lymphadenopathy and encephalitis (reactivation in patients with acquired immunodeficiency syndrome [AIDS]); congenital infection, with fetal death, chorioretinitis, meningoencephalitis
Entamoeba histolytica	50 million (100,000 deaths)	Asymptomatic infection, diarrhea, dysentery, or liver abscess
Giardia lamblia	200 million (most common in young children and immunocompromised persons)	Asymptomatic infection, chronic diarrhea
Cryptosporidium parvum and *C. hominis*	Prevalence 3%–10% in patients with diarrhea in developing countries	Self-limited diarrhea in immunocompetent persons, severe intestinal and biliary disease in patients with AIDS
Trichomonas vaginalis	170 million/year	Asymptomatic infection, vaginal discharge, urethritis

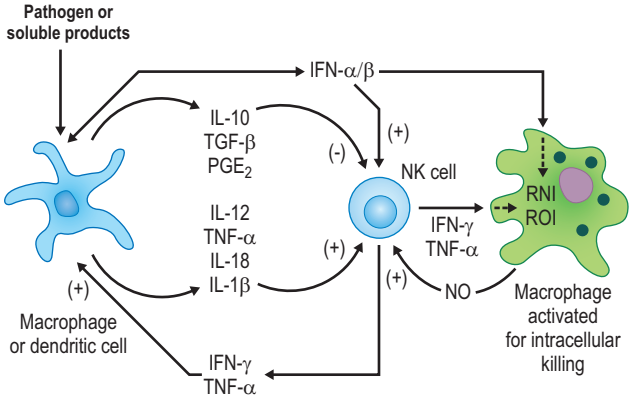


FIG. 29.1 Macrophage, Natural Killer *(NK)* Cell, and Cytokine Interactions in the Innate Immune Response to Intracellular Protozoa. Exposure of macrophages or dendritic cells to a pathogen or microbial product can result in the release of cytokines and inflammatory mediators that may stimulate (+) or suppress (–) NK cell activation. Activated NK cells produce cytokines that can then activate macrophages for intracellular killing. It must be recognized that this diagram is oversimplified and that these cytokines, most notably interferon *(IFN)*-α/β, interleukin *(IL)*-10, transforming growth factor *(TGF)*-β, and IL-12, may be produced by other types of cells, such as epithelial cells or enterocytes. *NO*, Nitric oxide; *PGE_2*, prostaglandin E_2; *RNI*, reactive nitrogen intermediates; *ROI*, reactive oxygen intermediates; *TNF-α*, tumor necrosis factor-α.

adhesion, and parasite sequestration in various organs causing symptoms of severe malaria. Early cytokine production has been linked to Toll-like receptor 2 (TLR2) on the macrophage surface, TLR7 and 9 in the endosome, and cytoplasmic STING and cGAS, leading to IFN-I production. Fever is linked to IL-1, which is NLRP3 and AIM2 inflammasome-dependent in *Plasmodium* infection. TNF family members play a strong role in severe malaria symptoms.[8]

Severe malaria includes placental malaria, severe anemia, respiratory distress, and cerebral malaria, with the latter causing up to 20% of deaths. In cerebral malaria, a combination of inflammation and vascular and multi-organ pathology leads to coma, seizures, and death. Pediatric patients may die as a result of severe edema of the brainstem leading to respiratory arrest. Severe anemia is due to the destruction of both infected and uninfected RBCs and dyserythropoiesis, which are increased by inflammatory cytokines, as is the damage in placental malaria. Immunoregulatory cytokines IL-10 and transforming growth factor-β (TGF-β) regulate the Th1 immune response, which drives parasite killing to limit pathology.[9]

KEY CONCEPTS

Immunopathogenesis of Severe Plasmodium falciparum Malaria

- Parasite components stimulate multiple innate recognition pathways to induce tumor necrosis factor (TNF) and lymphotoxin (LT), interferon type-I, and interleukin (IL)-1 production.
- TNF/LT induce vascular leakage and endothelial cell activation, with expression of adhesion molecules, platelet activation, leukocyte and parasite adhesion, and coagulation.
- Inflammatory cytokines amplify severe anemia caused by loss of infected red blood cells (RBCs) by inducing dyserythropoiesis and phagocytosis of uninfected RBCs, which can be coated with parasite antigens.
- Lethal outcomes for the fetus of infected women during pregnancy are caused by the local inflammation and vascular congestion induced by placenta-binding parasites.

Innate Immunity

While sporozoites rapidly transit from the skin to the liver, some remaining in hair follicles stimulate an immune response, including antibodies that can prevent re-infection. Complement can drive lysis of sporozoites and merozoites; however, C5 and C3 are also essential for pathogenesis of severe malaria in mice. Early production of IFN-γ by NK cells is associated with better outcomes of infection. γδ T cells are also activated

early in infection by phosphorylated lipid antigens correlating with immunity and protective vaccination. There appear to be innate specificities for both T cells and antibodies in naïve individuals that can recognize infection with *Plasmodium* spp. Macrophages stimulated by TNF and IFN-γ are important for parasite phagocytosis and killing. Animal studies have shown that phagocyte activation can limit initial growth of blood-stage parasite if strong enough, but this is regulated by the level of IFN-I triggered by each parasite strain. Antigen presentation by DCs can induce a strong adaptive immune response. These cells are replaced by inflammatory monocytes, which stimulate a less inflammatory response during prolonged infection, allowing the formation of germinal centers and parasite clearance.[7,8]

Adaptive Immunity

Immunity to *Plasmodium* spp. infection is acquired slowly. In areas of intense perennial *P. falciparum* transmission, the density of parasitemia and the incidence of severe malaria are highest in early childhood years, with severe disease declining after 1 to 2 infections. Disease immunity is likely to be due to subsequent production of a less pathogenic pattern of cytokines, though resistance to fever develops more slowly. Immunity to parasite growth is gained with exposure to multiple variants and development of strain-specific antibody. Therefore, most highly exposed adults are immune. However, in pregnant women, despite systemic immunity, the placenta can bind infected RBCs, causing mortality and morbidity in fetus and mother respectively.[8]

Adaptive immunity to the liver stage can be mediated by parasite-specific CD8 T cells via IFN-γ-induced NO-dependent killing of intra-hepatocytic parasites. While T-cell immunity to liver stage antigens is not high in exposed adults, CD8 T cells specific for pre-erythrocytic antigens can be protective. Attenuated sporozoite vaccines are in vaccine trials and are widely used in human challenge studies. One caveat is that since one escaped sporozoite can lead to malaria, anti-sporozoite vaccine induced protection would require very high antibody titers to quickly block invasion, or exceedingly high numbers of cytotoxic T cells to kill infected hepatocytes.[10]

Both humoral and cellular immunity is required against the erythrocytic stage of infection. Adoptive transfer of human immune serum is protective for naïve individuals. Antibodies directed against merozoite surface proteins can inhibit invasion by neutralization. The cytophilic isotype antibodies immunoglobulin G1 (IgG1) and IgG3 also play a significant role in naturally acquired immunity by opsonizing infected cells for phagocytosis in the spleen. However, an influx of blood-borne antigen induces a low-affinity IgM plasmablast response. Although there is an increase in atypical B cells, memory B cells and long-lived plasma cells are generated, leading to some strain-specific humoral immunity.[11] Parasite-antigen-specific B cells and CD4 T cells are required to generate germinal centers to make high-affinity antibody, to keep up with variant antigen switching, for complete clearance of parasites. CD4 T cells help B cells through a cognate interaction, and the production of IL-21 by T cells is crucial for isotype switching in *Plasmodium* infection. IFN-γ produced by CD4 and CD8 T cells, NK cells, and γδ T cells promotes phagocytosis of parasites, in synergy with TNF, IL-1, and opsonins. Although the response is characterized as Th1 due to the predominance of IFN-γ, and a little IL-4, the responding CD4 T cells express multiple cytokines. In addition to their role in helping B cells make antibody via IL-21, CD4 T cells are implicated in protection from severe pathology. CD4 T cells regulate the intense inflammatory response through production of antiinflammatory cytokines IL-10 and TGF-β, which is promoted by IL-27.[8]

Evasion of Host Immunity

Sporozoites and merozoites evade circulating antibody by rapidly entering hepatocytes or RBCs, respectively. Mature RBCs do not express MHC molecules on their surface and so avoid recognition by T cells. The few parasite proteins expressed on the erythrocyte surface either have hidden immunostimulatory epitopes or exist in multiple allelic forms that switch rapidly avoiding recognition by the adaptive immune system.[12] It is not yet clear if complement has a net protective effect, as infected RBCs express complement inhibitory receptors. Parasites also express a chemokine homologue and bind complement receptor 1, which leads to clumping of erythrocytes, a phenomenon called rosetting that probably helps parasite invasion, but also drives pathology.

TOXOPLASMA GONDII

Pathogenesis

Transmission occurs via the ingestion of oocysts, which are shed in the feces of felines or via tissue cysts present in undercooked meat. Following the oral ingestion of cysts, phagocytes recruited to the gut lumen facilitate transepithelial migration into the lamina propria. These phagocytes have been termed "Trojan horses," as they carry the parasite to new, unsuspecting host cells. Host-cell invasion starts with loose attachment facilitated by laminin. As in *Plasmodium*, intracellular tachyzoites replicate within a parasitophorous vacuole (PV) and ultimately leave the cell by using an active egress pathway. The released tachyzoites can disseminate to invade virtually any nucleated cell type, but mononuclear phagocytes are the preferred host cells. Under pressure from the host immune response, tachyzoite replication is controlled and tissue cysts, containing slowly replicating bradyzoites, are formed. The tissue cysts persist as a chronic latent infection as long as the host immune function is intact. If the latently infected person is immunosuppressed, reactivation occurs, and tachyzoites are released to infect more cells. Because tissue cysts are found in large numbers in the brain, reactivation of latent infection in the immunocompromised host most commonly manifests as encephalitis.

Innate Immunity

Similar to immunity to *Plasmodium* and *Leishmania*, type 1 immunity contributes to the control of the early stages of *T. gondii* infection.[1] Therefore, IL-12 is essential to host resistance. CD8 DCs are the primary producers of IL-12, through a MyD88-dependent mechanism. The parasite is recognized by TLR2 and TLR11 (in mice), and activation of CCR5, as well as intracellular STING.[13] IL-12-dependent activation of NK cells leads to their conversion into ILC1, which along with γδ T cells and neutrophils produce IFN-γ, which, in turn, activates macrophages to limit parasite replication.[14] In mice, the induction of immunity-related GTPases (IRG), which damage the PV membrane and kill *T. gondii* within the cytosol, is the primary macrophage effector mechanism. However, humans lack the IRG family and TLR11, and the mechanism(s) of macrophage activation in humans is unclear. Recognition by human $CD16^+$

monocytes and type 1 DCs is different from that in mice, as it comes after phagocytosis, suggesting a cytoplasmic sensor. The generation of reactive nitrogen and oxygen intermediates as well as degradation of tryptophan have all been implicated in the control of *T. gondii* in human macrophages.

Adaptive Immunity

Although serum antibodies can be used in the diagnosis of *T. gondii* infection, the systemic antibody response does not play a role in adaptive immunity. Mucosal IgA, however, does provide resistance to oral infection with *T. gondii* cysts. CD4 and CD8 T cells are highly activated during infection and are essential for adaptive immunity to prevent cyst reactivation during chronic latent infection. As such, patients with defects in T cell–mediated immune responses (e.g., patients with AIDS) are at risk for reactivation of latent infection. IL-12 also drives differentiation of Th1 cells and terminally differentiated effector CD8 T cells (Killer cell lectin-like receptor G1 positive [$KLRG1^+$]) in the response to *T. gondii* infection, and T-cell derived IFN-γ is required for protection. Parasite-specific cytolytic T cells have been demonstrated; however, CD8 T cells mediate protection primarily through the generation of IFN-γ. As in *Plasmodium* immunity, antiinflammatory molecules, particularly IL-10 and IL-27 made by Th1 cells, play an important role in modulating the adaptive immune response and restricting host tissue damage.[15]

Evasion of Host Immunity

T. gondii escapes early macrophage killing in a number of ways.[15] Virulent parasites are protected by the PV, which does not fuse with host cell lysosomes, or become acidified to kill the parasite. The infected macrophage is also a suboptimal target for T cell–induced immunity because of reduced expression of MHC class II and costimulatory molecules. Infection also induces the production of counterregulatory molecules, such as IL-10, TGF-β, IFN-I, and lipoxin A4, which both downregulate a potentially pathological host inflammatory response and inhibit macrophage antimicrobial activity. In addition, *T. gondii* interferes with normal macrophage signaling. For example, infection inhibits DNA binding of signal transducer and activator of transcription 1 (STAT1) and nuclear factor kappa B (NF-κB) activation and promotes antiinflammatory pathways downstream of the suppressor of cytokine synthesis proteins. The parasite also has several virulence proteins, including ROP5 and ROP16, which bind to the PV and reduce accumulation of the IRGs. ROP proteins also activate host STAT3 and STAT6, driving an M2 macrophage phenotype that is permissive to infection. The parasite also suppresses pDC activation through an ROP16 kinase-dependent mechanism.[15]

LEISHMANIA SPP.

Pathogenesis

The female phlebotomine sand fly becomes infected by ingesting *Leishmania* amastigotes during a blood meal. In the sand fly midgut, the amastigotes differentiate intracellularly into infectious extracellular metacyclic promastigotes that infect the vertebrate host during the next blood meal. Immunomodulatory factors present in sand fly saliva also enhance the infectivity of the parasite. Once injected into skin, promastigotes are phagocytosed by neutrophils, DCs, and macrophages (through complement–complement receptor–mediated coiling phagocytosis); they evade immunity and transform to amastigotes to replicate within the hostile acidic environment of the phagolysosome. A significant part of the pathogenesis following *T. cruzi* infection is its dissemination through the bloodstream to many tissues. The cytotoxic activity of CD8 T cells may promote cutaneous inflammation and lesion pathology.[16]

Innate Immunity

Much of what we know of immunity in leishmaniasis comes from studies of inbred mouse strains (B6 and BALB), which demonstrate genetically divergent innate and adaptive immune responses differentially shaping the outcome of infection. These studies also form the basis for much of what we know about the Th1/Th2 paradigm and cytokine-producing memory T cells in vivo. The innate immune response to *Leishmania* is mediated by complement, NK cells, cytokines, and phagocytes.[17] As in *Plasmodium* and *Toxoplasma*, IL-12 from DCs promotes crucial early IFN-γ production. Chemokines (IP-10, MCP-1, and XCL1), as well as interaction between LPG and TLR, can also promote early NK-cell activation, which can be cytolytic for *Leishmania*-infected macrophages. However, NK cell–derived IFN-γ plays a more prominent role in host defense by activating macrophages to kill the intracellular parasites through the generation of reactive oxygen intermediates (ROIs) or reactive nitrogen intermediates (RNIs). The generation of RNI by activated macrophages is the primary mechanism of parasite killing in the murine model. In human macrophages, inhibition of nitric oxide synthase 2 (NOS2) can impair killing of intracellular *Leishmania*. Parasite-induced, MyD88-dependent signaling through TLR2, TLR3, and TLR4, as well as type I IFN contribute to macrophage activation and NO production. Activated neutrophils can kill parasites through oxidative mechanisms, but their role in vivo depends on the timing of their recruitment and their interaction with other immune cells. Infiltrating neutrophils promote sand fly–transmitted infection through modulation of macrophage function following engulfment of apoptotic parasitized neutrophils.[17]

Adaptive Immunity

Within an endemic area there is acquisition of immunity in the population over time, and one infection (primary, subclinical, or healed) appears to be sufficient to induce immunity. Following primary infection, parasites persist for the life of the host and this persistence appears to be important for maintaining long-term immunity. Experimental models demonstrate that cellular immune mechanisms mediate adaptive resistance to *Leishmania* infection, and human studies have confirmed this. Because of the intracellular location of the parasite in the mammalian host, antileishmanial antibodies, which are produced at a low level in localized cutaneous leishmaniasis (LCL), and at a very high level in visceral leishmaniasis (VL), play no role in protection. The general mechanisms of cellular immunity in leishmaniasis are summarized in Fig. 29.2. Following infection in the skin, migratory dermal DCs phagocytose *Leishmania* and presumably transport the intracellular parasite to the regional lymph node, where they induce a T-cell response. Adaptive immunity is primarily mediated by parasite and IL-12-induced production of IFN-γ by CD4 T cells (Th1). CD4 T cells are absolutely required, but immunity to cutaneous disease is also mediated by CD8 T cells via production of IFN-γ.[16] Both CD4 and CD8 T cells are required for an effective defense against murine visceral *L. donovani* infection, but the precise role of CD8 T cells is unclear.

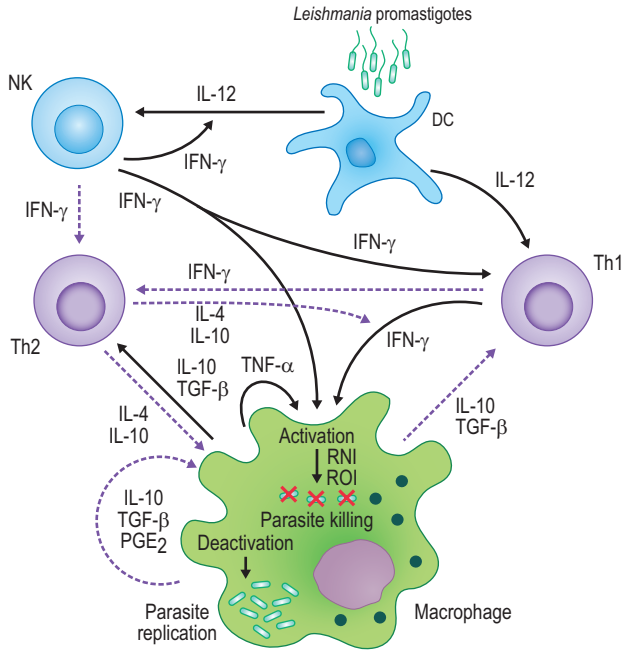


FIG. 29.2 Immunity in Leishmaniasis. Exposure of dendritic cells to parasites or parasite antigens leads to the release of interleukin *(IL)*-12, which induces natural killer *(NK)* cells to produce interferon *(IFN)*-γ and drives the adaptive immune response toward a protective T-helper 1 *(Th1)* phenotype. IL-12 production by dendritic cells and IFN-γ production by NK and Th1 cells negatively regulates the Th2 response. IFN-γ activates macrophages to kill the intracellular pathogen. In genetically susceptible individuals, a counterregulatory Th2 cytokine response can suppress the Th1 response and impair classic macrophage activation, leading to parasite replication and uncontrolled infection. Counterprotective macrophage-derived cytokines can also inhibit the Th1 response, stimulate the Th2 response, and impair classical activation through an autocrine loop. Activating stimuli are shown by *solid arrows*, and deactivating stimuli are shown by *dashed arrows*. *DCs*, Dendritic cells; *RNI*, reactive nitrogen intermediates; *PGE_2*, prostaglandin E_2; *ROI*, reactive oxygen intermediates; *TGF*, transforming growth factor.

The generation and maintenance of the protective Th1 response is critically dependent on CD40- CD40L-mediated IL-12 production by DCs. TNF also contributes to protective immunity by synergizing with IFN-γ to activate macrophages. Peripheral blood mononuclear cells (PBMCs) isolated from patients with LCL demonstrate a Th1 response to *Leishmania* antigens, and in the cutaneous lesion there is an exuberant Th1 and granulomatous response that mediates parasite killing and localized tissue damage, which usually leads to a scar. Patients with mucosal leishmaniasis (ML) exhibit vigorous cellular immune responses characterized by high levels of TNF and Th1 and Th17 cytokines contributing to the prominent tissue destruction of ML. Patients with diffuse cutaneous leishmaniasis resemble the progressive infection caused by *L. major* in BALB/c mice in that there are minimal or absent *Leishmania*-specific lymphoproliferative responses, and predominant Th2 cytokine expression. During active VL in humans, there is a marked depression of *Leishmania*-specific lymphoproliferative and IFN-γ responses, contraction of circulating memory T cells, and an absence of delayed-type hypersensitivity (DTH) response to parasite antigens.

Several subpopulations of memory CD4 T cells mediate immunity from cutaneous leishmaniasis. Effector memory cells, which are generated under conditions of antigen persistence, rapidly respond to secondary infection by migrating to the infected tissue and generating effector cytokines. Tissue resident memory cells (T_{rm}) also recruit other T cells to the site and contribute to efficient secondary anti-*Leishmania* immunity. Central memory T cells (T_{cm}) circulate throughout the lymphatic system and upon secondary challenge have a delayed response as they proliferate, and then migrate to the site of infection.[18]

Several adaptive immune mechanisms can also promote parasite replication and disease.[19] The progression of murine *L. major* infection has been correlated with the expansion of Th2 cells and the production of IL-4, IL-5, and IL-10. In susceptible mice, IL-4 production within the first day of infection was shown to downregulate IL-12 receptor β-chain expression and drive the response to a Th2 phenotype. However, other mouse strains appear to be able to overcome an early IL-4 response and develop a resistant phenotype, and susceptibility to some *L. major* strains is not strictly mediated by IL-4 (IL-13 and/or IL-10 may have a prominent role). The macrophage production of immune suppressive molecules, such as TGF-β or prostaglandin E_2 (PGE_2), may also contribute to susceptibility. The induction of T regulatory cells (Treg) in a PI3K-dependent way can also contribute to susceptibility by limiting the effectiveness of IFN-γ on infected macrophages.[2]

Evasion of Host Immunity

The *Leishmania* parasite has numerous ways in which it adapts to and survives within the vertebrate host.[20] In skin, the promastigotes may be phagocytosed by neutrophils and macrophages, which, unlike DCs, do not actively participate in T-cell priming. Furthermore, the clearance of apoptotic neutrophils is likely to make macrophages more permissive to infection. The expression of *Leishmania* surface protease GP63 and lipophosphoglycan (LPG) prevent the fusion of phagosome and lysosome, thereby enabling the parasite to replicate in phagocytes. They also confer complement resistance and facilitate the entry of complement-opsonized parasites into the macrophage without triggering a respiratory burst.

Leishmania-infected macrophages have diminished capacity to initiate and respond to a T-cell response, and the impaired antimicrobial effector activity provides a safe haven for the intracellular parasite.[20] Infected macrophages have decreased synthesis of IL-1 and IL-12, and blunted IFN-γ-mediated activation through the disruption of signal transduction pathways. Conversely, there is increased synthesis of the immunosuppressive molecules IL-10, TGF-β, and PGE_2. Parasite factors and IL-4/IL-13 enhance expression of arginase by infected macrophages. Both host and parasite arginase promote infection and pathogenesis through depletion of L-arginine, enhanced production of polyamines, and reduced NO production.[2] IL-10 produced by $CD4^+CD25^+$ Treg has an essential role in parasite persistence. Successful treatment of active disease restores an antigen-specific Th1 response.[19] Metastasizing parasites that cause mucosal disease also harbor *Leishmania* RNA virus, which subverts the host immune response and promotes parasite persistence through activation of TLR3 and the NLRP3 inflammasome.[21]

TRYPANOSOMA CRUZI

Pathogenesis

T. cruzi is transmitted to the mammalian host when the infectious metacyclic trypomastigote, which is deposited on skin in the feces of the reduviid insect vector while it takes a blood meal, is scratched into the wound or transferred to a mucous membrane (e.g., the eyes). The trypomastigotes can infect almost any cell type and replicate as amastigotes in the cytoplasm. Eventually, the amastigotes transform back into trypomastigotes and rupture the cell to enter the bloodstream, from where they invade other cells or are picked up by another insect vector. Muscle and glial cells are the most frequently infected cells, and acute myocarditis or meningoencephalitis can develop. In most cases, however, primary infection occurs without clinical symptoms (70% to 90%), and the infected individual may enter an indeterminate phase of asymptomatic seropositivity. Chronically infected individuals can develop symptomatic Chagas disease, usually involving the heart or the gastrointestinal (GI) tract. Pathologically, there are few parasites observed in cardiac tissue, but an intense chronic inflammatory infiltrate with fibrosis and loss of muscle fibers is evident. In the digestive tract, there is lymphohistiocytic infiltration of the myenteric plexuses, with reduction in the number of ganglion cells.

The tissue damage of acute *T. cruzi* infection is the result of a direct effect of the parasite and an indirect effect of the acute inflammatory and oxidative response. In chronic infection, the balance between immune-mediated parasite containment and inflammation-driven oxidative damage determines the course of disease. Whether tissue damage is caused directly by parasites or indirectly through parasite-driven inflammatory or autoimmune mechanisms, parasite persistence is a significant driver of disease (Table 29.2).[22] Autoimmunity could arise from molecular mimicry of self by parasite antigens or by the release of self molecules from damaged or dying host cells within the environment of an activated immune response.

Innate Immunity

As in the intracellular parasite immune reactions above, the early innate immune response to *T. cruzi* infection is mediated by NK cells, DCs, and macrophages.[23] Macrophages and DCs exposed to *T. cruzi* trypomastigote antigens produce IL-12 and TNF through a MyD88-dependent mechanism. A number of trypomastigote antigens, including free GPI anchors, glycoinositol phospholipids (GIPLs), GPI-linked glycoproteins, and GPI-mucins activate the innate immune response, through TLR2 and possibly TLR4. Immune activation through the Toll/IL1R domain-containing adapter protein–inducing IFN-β (TRIF), another protein involved in TLR2 signaling, promotes resistance through IFN-β and of IFN-β-inducible genes, such as the p47 guanosine triphosphatases (GTPase) *IRG47*. CpG DNA motifs in the *T. cruzi* genome also activate TLR9 and nucleotide-binding oligomerization domain (NOD)–like receptors are also activated intracellularly.[23] NK cell IFN-γ synergizes with TNF to activate macrophages to control parasite replication. The generation of NO is the primary trypanocidal mechanism in murine macrophages.

Adaptive Immunity

The innate immune response, through production of IL-12, type I IFNs, and other proinflammatory mediators, is critically linked to the generation of an effective adaptive immune response. Antibodies contribute to immunity within the bloodstream through opsonization, complement activation, and antibody-dependent cellular cytotoxicity. Several lines of evidence establish the importance of CD4 and CD8 T cells in adaptive immunity to *T. cruzi* infection. CD8 T cells specific to *T. cruzi*–infected cells confer protection when passively transferred to naïve mice. IFN-γ and TNF production by parasite-specific CD8 T cells is more important than cytolytic activity in the control of infection. CD4 T cells, followed by CD8s, are the predominant infiltrating cells in cardiac tissue.[24]

T. cruzi infection leads to a regulated Th1 cytokine response as described for *Plasmodium* and *Toxoplasma*. While IL-12/STAT4-dependent IFN-γ production is critically important to protection, the regulatory cytokines IL-10 and TGF-β can promote parasite replication by inhibiting macrophage activity, and also play a critical role in minimizing inflammation-mediated tissue pathology by regulating Th1 and TNF responses. Polymorphisms in the IL-10 gene that lead to reduced IL-10 production are associated with increased severity of cardiomyopathy in patients with Chagas disease. In addition to these regulatory

TABLE 29.2 Evidence for Autoimmune and Parasite-Induced Inflammatory Mechanisms in Chronic Chagas Disease

Evidence for Autoimmune-Mediated Disease	Evidence for Parasite-Induced Inflammatory Disease
Inflammatory disease presents in tissues with few or no parasites on routine histopathology studies	Sensitive parasite detection techniques (polymerase chain reaction [PCR], immunohistochemistry) correlate with the presence of parasites (or parasite material) and severity of inflammatory disease
Peculiar pattern of organ involvement (heart and gastrointestinal tract) in patients with chronic disease	Organs free of parasites (by sensitive parasite detection techniques) are also free of disease
Long delay in the onset of chronic disease following infection; only a minority of infected persons develops disease	Absence of effective cellular immune response (in mice or humans) almost invariably exacerbates rather than reduces the parasite burden and disease
Wide variability in the expression of disease among infected people	In chronically infected mice, the destruction of a transplanted heart is dependent on parasite infiltrating the transplanted tissue
Self-reactive antibodies and T cells demonstrable in infected people and in experimental animals. Level of antibodies to the ribosomal P protein (R13 peptide) and cardiac myosin (B13 antigen) correlate with cardiac disease	Degree of disease in hearts transplanted into chronically infected mice correlates with level of parasite burden in transplanted tissue
Transient or limited disease reported in experimental models following lymphocyte transfer	Reduction of parasite burden by chemotherapy usually leads to decreased tissue inflammation and disease

cytokines, the secretion of prostaglandins and the expansion of a myeloid suppressor cell population serve to control the intensity of the immune response.

Evasion of Host Immunity

Trypanosomal clonal variation and host genetic polymorphisms both contribute to tissue tropism, parasite persistence, and disease severity. *T. cruzi* bloodstream trypomastigotes resist complement lysis via a complement regulatory protein (GP-160). This parasite protein is functionally similar to mammalian decay-accelerating factor in that it inhibits C3 convertase formation and activation of the alternate complement pathway. *T. cruzi* invades host cells, in particular cardiomyocytes, by subverting a host plasma membrane repair pathway that promotes parasite persistence and tissue tropism. The rapid escape of the parasite from the phagosome into the cytoplasm through the action of acid-activated porins enables the organism to avoid phagolysosomal enzymatic destruction.

The establishment of chronic infection by *T. cruzi* is favored by a generalized depression of T-cell responses. A number of different mechanisms may contribute to this, including low IL-2 production or IL-2 receptor expression; downregulation of components of the T cell–receptor complex; T-cell receptor dysfunction; apoptosis of T cells; defects in the processing and presenting of antigens in the MHC class II (but not the class I) pathway; T-cell or macrophage suppressor activity; and PGE_2 production. Within the foci of myocarditis, apoptosis of both parasites and host cells occurs. The phagocytosis of these apoptotic cells by macrophages leads to their acquisition of an M2 phenotype, which enables parasite replication and persistence.

ENTAMOEBA HISTOLYTICA

Pathogenesis

E. histolytica causes asymptomatic intestinal colonization, acute diarrhea, dysentery, colitis, liver abscess, and, rarely, disseminated disease. Transmission of *E. histolytica* is through ingestion of cysts in fecal-contaminated food or water. When the cysts reach the intestine, trophozoites are released and migrate to the colon, where they adhere to the outer mucus layer covering the epithelium. In asymptomatic infections, trophozoites are confined to the outer mucus layer and survive by feeding on commensal bacteria and sugars from mucus. While mechanisms of invasive disease are unclear, *E. histolytica* is dependent on commensal bacteria for survival, but the composition of the microbiome (commensal bacteria, archaea, viruses, and eukaryotes) also contributes to driving disease pathology.[25] Risk factors for symptomatic disease, including the host's nutritional status, genetics, age, and sex, can alter the composition of the intestinal microbiome. When the microbiome is altered, trophozoites degrade the colonic mucus layers through secretion of proteases and glycosidases, allowing them to adhere directly to colonic epithelial cells by a galactose/*N*-acetylgalactosamine-inhibitable lectin (Gal/GalNAc). Exposure to the parasite induces epithelial cell lysis via contact-dependent mechanisms or through the release of lytic factors. Following epithelial damage, the parasite penetrates the mucosa and degrades connective tissue by releasing cysteine proteases, resulting in ulceration of the mucosa and submucosa (Fig. 29.3). *E. histolytica* is also capable of lysing a variety of other host cell types, including neutrophils, which release enzymes that further damage the tissue. Inflammation is a hallmark of invasive disease and characterized by the secretion of proinflammatory cytokines IL-8 and TNF and infiltration of immune cells. Amebic liver abscesses develop when trophozoites erode through the intestinal submucosa, enter the portal circulation, and disseminate to the liver.[26]

Innate Immunity

Through adherence and secretion of immunomodulatory proteins, trophozoites stimulate intestinal epithelial cells to release a variety of proinflammatory mediators, including IL-1α, IL-1β, IL 6, IL-8, and TNF that trigger the recruitment of neutrophils and macrophages to the site of invasion. Following activation by IFN-γ and TNF, neutrophils and macrophages become amebicidal through the release of reactive oxygen species and production of nitric oxide, respectively. Critical for controlling amebic liver abscesses, invariant natural killer T cells (iNKT) are specifically stimulated by amebic lipopeptidephosphoglycan (LPPG) to release IFN-γ. Production of IFN-γ reduces parasite burden and controls abscess formation.[26]

Adaptive Immunity

Secretory IgA responses against *E. histolytica* have been well characterized and shown to correlate with protection against infection and disease. For this reason, vaccines designed to generate IgA antibodies against the Gal/Gal/NAc lectin have been highly efficacious in preventing experimental *E. histolytica* infection in mice and baboons. However, the role of anti-lectin IgG is unclear, and protection may be subclass-dependent.[27] Cell-mediated immunity is also critical for host defense against *E. histolytica*. IFN-γ–producing CD4 T cells provide protection through their ability to activate macrophages and neutrophils to the parasite. CD8 T cells mediate protection through the secretion of IL-17, a key player in the secretion of mucin and antimicrobial peptides, recruitment of neutrophils, and IgA transport across the epithelial barrier. In chronic infection, *E. histolytica* promotes the development of Tregs that suppress the proliferation of responder T cells by releasing IL-10, TGF-β, and IL-35.[26] Impairment of cell-mediated immunity results in parasite dissemination. Indeed, patients with human immunodeficiency virus infection and coinfected with *E. histolytica* have high rates of invasive amebiasis, liver abscesses, and seroconversion.[27]

Evasion of Host Immunity

E. histolytica utilizes a number of strategies to circumvent the immune defenses of the host. It resists complement-mediated lysis during hematogenous spread by proteolytic degradation of C3a and C3b. In addition, the Gal/GalNAc lectin binds to C8 and C9, preventing assembly of the C5b-9 membrane attack complex. The cytolytic capability of *E. histolytica* affords protection from neutrophils, macrophages, and eosinophils, unless these cells are activated. Cytolysis by *E. histolytica* can occur via necrosis and apoptosis. Trophozoites also inhibit the macrophage respiratory burst and the production of IL-1 and TNF. A protective antibody response is subverted by the degradation of IgA and IgG by amebic cysteine proteases and by capping, ingesting, or shedding ameba-specific antibodies. Amebic proteases can also cleave the Fc region so that interaction with host cell-surface receptors is avoided. The suppression of host macrophage NO production by an array of trophozoite secretory products, including parasite-derived PGE_2, is a major factor in the persistence of amebic liver abscesses. *E. histolytica* expresses a homolog of the proinflammatory cytokine macrophage

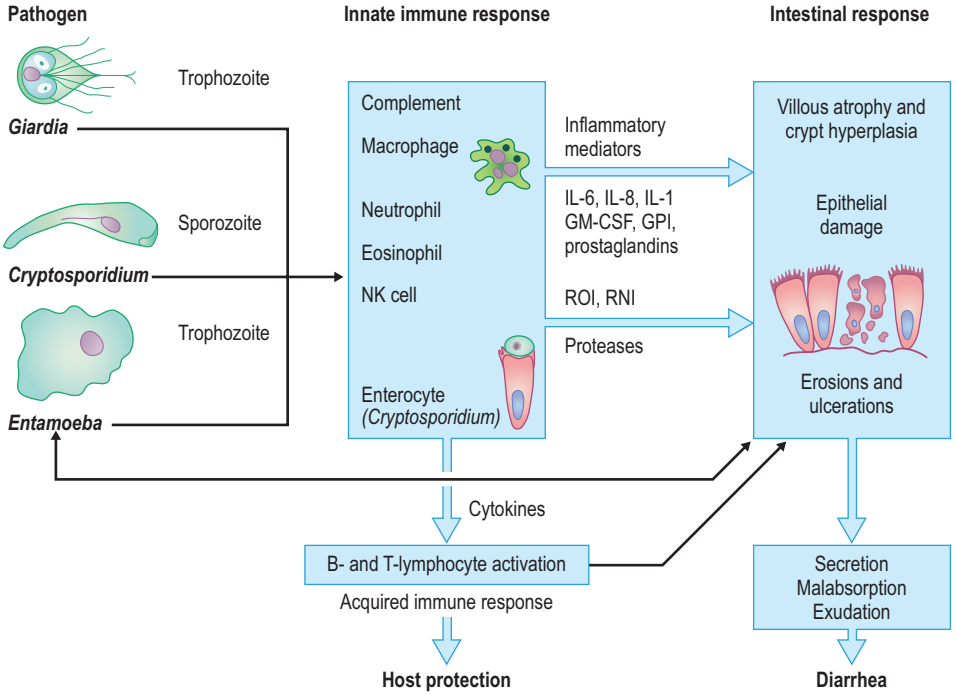


FIG. 29.3 Immunopathogenesis of Intestinal Protozoal Pathogens. After adherence (*Giardia* and *Entamoeba*) or epithelial invasion (*Entamoeba* and *Cryptosporidium*), there is release of various inflammatory mediators from macrophages and neutrophils. This causes the activation of resident phagocytes and recruitment of phagocytes into the lamina propria. Enterocyte death can be due to direct action of the parasites or to immune-mediated damage from complement, cytotoxic lymphocytes, proteases, and reactive oxygen and nitrogen intermediates (*ROI* and *RNI*, respectively). The inflammatory mediators also act on enterocytes and the enteric nervous system, inducing the secretion of water and chloride. In response to enterocyte damage, under the influence of activated T lymphocytes, the crypts undergo hyperplasia, and the villi become shorter (villous atrophy). The immature hyperplastic cells have poor absorptive ability but retain secretory ability. Damage to the epithelium can cause leakage (exudation) from lymphatics and capillaries. Similar mechanisms are probably responsible for the diarrhea that occurs in infection with *Cyclospora* and *Isospora*. *Isospora* is unique in causing an eosinophilic infiltrate. *GM-CSF*, Granulocyte macrophage–colony-stimulating factor; *GROα*, growth-related oncogene alpha; *IL*, interleukin; *NK*, natural killer.

migration inhibitory factor (MIF) and also monocyte locomotion inhibition factor (MLIF), which respectively promote gut inflammation and tissue invasion and phagocytic respiratory burst and other monocyte functions.

GIARDIA DUODENALIS

Pathogenesis

G. duodenalis (syn. *Giardia lamblia*, *Giardia intestinalis*) is a complex of eight genetic assemblages (designated A to H), two of which are infectious to humans and a variety of other mammalian hosts (assemblages A and B). The severity of giardiasis ranges from asymptomatic carriage to chronic watery diarrhea, epigastric pain, nausea, vomiting, and weight loss, depending on host factors and the virulence of the *Giardia* strain. Recent studies suggest even with treatment the parasite can elicit intestinal complications that persist for years.[28] Younger age, malnutrition, and immunodeficiency increase the risk of severe disease. Infection is initiated following the ingestion of food or water contaminated with *G. duodenalis* cysts. Exposure to stomach acids induces the excystation process that releases two trophozoites into the lumen of the proximal small intestine. Colonization occurs when the parasite attaches to the intestinal epithelium and begins to reproduce by binary fission. Trophozoites remain within the lumen and do not invade the epithelial barrier. Parasite migration into the lower intestine triggers encystation, allowing the organism to survive when excreted into the environment.

The *G. duodenalis* trophozoite initiates adherence to the intestinal epithelium via a surface mannose-binding lectin (MBL). Histopathological changes in symptomatic giardiasis range from a normal appearance to increased crypt-villous ratios, epithelial damage, and chronic inflammatory infiltrate in the lamina propria (see Fig. 29.3). The factors responsible for the structural changes in the small bowel may include injury from adherence, parasite-induced apoptosis of epithelial cells, and the release of cytotoxins, including proteases. Additional epithelial damage may be mediated by the host cellular immune response and changes in the composition of commensal microbiota. Diarrhea arises from epithelial barrier dysfunction, reduction in microvillous surface area, chloride hypersecretion, and glucose and sodium malabsorption.[29]

Innate Immunity

As *G. duodenalis* does not invade the intestinal epithelium, host defense and immune factors present within the lumen are essential for preventing and controlling tissue infection. Antimicrobial peptides made by Paneth cells, including cryptdins, neutrophil defensin, cathelicidin, and lactoferrin effectively

kill *G. duodenalis* trophozoites in vitro. NO, produced by both epithelial cells and macrophages, inhibits trophozoite division; however, the parasite can circumvent this defense by competing with host cells for arginine uptake. Mast cells have a significant role in protecting against the parasite. Mice deficient in mast cells fail to clear *G. duodenalis* infection, in part because they are unable to mount parasite-specific IgA. Mast cells also contribute to B-cell survival, activation, and differentiation into plasma cells, and together with NO induce peristalsis. Increased intestinal motility contributes to *G. duodenalis* clearance, by impairing the ability of the parasite to attach to the epithelium and resist the luminal bulk flow.[29]

Adaptive Immunity

Several lines of evidence suggest the importance of the humoral immune response in the control of giardiasis. Infection with *Giardia* results in the production of anti-*Giardia* antibodies in the serum and mucosal secretions. Patients with severe B-cell defects or selective IgA deficiency have an increased risk of developing chronic infections. Mouse studies showed key functions of secretory IgA and the polymeric immunoglobulin receptor, which transports IgA into the intestinal lumen, in controlling parasite burden.[29] A reduction or absence of CD4 T cells can also lead to chronic infection. IFN-γ- and IL-17A-producing CD4 T cells developed following infection[30] contribute to parasite clearance. IL-17A is likely involved in modulating transport of IgA into the intestinal lumen. Epidemiological studies indicate that partial immunity is acquired from *Giardia* infection, which leads to reduced risk and severity of subsequent infections.

Evasion of Host Immunity

Giardia evades the host humoral immune response with surface antigenic variation using a group of variant-specific surface proteins (VSPs). VSP switching occurs when intestinal anti-VSP IgA responses are first detected. *G. duodenalis* also produces cysteine proteases that degrade immunoglobulins, inflammatory mediators, and host defensins.[28] Although *Giardia* activates DCs for antigen presentation, it also inhibits IL-12 production, in part by enhancing IL-10 release; the net result is the dampening of a local antiparasitic inflammatory response. The trophozoite also releases arginine deiminase, which degrades arginine, making it less available for host NO production.[29]

CRYPTOSPORIDIUM SPP.

In humans, there are four intestinal coccidians that are intracellular parasites of enterocytes: *Isospora belli*, *Cyclospora cayetanensis*, and two species of *Cryptosporidium*, *C. parvum* and *C. hominis*. Of the four coccidians, *Cryptosporidium* has the greatest epidemiological significance: in 1993, an outbreak involving 403,000 persons occurred in Milwaukee, in the United States. Because of their similarity, only the immunology of cryptosporidiosis will be discussed.

Pathogenesis

Cryptosporidium typically causes self-limited (but often prolonged) diarrhea in the immunocompetent host. However, in the immunocompromised host *Cryptosporidium* can cause severe diarrhea, with malabsorption and wasting, and cholangiopathy. Infection begins with the ingestion of food or water contaminated with oocysts. The acidic environment of the stomach induces excystation and the release of four sporozoites per parasite into the small intestines. After entry into epithelial cells using its surface glycoproteins, the parasite resides within a unique intracellular but extracytoplasmic vacuole, which protects the pathogen. Here, the sporozoites grow and undergo schizogony. Individual merozoites emerge and invade neighboring epithelial cells. The merozoites may continue an asexual cycle or develop into macro- or microgametes that fuse to form oocysts. Before being excreted in feces, oocysts undergo sporulation to become infectious. Histologically, infection causes villous atrophy and blunting, and crypt hyperplasia with increased infiltration of lymphocytes, macrophages, and plasma cells.[31] Intraepithelial lymphocytes (IELs) are uncommon; neutrophils and occasional eosinophils are present between the epithelium and the lamina propria. Disorganized cells undergoing necrosis replace normal enterocyte architecture (see Fig. 29.3). There is an association between the degree of intestinal injury and malabsorption and the intensity of infection, as measured by oocyst excretion.

The neuropeptide substance P, which is produced by endothelial cells, lymphocytes, and monocytes in the lamina propria, contributes to diarrhea in patients with AIDS as well as cryptosporidiosis by increasing intestinal chloride secretion and glucose malabsorption.[32]

Innate Immunity

IFN-I and IFN-γ are critical in the innate protective response against *Cryptosporidium*. Because of the parasite's intracellular location, near the luminal surface of the enterocyte, the macrophages of the lamina propria are spatially separated from the parasite, rendering them inert and the intestinal epithelium's response critical. Endothelial cells are activated by TLR2/TLR4–dependent NF-κB activation and release of the microbicidal peptide β-defensin-2, TNF, and the chemokines IL-8, and RANTES, growth-regulated oncogene α (GROα), which act as chemoattractants and activators of neutrophils.[31]

IL-15, produced by activated monocytes, stimulates NK-cell proliferation, cytotoxicity, and cytokine production, including IFN-γ. IL-15 levels in the jejunal mucosa in immunocompetent patients inversely correlate with parasite burden. However, in patients with AIDS who have chronic uncontrolled cryptosporidiosis, IL-15 is undetectable. Infected intestinal cells also release TGF-β, which decreases necrosis and stimulates the synthesis of extracellular matrix proteins, thereby limiting epithelial damage. Prostaglandins E_2 and $F_{2\alpha}$, released by infected enterocytes, not only promote secretory diarrhea but also upregulate mucin production, which may hinder parasite attachment. They also stimulate the release of β-defensin-2, which has direct anticryptosporidial activity and is chemotactic for T cells and DCs. MBL is a serum complement protein that binds to various pathogens, including *Cryptosporidium*, and opsonizes them. Low serum levels of MBL, the result of malnutrition or genetic polymorphisms, increases susceptibility to cryptosporidiosis.[31]

Adaptive Immunity

Activated by macrophages and DCs, cell-mediated immunity plays an important role in the resolution of cryptosporidiosis and protection from re-infection.[6] In immunocompetent, adult mice CD4 IELs initiate early control of infection, whereas cytotoxic CD8 IELs appear later and function in parasite elimination. Resolution of infection depends on a balance of Th1 cytokines (IFN-γ, IL-18) needed to control the infection and

Th2 cytokines (IL-4, IL-10, and IL-13) that limit immunopathological damage. In mice, γδ T cells are rapidly recruited to control cryptosporidial infection, but their role in human infection is unknown. Severe intestinal disease or biliary involvement is usually seen in patients with AIDS whose CD4 count is less than 50/μL.[31]

The role of humoral immunity is less clear. Immunoglobulin deficiencies are often associated with persistent or recurrent infections. However, specific anti-*Cryptosporidium* IgA levels have been reported in patients with AIDS who are heavily infected with the parasite, and could play a role in neutralizing parasites, preventing attachment to the epithelium, but clearly are not sufficient.[31] In experimentally infected human volunteers, serum IgM and IgG directed toward sporozoite proteins protect against the development of symptoms, but not infection.

Evasion of Host Immunity

Cryptosporidium evades host defenses primarily by exerting control over infected enterocyte apoptosis. One of the upregulated genes is osteoprotegerin, which inhibits apoptosis by acting as a decoy receptor for TNF-related apoptosis inducing ligand (TRAIL). Control of host apoptosis is complex; early inhibition of apoptosis by NF-κB activation allows the parasite to complete its life cycle, whereas the late promotion of apoptosis facilitates merozoite release. Nevertheless, the infected cells secrete FasL, which promotes apoptosis in uninfected bystander cells. In this way, the host counters the antiapoptotic activity of the parasite by surrounding the parasitized cells with a zone of apoptotic cells. In patients with AIDS and cryptosporidiosis, the HIV Tat protein may sabotage host defense against *Cryptosporidium* by inhibiting cholangiocyte TLR4 expression.

TRICHOMONAS VAGINALIS

Pathogenesis

T. vaginalis is a flagellated protozoan parasite of the human urogenital tract that exists only as a trophozoite. It causes vaginitis, cervicitis, and urethritis. Its adherence to the vaginal squamous epithelium is facilitated by a number of adhesins. *Trichomonas* causes tissue damage by contact-dependent cytolysis caused by pore-forming proteins and proteases, and secretion of a glycoprotein cell-detaching factor that causes sloughing of the vaginal epithelium. Levels of the cell-detaching factor correlate with the severity of the disease, and vaginal antibodies directed against this factor modulate its effects. Inflammation in the genital mucosa and submucosa leads to copious secretions, and the surface epithelium may slough, causing focal erosions and hemorrhage.

The increased risk of HIV transmission in women with trichomoniasis may result from increased recruitment of inflammatory cells, mucosal erosion, or degradation of secretory leukocyte protease inhibitor (SLPI) by trichomonal proteases. Lower levels of SLPI are found in the vaginal fluid of women with trichomoniasis, which can lead to increased tissue damage and HIV transmission.[33] The LPG of *Trichomonas* induces production of the chemokines IL-8 and CCL20, which can also facilitate HIV infection by promoting DC recruitment.

Innate Immunity

Although both innate and adaptive immune responses are generated during *T. vaginalis* infection, evidence suggests that innate immunity plays a more prominent role in host protection and parasite elimination.[34] *Trichomonas* secretes a factor that promotes neutrophil chemotaxis, causing profuse leukorrhea, but the oxidative microbicidal mechanisms of the neutrophils have decreased efficacy in the anaerobic vaginal environment. Activated macrophages can destroy trichomonads in a T and B cell–independent manner and release IL-lβ and TNF. *Trichomonas* induces neutrophil apoptosis, and macrophage clearance of these apoptotic cells causes release of IL-10, which may contribute to resolution of the inflammatory response.[35]

Adaptive Immunity

Repeated infections with *T. vaginalis* do not induce immunity; however, the infection is self-limited in most cases, demonstrating that there are effective mechanisms of host defense. *T. vaginalis* induces the production of antibodies in both the serum and vaginal secretions. The serum antibody response correlates with active infection, and serum, but not vaginal, IgG from infected patients displays complement-mediated lytic activity against trichomonads in culture.[35]

Evasion of Host Immunity

Although *T. vaginalis* activates the alternative pathway of complement, the cervical mucus and menstrual blood are low in complement. Parasite virulence is enhanced, and the symptoms are exacerbated during menses. Menstrual blood supplies iron, which upregulates trichomonal adhesins and cysteine proteases, causing the degradation of complement component C3 bound to the surface of the parasite. Cysteine proteases secreted by *T. vaginalis* also degrade immunoglobulins, sabotaging the antibody response.[34] To further limit humoral immunity, the parasite secretes soluble antigens that are lytic to B and T cells or act as decoys for neutralizing antibodies. Phenotypic variation of surface antigens and ability to disguise itself by binding to host plasma proteins are additional mechanisms utilized by the parasite to escape immune detection.

ON THE HORIZON

Anticipated Approaches to Improved Control of Protozoal Diseases

- Gaining insight into the role of host genetics, gender, intestinal microbiota, and nutritional status in the outcome of protozoal infections.
- Defining host immune and inflammatory mechanisms that promote disease caused by protozoal pathogens.
- Understanding mechanisms for sustained generation of protective immunity against protozoal pathogens.
- Developing vaccines for several blood and tissue protozoa.
- Developing new host-directed treatment strategies to overcome immune evasion by protozoal pathogens.

REFERENCES

1. Butler NS, Harris TH, Blader IJ. Regulation of immunopathogenesis during *Plasmodium* and *Toxoplasma* infections: more parallels than distinctions? *Trends Parasitol.* 2013;29(12):593–602.
2. Ikeogu NM, Akaluka GN, Edechi CA, et al. *Leishmania* immunity: advancing immunotherapy and vaccine development. *Microorganisms.* 2020;8(8):1201.
3. Hotez PJ, Pecoul B, Rijal S, et al. Eliminating the neglected tropical diseases: translational science and new technologies. *PLoS Negl Trop Dis.* 2016;10(3):e0003895.

4. Cockburn IA, Seder RA. Malaria prevention: from immunological concepts to effective vaccines and protective antibodies. *Nat Immunol*. 2018;19(11):1199–1211.
5. Vazquez-Chagoyan JC, Gupta S, Garg NJ. Vaccine development against *Trypanosoma cruzi* and Chagas disease. *Adv Parasitol*. 2011;75:121–146.
6. Lemieux MW, Sonzogni-Desautels K, Ndao M. Lessons learned from protective immune responses to optimize vaccines against cryptosporidiosis. *Pathogens*. 2017;7(1):2.
7. Gazzinelli RT, Kalantari P, Fitzgerald KA, Golenbock DT. Innate sensing of malaria parasites. *Nat Rev Immunol*. 2014;14(11):744–757.
8. Gbedande K, Carpio VH, Stephens R. Using two phases of the CD4 T cell response to blood-stage murine malaria to understand regulation of systemic immunity and placental pathology in *Plasmodium falciparum* infection. *Immunol Rev*. 2020;293(1):88–114.
9. Moxon CA, Gibbins MP, McGuinness D, et al. New insights into malaria pathogenesis. *Annu Rev Pathol*. 2020;15:315–343.
10. Lefebvre MN, Harty JT. You shall not pass: memory CD8 T cells in liver-stage malaria. *Trends Parasitol*. 2020;36(2):147–157.
11. Portugal S, Obeng-Adjei N, Moir S, et al. Atypical memory B cells in human chronic infectious diseases: an interim report. *Cell Immunol*. 2017;321:18–25.
12. Wahlgren M, Goel S, Akhouri RR. Variant surface antigens of *Plasmodium falciparum* and their roles in severe malaria. *Nat Rev Microbiol*. 2017;15(8):479–491.
13. Lopez-Yglesias AH, Camanzo E, Martin AT, et al. TLR11-independent inflammasome activation is critical for CD4$^+$ T cell-derived IFN-gamma production and host resistance to *Toxoplasma gondii*. *PLoS Pathog*. 2019;15(6):e1007872.
14. Park E, Patel S, Wang Q, et al. *Toxoplasma gondii* infection drives conversion of NK cells into ILC1-like cells. *Elife*. 2019;8:e47605.
15. Hunter CA, Sibley LD. Modulation of innate immunity by *Toxoplasma gondii* virulence effectors. *Nat Rev Microbiol*. 2012;10(11):766–778.
16. Novais FO, Carvalho AM, Clark ML, et al. CD8+ T cell cytotoxicity mediates pathology in the skin by inflammasome activation and IL-1beta production. *PLoS Pathog*. 2017;13(2):e1006196.
17. Gurung P, Kanneganti TD. Innate immunity against *Leishmania* infections. *Cell Microbiol*. 2015;17(9):1286–1294.
18. Glennie ND, Volk SW, Scott P. Skin-resident CD4$^+$ T cells protect against *Leishmania major* by recruiting and activating inflammatory monocytes. *PLoS Pathog*. 2017;13(4):e1006349.
19. Soong L, Henard CA, Melby PC. Immunopathogenesis of non-healing American cutaneous leishmaniasis and progressive visceral leishmaniasis. *Semin Immunopathol*. 2012;34(6):735–751.
20. Geiger A, Bossard G, Sereno D, et al. Escaping deleterious immune response in their hosts: lessons from trypanosomatids. *Front Immunol*. 2016;7:212.
21. de Carvalho RVH, Lima-Junior DS, da Silva MVG, et al. *Leishmania* RNA virus exacerbates leishmaniasis by subverting innate immunity via TLR3-mediated NLRP3 inflammasome inhibition. *Nat Commun*. 2019;10(1):5273.
22. Bonney KM, Luthringer DJ, Kim SA, et al. Pathology and pathogenesis of Chagas heart disease. *Annu Rev Pathol*. 2019;14:421–447.
23. Tarleton RL. Immune system recognition of *Trypanosoma cruzi*. *Curr Opin Immunol*. 2007;19(4):430–434.
24. Acosta Rodriguez EV, Araujo Furlan CL, Fiocca Vernengo F, et al. Understanding CD8(+) T cell immunity to *Trypanosoma cruzi* and how to improve it. *Trends Parasitol*. 2019;35(11):899–917.
25. Uribe-Querol E, Rosales C. Immune response to the enteric parasite *Entamoeba histolytica*. *Physiology (Bethesda)*. 2020;35(4):244–260.
26. Nakada-Tsukui K, Nozaki T. Immune response of amebiasis and immune evasion by *Entamoeba histolytica*. *Front Immunol*. 2016;7:175.
27. Ghosh S, Padalia J, Moonah S. Tissue destruction caused by *Entamoeba histolytica* parasite: cell death, inflammation, invasion, and the gut microbiome. *Curr Clin Microbiol Rep*. 2019;6(1):51–57.
28. Allain T, Buret AG. Pathogenesis and post-infectious complications in giardiasis. *Adv Parasitol*. 2020;107:173–199.
29. Lopez-Romero G, Quintero J, Astiazaran-Garcia H, Velazquez C. Host defences against *Giardia lamblia*. *Parasite Immunol*. 2015;37(8):394–406.
30. Saghaug CS, Sornes S, Peirasmaki D, et al. Human memory CD4$^+$ T cell immune responses against *Giardia lamblia*. *Clin Vaccine Immunol*. 2016;23(1):11–18.
31. Kothavade RJ. Challenges in understanding the immunopathogenesis of *Cryptosporidium* infections in humans. *Eur J Clin Microbiol Infect Dis*. 2011;30(12):1461–1472.
32. Pantenburg B, Dann SM, Wang HC, et al. Intestinal immune response to human *Cryptosporidium* sp. infection. *Infect Immun*. 2008;76(1):23–29.
33. Thurman AR, Doncel GF. Innate immunity and inflammatory response to *Trichomonas vaginalis* and bacterial vaginosis: relationship to HIV acquisition. *Am J Reprod Immunol*. 2011;65(2):89–98.
34. Nemati M, Malla N, Yadav M, et al. Humoral and T cell-mediated immune response against trichomoniasis. *Parasite Immunol*. 2018;40(3).
35. Petrin D, Delgaty K, Bhatt R, Garber G. Clinical and microbiological aspects of *Trichomonas vaginalis*. *Clin Microbiol Rev*. 1998;11(2):300–317.

30

Host Defenses to Helminth Infection

Subash Babu and Thomas B. Nutman

Parasitical helminths are complex eukaryotic organisms, characterized by their ability to maintain long-standing, chronic infections in human hosts, sometimes lasting decades. Hence, parasitical helminths are a major healthcare problem worldwide, infecting more than 1.5 billion people, mostly in resource-constrained countries (Fig. 30.1). Common helminth infections include those with intestinal helminths, and filarial and schistosome infections are a major medical, social, and economic burden to the countries in which these infections are endemic. Chemotherapy, although highly successful in some areas, still suffers from the disadvantages of the length of treatment, the logistics involved in the distribution of drugs, and, in some cases, the emergence of drug resistance. Vector control measures are at best an adjunct measure in the control of helminth infections but also suffer from the same social, logistic, and economic obstacles as those for mass chemotherapy. Therefore, the study of the immune responses to helminth infections attains great importance both in terms of understanding the parasite strategies involved in establishing chronic infection and in the delineation of a successful host immune response to develop protective vaccines against infection.

SPECTRUM OF HOST–PARASITE INTERACTIONS

Helminths have characteristically complex life cycles, with many developmental stages.[1] Transmission to humans occurs by ingestion of eggs or larvae, penetration of intact skin by larvae, or inoculation of larvae by biting insects. Thus, the host is exposed during the course of a single infection to multiple life-cycle stages of the parasites, each stage with a shared as well as a unique antigenic repertoire. For example, *Schistosoma mansoni* infection begins with penetration of the skin of humans exposed to infested waters by the free-swimming cercariae, which then develop into tissue-dwelling schistosomula. In the liver and mesenteric veins, schistosomula differentiate into sexually dimorphic adult worms, which then mate, and the resultant eggs produced migrate through tissues into the lumen of the intestine or bladder for environmental release. Similarly, in lymphatic filarial infection, the host is exposed to infective-stage larvae in skin, lymph nodes, and lymphatics; to adult worms in lymph nodes and lymphatics; and finally, to microfilariae in the peripheral circulation. Hence, the host–helminth interaction is complex not only because of the multiple life-cycle stages of the parasite but also because of the tissue tropism of the different stages.

Antigenic differences among the life-cycle stages can lead to distinct immune responses that evolve differentially over the course of a helminth infection. In addition, depending on the location of the parasite, the responses are compartmentalized (intestinal mucosa and draining lymph nodes in intestinal nematode infection, or skin/subcutaneous tissue and draining lymph nodes in onchocerciasis) or systemic (lymphatic filariasis or schistosomiasis). Moreover, the migration patterns of the parasite might elicit varied cutaneous, pulmonary, and intestinal inflammatory pathologies, as seen, for example, in *Ascaris* or *Strongyloides* infection during their migratory phase. This is further complicated by the fact that human hosts are often exposed to multiple life-cycle stages of the parasite at the same time. Thus, a patient with chronic infection with lymphatic filariasis harboring adult worms and microfilariae might be exposed to insect bites, thereby transmitting the infective-stage parasite. The immune response that ensues will not only be a reaction to the invading organism but will also bear an imprint of the previous exposures and the concurrent infection.

KEY CONCEPTS

Helminth Infection

- Divided into nematodes, trematodes, and cestodes.
- Produce chronic infections that can persist for decades.
- Characteristically cause morbidity rather than mortality.
- Multicellular parasites that do not multiply in the definitive host but can reproduce sexually to produce larval stages that ensure continued transmission.

Helminth infections can elicit a spectrum of clinical manifestations mirroring diversity in host immune responses.[2] For example, in lymphatic filariasis, most infected individuals remain clinically asymptomatic despite harboring significant worm burdens; this is thought to reflect the induction of parasite-specific tolerance in the immune system. Others exhibit acute manifestations, including fever and lymphadenopathy, and this is thought to reflect inflammatory processes induced by incoming larvae, dying worms, or superadded infections. Individuals who mount a strong but inappropriate immune response end up with lymphatic damage and subsequent immune-mediated pathology—hydrocele and elephantiasis. Finally, a group of infected individuals mount exuberant immune responses that often result in unusual pathology, such as tropical pulmonary eosinophilia. Thus, the clinical manifestations of lymphatic filariasis exemplify the spectrum of host–parasite interactions that occur during helminth infections (Fig. 30.2). Helminths also cause disease by a variety of mechanisms, including mechanical effects such as intestinal obstruction

(e.g., ascariasis), invasion of host cells or tissue with damage or loss of function (e.g., trichinellosis), or competition for nutrients (e.g., vitamin B_{12} deficiency from fish tapeworm infection).

Another hallmark of all helminth infections is their chronic nature, with many helminths surviving in the host for decades. For example, adult schistosomes and filariae may survive in host tissues for as long as 30 years, producing eggs and larval stages throughout this time. Similarly, *Strongyloides stercoralis*, with its ability to "autoinfect," can maintain its life cycle for decades. Chronic infections certainly reflect an adaptation that leads to "parasitism," as mortality induced in the host would prevent parasite transmission if the host died before larval release or egg production could occur. In addition to the long-lived nature of the infection, helminths appear to reflect a harmonious host–parasite interface such that relatively asymptomatic carriers are available as reservoirs for ongoing transmission. Of course, failure to establish this harmonious coexistence does occur, leading to pathological conditions exemplified by cirrhosis and portal hypertension in schistosomiasis and elephantiasis associated with lymphatic filariasis.

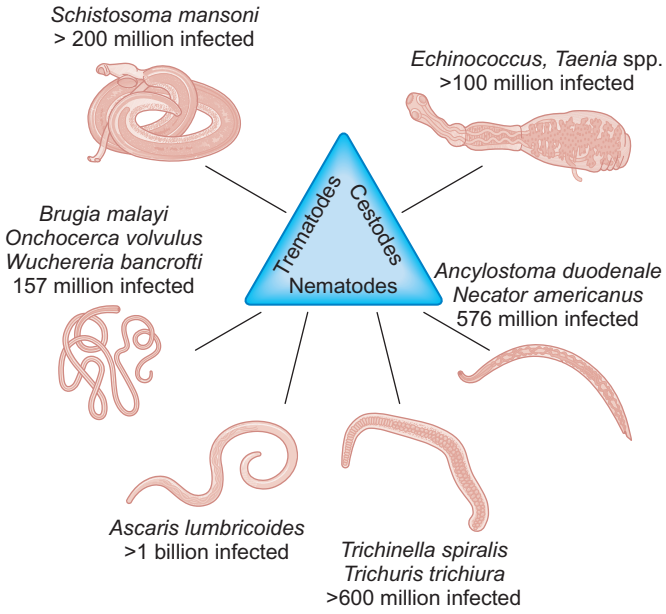


FIG. 30.1 Common and medically relevant helminth infections and their global prevalence.

PROTOTYPICAL HOST RESPONSES TO HELMINTHS

The canonical host immune response to all helminths is of the T-helper 2 (Th2) type and involves the production of cytokines, interleukin (IL)-4, IL-5, IL-9, IL-10, and IL-13; the antibody isotypes immunoglobulin G1 (IgG1), IgG4, and IgE; and expanded populations of eosinophils, basophils, mast cells, type 2 innate lymphoid cells (ILC2), and alternatively activated macrophages (AAMs).[3] However, it is also being increasingly recognized that while the predominant response is Th2 in nature, a large regulatory component involving both regulatory cytokines and cells is also part of this repertoire.[4] The Th2 response induced by helminth parasites is quite stereotypical, but its initiation, progression, and culmination requires interaction with many

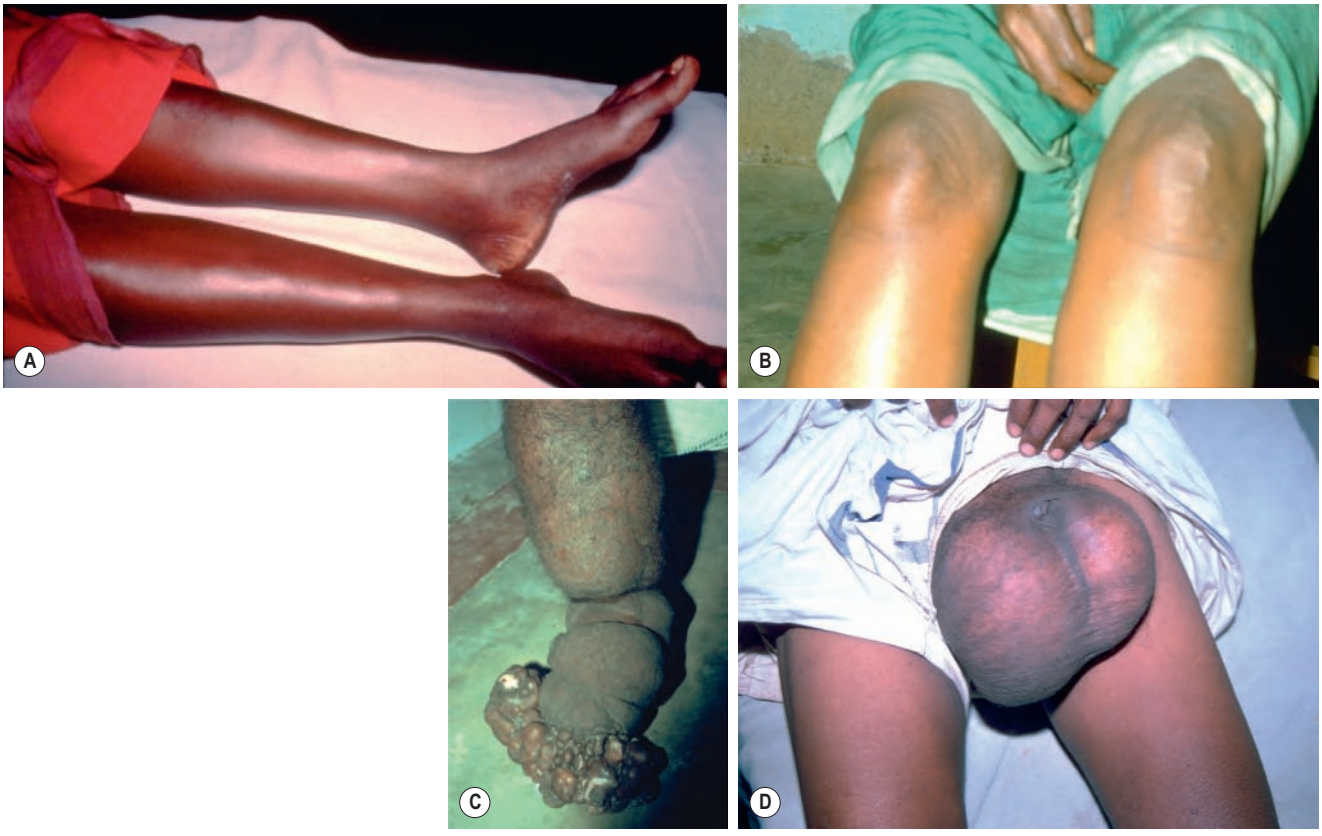


FIG. 30.2 The clinical manifestations of lymphatic filariasis, including (A) mild lymphedema, (B) severe lymphedema, (C) elephantiasis, and (D) hydrocele.

different cell types, most notably *(i)* epithelial/stromal cells, *(ii)* innate lymphoid cells (ILCs), *(iii)* dendritic cells (DCs) and macrophages, *(iv)* T cells, *(v)* B cells, *(vi)* eosinophils, *(vii)* mast cells/basophils, and *(viii)* neutrophils. The Th2 response plays a role in both resistance and tolerance phenotypes in response to helminth infection.[5] Resistance mechanisms promote resistance to helminth infections, while tolerance mechanisms (such as wound healing) operate to reduce the effect of helminth infections on host immunity without causing pathology. In addition, the host–helminth interactions can lead to a variety of modulated immune responses that are mediated largely by the induction of regulatory T cells (Treg) and AAMs (Fig. 30.3).

Helminths and Epithelial Cells

Epithelial cells are the first barrier layer exposed to or breached by most helminths, and the capacity of these cells to respond by initiating an "alarm" response has recently been recognized.[6] These epithelial cells mount a prototypical response comprising of chemokines and cytokines, such as IL-25, IL-33, and thymic stromal lymphopoietin (TSLP), as well as alarmins, such as uric acid, ATB, HMGB1, and S100 proteins.[7] These signals program dendritic cells (DCs) to mount Th2 cell–mediated immunity and in doing so boost ILC2, basophil, and mast cell function. IL-25 and IL-33 are necessary for protective immunity to a variety of helminth infections by inducing the production of Th2 cytokines, while TSLP is important for promotion of Th2 cell differentiation. Epithelial cells produce chemokines, including CCL17 and CCL22 (acting on ILC2, basophils, Th2 cells, and Tregs), and eotaxins, such as CCL11, CCL24, and CCL26 (acting on eosinophils and Th2 cells). They also produce prostaglandin D_2 (PGD_2), which acts on the CRTH2 receptor to recruit ILC2, basophils, mast cells, and Th2 cells. More recently, tuft cells, a specialized secretory cell type of the intestinal epithelium, have been identified as a major player in anthelmintic immunity. Tuft cells are the sole intestinal source of IL-25, which further promotes IL-33 production.[8] Moreover, tuft cells exhibit a requirement for signaling via chemosensory receptors, raising the possibility that they use chemosensation to recognize helminths. In addition, epithelial cells in the intestine, for instance, are in constant contact with both beneficial and pathogenic bacteria and hence ideally located for immunological surveillance of the intestinal lumen. This recognition of signals by intestinal epithelial cells is essential to mucosal homeostasis, implicating these cells as central modulators of inflammatory responses. Finally, the production of mucus and mucus-associated bioactive molecules (Mucin5AC, trefoil factor-2, and resistin-like molecule-β [RELM-β]) are important in promoting protection against intestinal helminth infection.

Helminths and Innate Lymphoid Cells

The ILC family includes ILC1, which predominantly express IFN-γ; ILC2, which predominantly express IL-5 and IL-13; and ILC3, which predominantly express IL-22 and/or IL-17.[9] ILC2 are defined by their expression of the IL-33 receptor (IL-33R) and the transcriptional regulators, Id2, RORα, GATA-3, and Bcl11b. Unlike T cells, ILC2 rely on cytokines to drive activation rather than on cognate interactions mediated by antigen-specific receptors. ILC2 are a critical innate source of type 2 cytokines, including moderately large quantities of IL-5 and IL-13, but also of IL-4, IL-9, granulocyte macrophage–colony-stimulating factor (GM-CSF), and amphiregulin. These cytokines potently induce eosinophilia, mucus production from goblet

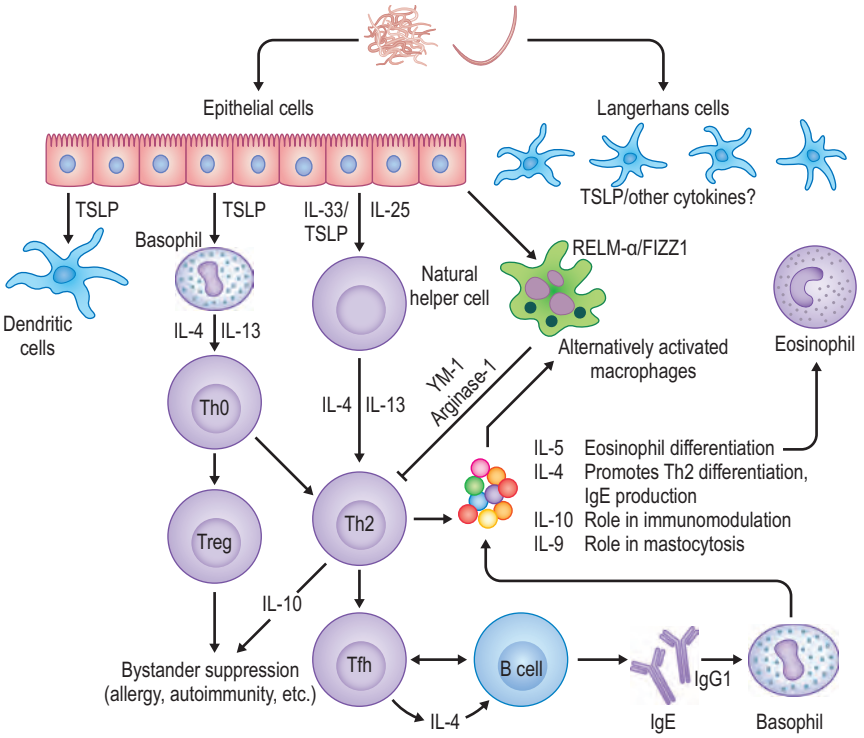


FIG. 30.3 Regulation of the T-cell response in helminth infection. *Chi3l*, Chitinase 3 like protein; *IDO*, indoleamine 2,3-dioxygenase; *Ig*, immunoglobulin; *IgE*, Immunoglobulin E; *IgG1*, Immunoglobulin G1; *IL*, interleukin; *RELM-α*, resistin-like molecule-α; *Tfh*, T-follicular helper cell; *Th0*, precursor T-helper cell; *Th2*, T-helper 2 cell; *Treg*, regulatory T cell; *TSLP*, thymic stromal lymphopoietin; *Relm(alpha)*, resistin-like molecule alpha, synonymous with *FIZZ1*, inflammatory zone 1.

cells, activation of AAM, muscle contractility, mast cell proliferation, and tissue repair.[10] IL-25 and IL-33 play crucial roles in promoting ILC2 responses in the lung and intestine, while TSLP is necessary for ILC2 responses in the skin. In addition, the transcription factors GATA-3 and RORα, as well as Notch signaling, have been found to be essential for the development of ILC2. Also, ILC2 are regulated by the nervous system through neurotransmitters and neuropeptides (including stimulation by neuromedin U, vasoactive intestinal peptide, and calcitonin gene-related peptide, and inhibition by catecholamines and acetylcholine receptor agonists). Although the function of ILC2 and Th2 cells appear to be largely overlapping, the kinetic differences in the ability to secrete cytokines rapidly and in profuse amounts allows for a coordinated interaction between the two cell types. Moreover, ILC2 can directly regulate the activation of T cells through their expression of major histocompatibility complex (MHC) class II molecules and the accessory molecules, CD80 and CD86, albeit less efficiently compared with DCs. Finally, recent reports have linked ILC2 with metabolic homeostasis, obesity, and dietary stress, providing an indirect link by which helminths might modulate host metabolic function. Also, ILC2 can interact with neuronal cells, which produce a chemical called neuromedin U in response to helminths, and produce IL-5 and IL-13 directly.

Helminths and Dendritic Cells

DCs are professional antigen-presenting cells (APCs) that play an essential role in presenting antigen to T cells to initiate immune responses. Although the role of DCs in inducing Th1, Th17, and Treg responses is well established, their role in inducing Th2 responses has remained relatively unclear.[11] Nevertheless, a series of studies have shown that DCs are required for optimal Th2 responses in vivo. Thus in vivo depletion of DCs has been shown to inhibit the induction of Th2 responses to *S. mansoni* or *Heligmosomoides polygyrus bakeri*. Helminth products can prime DCs for the induction of Th2 responses by interaction with pattern recognition receptors (PRRs), such as Toll-like receptors (TLRs) and C-type lectin receptors (CLRs). This interaction, which depends on TLR and CLR signaling, can promote Th2 responses by suppressing antigen presentation, costimulation, and/or expression of Th1-promoting cytokines by directly interfering with these pathways. DCs that drive Th2 responses typically exhibit specialized markers, such as CD301b, PDL2, and CD11b, and several receptors for the Th2-related cytokines IL-4R, IL-13R, IL-25R, TSLP-R, and IL-33R. These DCs also express the transcription factors IRF4 and KLF4. Additionally, the extracellular signal-regulated kinase (ERK) and signal transducer and activator of transcription 4 (STAT4) pathway upregulates the costimulatory molecules CD40, OX40L, and Jagged. Activation of the major transcription factors, IRF4 and KLF4, inhibits IL-12 production and increased IL-10 secretion. In addition, DCs expressing FcεRIII can induce murine IgG1-related Th2 responses. These factors typically act individually or in concert to orchestrate Th2 responses in helminth infections. Although Th2 cell–mediated immunity requires IRF4-dependent $CD301b^{+}CD11b^{+}$ DCs in the mouse, Langerhans cells are the predominant inducers of Th2 cells ex vivo in humans. The modulation of DC function by helminth antigens appears to be generalizable and has been shown to impair their ability to respond to other infectious stimuli (e.g., *Mycobacterium tuberculosis*).

Helminths and Macrophages

Macrophages are the other important class of APCs that can serve as protective effector cells in bacterial and protozoan infections by their production of nitric oxide and other mediators. Helminth interaction with macrophages induces a population of cells preferentially expressing arginase instead of nitric oxide as a result of increased activation of *arginase-1* by IL-4 and IL-13.[12] These AAMs are characterized by their ability to upregulate arginase-1, chitinase 3-like proteins 3 and 4 (also known as Ym1 and Ym2, respectively), and RELM-α.[5] These AAMs are known to be important in wound healing and have been postulated to play a potential role in repairing wound damage that occurs during tissue migration of helminth parasites. In fact, there appear to be two distinct populations of AAMs, one derived from blood and functioning in an immune regulatory role and the other derived from tissue-resident macrophages apparently responsible for much of the fibrosis seen in chronic helminth infections. Different tissue environments have specific amplification mechanisms for AAM proliferation (e.g., surfactant protein A in the lung and complement component, C1q in the peritoneal cavity and liver). By virtue of the expression of regulatory molecules, such as IL-10, TGF-β, and PDL2, these AAMs may have a predominantly regulatory role in helminth infections. These antiinflammatory macrophages function through arginase-1, PDL2, triggering receptor expressed on myeloid cells 2 (TREM2) and RELM-α to inhibit classic macrophage inflammation and recruitment and T-cell responses. RELM-α–expressing macrophages can also reduce parasite burdens and protect lung damage during helminth infections. Similarly, macrophage-derived human resistin is induced by helminth infection and promotes inflammatory responses and increased susceptibility. AAMs are also a source of retinoic acid and can expand thymus-derived regulatory T cells or inducible regulatory T cells at the site of infection.

Helminths and T Cells

Typically, infections with helminths induce a robust Th2 response manifested by enhanced expression of IL-4, IL-5, IL-9, IL-10, and IL-13 in response to live parasites, parasite antigens, or mitogens.[7] The central player in Th2 immunity is certainly the CD4 Th2 cell.[13] It is clear that IL-4Rα, a component of both the IL-4 and IL-13 receptors, is at the epicenter of Th2 immunity, since IL-4 and IL-13, together or individually, are absolutely critical for resistance to most helminth parasites. Recent work has reported that the Th2 cell population is heterogeneous, containing both $IL\text{-}5^{+}$ and $IL\text{-}5^{-}$ Th2 cells that express IL-4 and IL-13. In addition, IL-4 and IL-13 production is spatially separated, with IL-13 expression being marked in tissues and IL-4 expression being pronounced in the lymph nodes within the Th2 cell compartment. More recently, it has been reported that this heterogeneity extends to involve the expansion of a $CD161^{+}$, $CD27^{-}$ Th2 cell compartment, which diminishes following anthelmintic treatment.[14] The major transcription and accessory factors involved in the differentiation of Th2 cells are GATA3, STAT6, STAT5, STAT3, Gfi-1, c-Maf, and IRF4. Interestingly, chronic helminth infections are associated with downmodulation of parasite antigen-specific proliferative responses as well as IFN-γ and IL-2 production, but with intact IL-4 responses to parasite antigens and global downregulation of both Th1 and Th2 responses to live parasites. Finally, the receptor NLRP3 has been shown to be a key transcription factor in Th2 differentiation.

Central memory T cells are typically associated with resistance to helminth infections and are often found at lower frequencies in helminth-infected individuals.[15] This is associated with lower systemic levels of IL-7 and IL-15. In addition, helminth infections are also associated with the expansion of a new type of memory CD8 T cells called virtual memory or bystander memory CD8 T cells, whose induction is dependent on IL-4. Although the role of tissue-resident memory T (T_{RM}) cells is well established in viral and bacterial infections, very little is known about the role of these cells in helminth infections. However, it has been shown that tissue-resident Th2 cells can exert innate (TCR-independent and IL-33–dependent) functions upon appropriate stimuli and confer protection against helminth infection. In addition, although multifunctionality (ability to produce two or more cytokines) has not been well described in the Th2 cell compartment, helminth infections are known to be associated with an antigen-dependent enhancement of mono- and dual-functional Th2 cells and its reversal after treatment. Of interest, a stable subset of parasite induced T-bet$^+$, GATA-3$^+$, Th1/Th2 hybrid T cells has been described to develop directly from naïve precursors and to play a role in limiting pathological inflammation in animal models of helminth infection.

Recently, a new subset of T cells expressing IL-9 and IL-10, but not IL-4 (and therefore different from Th2 cells), has been described in allergic inflammation and in response to intestinal parasites.[16] These cells appear to be under the control of TGF-β and IL-4 and are dependent on STAT6, GATA-3, IRF4, and PU.1. Th9 cells have been recently shown to be associated with host protection in *Nippostrongylus brasiliensis* and *Trichuris muris* infection. Finally, Th9 cells have also been shown to be predominantly associated with lymphatic pathology in filariasis. T-follicular helper (Tfh) cells are a subset of CD4 T cells that migrate to B-cell follicles after activation and promote germinal center formation and B-cell isotype switching.[17] These cells, which form an independent lineage of CD4 T cells, have been recently identified to be the predominant IL-4–producing T cells early in helminth infection. In addition, Tfh are major producers of IL-21, a cytokine that plays a crucial role in supporting polarized Th2 responses in vivo.

Th17 cells, another subset of CD4 T cells, express the prototypical cytokine—IL-17.[16] In terms of helminth infections, the role of Th17 cells has been primarily studied in animal models of *S. mansoni*, where it has been strongly associated with infection-induced, immune-mediated pathology. More recently, it has also been demonstrated in human infections, in which children with *Schistosoma haematobium*–associated pathology have higher Th17 responses compared with those who are pathology-free. Similarly, a strong association of Th17 responses with pathological responses has also been demonstrated in lymphatic filariasis. Finally, Th22 cells are yet another subset of CD4 T cells that typically secrete IL-22.[16] To date, only a few studies have examined the role of Th22 cells in helminth infections. IL-22 was shown to be induced in the intestinal mucosa after infection with *Trichuris trichiura* or *Necator americanus* in humans, whereas the frequency of Th22 cells was shown to be higher in individuals with filarial infection compared with endemic healthy controls.

Tc2 cells, which are CD8 T cells expressing IL-4 and IL-13 are also expanded in helminth infections, which in turn is regulated by the co-inhibitory molecules, CTLA-4 and PD-1. Finally, γδ T cells that express Th2 cytokines have also been described in helminth infections.

Helminths and B Cells

Helminth interactions with B cells occur both at the B-cell cytokine level and at the level of antibody production.[18] Interactions at the cellular level primarily result in B-cell activation and cytokine production, most notably by the induction of IL-10. B cells have been shown to be important for the Th2 responses to certain helminths, with IL-2–producing B cells supporting optimal development of effector and memory Th2 cells and LTα1β2-expressing B cells supporting the recruitment of Th2 promoting DCs. Immune regulation by B cells has also been recognized in schistosome infection, where B-cell deficiency leads to enhanced Th2 cell–dependent immunopathology. However, it is at the level of antibody production that B cells play a profound role in helminth infections. Susceptibility to secondary infection is increased in the absence of B cells in infection with *Litomosoides sigmodontis, S. mansoni, T. muris*, and *H. polygyrus bakeri*. IgG is reported as an antibody isotype that is important for protection against intestinal helminths, and IgM (typically produced in a T cell–independent manner) has been linked to timely elimination of filarial parasites. One of the most consistent findings in helminth infections, both in mice and humans, is the elevated level of IgE that is observed after exposure to helminths.[19] Most of the IgE produced is not antigen-specific, perhaps representing nonspecific potentiation of IgE-producing B cells or deregulation of a normally well-controlled immune response. Interestingly, these IgE antibodies persist many years after the infection has been treated, indicating the presence of long-lived memory B cells or plasma cells in helminth infections. IgE production both in mice and humans is absolutely dependent on IL-4 or IL-13. Recent findings have shown that IL-4–expressing Tfh cells induce low-affinity IgE production, while IL-13–expressing Tfh cells induce high-affinity IgE production. Other isotypes that are commonly elevated in humans with chronic helminth infection are IgG4 and IgG1, the former being most dependent on both IL-4 and IL-10.

Recent studies have highlighted the role of regulatory B cells in suppression of immune responses to helminth parasites.[20] This B-cell function involves the secretion of IL-10 and IL-35 and is similar to the regulatory activity of B cells in autoimmune diseases. These cells mostly express CD11c and T-bet. B1 cells are a subset of B cells that produce natural antibodies. B1 cells have been demonstrated to be necessary for immunity to helminth infections in animal models. Moreover, B1 cells are known to produce polyclonal IgE during helminth infections.

Helminths and Eosinophils

Blood and tissue eosinophilia is characteristic of helminth infection and is mediated by IL-5 (probably in concert with IL-3 and GM-CSF).[21] Recruitment of eosinophils to the site of infection occurs very early in experimental helminth infection—as early as 24 hours after exposure. Kinetics of blood eosinophilia in humans is harder to determine but is postulated to occur as early as 2 to 3 weeks after infection, as demonstrated in experimental infections of volunteers. Both basal eosinophil levels and tissue accumulation during helminth infection appear to be under the influence of ILC2. Apart from the rapid kinetics of recruitment, eosinophils in blood and tissue also exhibit morphological and functional changes attributable to eosinophil activation. Eosinophils possess a range of immunomodulatory factors that are released upon cell activation, including cytokines, growth factors, and chemokines. Eosinophil granular proteins, such as eosinophil cationic protein, eosinophil-derived neurotoxin,

eosinophil peroxidase (EPO), and eosinophil major basic protein (MBP), are elevated in helminth infections and diminish following anthelmintic therapy. In general, eosinophils are able to kill larval stages of several helminth species in vitro, but their role in helminth control in vivo is not clear. By using eosinophil-depleted mice, it was shown that eosinophils were not required for primary or secondary immunity to a variety of helminth infections. Unlike T and B cells, eosinophils can rapidly release cytokines within minutes in response to stimulation since most of the cytokines are stored in a preformed fashion in secretory vesicles. Moreover, eosinophils can participate in the regulation of IgE and goblet cell mucus production; they also serve as effector cells in protective immune responses and as regulatory cells influencing both innate and adaptive immunity in helminth infections. The most important chemokine for eosinophil trafficking to tissues is Eotaxin-1, and to a lesser extent Eotaxin-2, while IL-5 enhances the development and maturation of eosinophils.

Helminths and Basophils/Mast Cells

Basophils are an important component of the immune response to helminth infections.[22] Basophils are capable of secreting a variety of mediators, including histamines, cytokines, chemokines, and lipid mediators that promote Th2 responses. Basophils in humans and mice also readily generate large quantities of IL-4 in IgE-dependent and IgE-independent manners and can also secrete IL-25 and TSLP. Basophils appear to play an important role in protective immunity to secondary infection (similar to eosinophils) with *N. brasiliensis, H. polygyrus bakeri*, and *L. sigmodontis;* they also play an active role in resistance to primary infection (through secretion of IL-4 and IL-13) with *T. muris* and *Trichinella spiralis*. In addition, basophils have been shown to be critical APCs (by acquiring MHC class II molecules from DCs by a process called trogocytosis) for driving Th2 cell differentiation in different models of helminth infection and to prime ILC2 for neuropeptide-mediated inhibition.

Mast cells may contribute to inflammatory reactions directed against invasive helminth parasites.[19] These cells express high-affinity Fcε receptors that are sensitized with parasite antigen–specific IgE and can be triggered by parasite antigens. It has been postulated that cytokines and other mediators released by sensitized mast cells contribute to *(i)* the recruitment and activation of effector eosinophils; *(ii)* increased local concentrations of antibody and complement; and *(iii)* enhanced mucus hypersecretion and increased peristalsis of the gastrointestinal (GI) tract, which plays an important role in resistance to certain GI nematode infections. More recently, a role for mast cells (in an IgE-independent manner) in mediating the secretion of epithelial-derived cytokines (IL-25, IL-33, and TSLP) and optimal migration of DCs was shown in *H. polygyrus bakeri* infection. Mast cells are also crucial for induction of ILC2 by production of IL-33 and for elimination of helminth infection by promoting ILC2 expansion and goblet cell hyperplasia. Finally, mast cell granular proteins, such as mast cell tryptase and carboxypeptidase A3, are elevated in helminth infections and diminish following anthelmintic therapy.

Helminths and Neutrophils

Although neutrophils are typically considered more important in bacterial and fungal infections, a number of studies have revealed that neutrophils can act in conjunction with macrophages to contain or kill helminth parasites.[7] Thus, neutrophils are major components of the granulomas forming around filarial parasites and the cysts containing larvae of intestinal helminths. Neutrophils have been demonstrated to collaborate with macrophages in the immobilization and killing of *S. stercoralis* larvae in a process that is complement-dependent and involving neutrophil extracellular traps (NETs). Similarly, neutrophils contribute in the early antifilarial response through oxidative burst, degranulation, and NETosis and protect against infective larvae in skin. Indeed, NETs can mediate direct killing of hookworms and hookworms in turn secrete a deoxyribonuclease that degrades NETs to evade host immunity. A seminal study reported that neutrophils adopt an "N2" phenotype during experimental infection with *N. brasiliensis* in the lung and express the genes for IL-13, IL-33, RELM-α, and Ym1. These "N2" neutrophils can train macrophages to acquire a memory phenotype that protects against secondary infection. Neutrophils can rapidly upregulate the expression of Th2-related genes, including IL-13, IL-33, RELM-α, and chitinase-3-like. Finally, it was also shown that even during primary infection, the absence of neutrophils resulted in greater worm burdens because of lack of immunity in the lungs. Thus, neutrophils appear to play an unexpected role in immunity to helminths that certainly merits further investigation. Moreover, neutrophil granular proteins, such as neutrophil elastase and myeloperoxidase, are increased in helminth infections and diminish following anthelmintic treatment.

PROTECTIVE IMMUNITY AGAINST HELMINTHS

The mechanism of protective immunity to helminths is dependent on the location of the helminth infection.[7] Clearly, T cells are central to resistance against helminths.[23] For example, T cells are essential in mediating the expulsion of GI nematodes. Mice lacking T cells are defective in their ability to expel *T. muris*, but resistance can be reconstituted by transfer of T cells from normal mice. In addition, CD4 T cells from infected mice can transfer protective immunity to severe combined immunodeficiency (SCID) mice (lacking both B and T cells), indicating that CD4 T cells, not CD8 T cells, are important for protective immunity. Similarly, T cells were shown to be required for expulsion in *N. brasiliensis* infection. Both nude mice (lacking T cells alone) and SCID mice are susceptible to infection with Brugian parasites, whereas mice that lack either CD4 T cells or CD8 T cells are not. In schistosome-infected mice, T cells are essential in forming host-protective granulomas around the eggs deposited in the liver.

The role of cytokines in protective immunity has been extensively studied in murine models of both GI helminths and tissue-invasive helminths.[5] In general, type 2 (Th2) cytokines target epithelial cells, goblet cells, smooth muscle cells, and macrophages, which together coordinate parasite expulsion by increasing fluid and mucus production, encapsulation and barrier formation, epithelial cell turnover, smooth muscle cell contraction, and production of anthelmintic effector molecules, such as RELM-β. The cytokines involved in both responses are IL-4, IL-5, IL-9, and IL-13.[15] Most of the studies examining resistance to intestinal helminths involve four parasitical GI nematode infections of rodent models—*T. spiralis, H. polygyrus bakeri, N. brasiliensis*, and *T. muris*.[13] These studies show that *(i)* CD4 T cells are crucial for host protection; *(ii)* IL-4 is required for host protection and limiting host pathology; *(iii)* IL-13 can substitute for IL-4 in some but not all infections; *(iv)* IL-2 and

IFN-γ inhibit protective immunity; and *(v)* IL-4 and IL-13 have multiple effects on the immune system and gut physiology, leading generally to protection. Type 2 cytokines mobilize a broad range of downstream effector mechanisms. Epithelial cells in the gut, specifically tuft cells, promote goblet cell differentiation, enhancement of mucus secretion, and the production of RELM-β, an innate effector molecule with direct anthelmintic activity. Goblet cells can also secrete gel-forming mucins, which are major macromolecular components of the mucus barrier. Two of these mucins have been shown to be critical in resistance to intestinal nematode infection—Muc2 and Muc5AC. Another group of molecules that have been found important in immunity to helminths are the trefoil factors, especially trefoil factor 2. IL-4Rα activation also leads to increased intestinal smooth muscle hypercontractility and accelerated epithelial turnover to promote an effector response akin to an "epithelial escalator," which, together with epithelial secretions, helps expel intestinal helminths. This smooth muscle activity is now known to occur in response to neurotransmitters or nerve stimuli, with serotonin and M3 muscarinic acetylcholine receptors being involved in the process. Enteroendocrine cells are also known to contribute to increased intestinal motility for parasite expulsion. Mucosal mast cells release proteases that can degrade epithelial tight junctions, thereby increasing fluid flow as part of the "weep and sweep" response, with contribution from goblet cell production of mucus. AAMs in the gut can also entrap intestinal worms and cause death by compromising worm vitality.[23]

Although the role of Th2 cytokines in immunity to GI helminth infection is well defined, their role in protective immunity to tissue-invasive helminths is not as clear. In murine models of schistosomiasis, protective immune responses can be generated by vaccination with irradiated cercariae. This resistance is dependent on a Th1-mediated immune response consisting of macrophages and endothelial cells activated by IFN-γ and TNF-α, producing nitric oxide and Th1-associated antibodies—IgG2a and IgG2b. In contrast, studies in rats and epidemiological studies in humans suggest Th2-mediated effector mechanisms involving IgA and IgE antibodies, as well as eosinophils are thought to be central to protective immunity. Protective immunity to filarial infections in mice is dependent primarily on Th2 responses in mice. Thus, mice lacking IL-4, IL-4R, or Stat6 are all susceptible to infection with *Brugian* parasites.

In tissue-invasive helminth infections, effector mechanisms involve multiple innate immune cells, with antibodies acting as initiators of immunity by activating Fc-receptor–expressing cells. Basophils, by their ability to produce high levels of IL-4, act as effectors to promote helminth killing in secondary or challenge infections. For example, basophils are important in immunity to the skin-invasive stages of intestinal helminths and, through IL-4 release, promote the activation of macrophages that trap larvae in an arginase-dependent manner. Although eosinophils are crucial players in producing IL-4 early in infection, they are also amplifiers of immune responses, rather than being critical mediators of primary immunity, since depletion of eosinophils does not alter the course of many helminth infections in murine models. The mechanism of protection mediated by eosinophils is thought to be by antibody-dependent, cell-mediated cytotoxicity, as observed in *S. mansoni* studies in vitro or through release of eosinophil granule contents. In addition, eosinophils play an important role in the protective immunity against primary infection with *Brugia malayi* and/or secondary infection with either *T. spiralis* or *N. brasiliensis*. The two most abundant granular proteins, MBP and EPO, are required for protective immunity against *S. stercoralis* and *L. sigmodontis*. Similarly, neutrophils can attack helminth larvae in response to IL-4 and IL-5, but their importance in resistance to primary helminth infections is not known.

Antibodies play a major role in mediating protection to some, but not all, helminth infections.[24] Antibody-mediated passive immunity has been demonstrated in animal models for *Ancylostoma caninum*, *Schistosoma* spp., *Taenia* spp., *Ascaris suum, Streptococcus ratti, T. muris, N. brasiliensis*, and *H. polygyrus bakeri*. Passive immunity has also been shown by using IgG monoclonal antibodies (mAbs) specific for *Fasciola hepatica* and *S. mansoni;* IgM (mAbs) specific for *B. malayi;* and IgG or IgA mAbs specific for *T. spiralis*. Using genetically-manipulated mouse models, IgM has been shown to be crucial for host protection against *B. malayi* and to *S. stercoralis*. B1B cells, a subset of B cells that secrete IgM, appear to be an important component of this protective axis. Finally, antibodies have the capacity to trap tissue-migrating helminth larvae and prevent tissue damage by driving an IL-4Rα–independent alternative differentiation of macrophages, in a process dependent on CD11b and FcγR1.

KEY CONCEPTS

Helminth-Induced Immune Responses

- Characterized by immunoglobulin E (IgE) and IgG4 antibody production, tissue and peripheral blood eosinophilia, mast cell involvement, innate lymphoid cell type 2 and Th2 cell expansion, and production of type 2 cytokines.
- Implicated both in pathogenesis of helminth infections and in mediating immunological protection.
- In mucosal immunity to helminths, T-helper 2 (Th2) cell responses are initiated and sustained by innate populations (including tuft cells and innate lymphoid cells) through interleukin (IL)-25, IL-33, and thymic stromal lymphopoietin (TSLP).
- In tissues, helminths are acted upon by the host innate effectors, including macrophages, neutrophils, eosinophils, and basophils.
- Regulated by T cells and other cells producing IL-4, IL-5, IL-9, IL-10, and/or IL-13.
- Characterized by the induction of regulatory T cells (Tregs) that mediate downmodulation of immune responses to helminth infections and impact bystander phenomena, such as allergy and autoimmunity.

In terms of protection, a major mechanism appears to be the formation of multicellular, immune cell aggregates, called *granulomas*, around incoming infectious larvae or eggs.[25] In murine models of schistosomiasis and filariasis, granulomas are primarily composed of T cells (which help in the recruitment of other cell types and mediate alternative activation of macrophages), B cells (particularly the B1 subset), and macrophages and eosinophils. Although the exact mechanism by which granulomas mediate killing of the parasite remains unknown, it is clear that formation of these structures is an important host defense mechanism. In neurocysticercosis, the dying *Taenia solium* cysticercus is enclosed in a granuloma composed of eosinophils, multinucleated giant cells, epithelioid cells, macrophages, T cells, and collagen fibers. One cell type that can mediate effector functions within granulomas is the AAM, which is exemplified by the targeting of the glycan chitin that is frequently expressed by helminths but not by the host. The chitinase and

fizz family proteins (ChaFFs), which include chitinase and chitinase-like secreted proteins, are prime candidates for mediating host resistance. These proteins include acidic mammalian chitinase (AMCase) and the RELM family proteins and are capable of enzymatic activities that potentially damage certain helminths. AAMs isolated from granulomas express high levels of surface IgG1, IgG3, FcγRs, and CD11b. Eosinophils within granulomas have been described to function by degranulation-mediated damage of the helminths as well as by enhancing antibody-dependent cell-mediated cytotoxicity. Myelin basic protein and eosinophil-stimulator-protein-mediated lysis have also been described.

PATHOLOGY ASSOCIATED WITH IMMUNE RESPONSES IN PARASITIC HELMINTH INFECTION

Typically, pathological findings associated with each parasitic infection are different and relate to the presence of the parasites in host tissues, but there are pathological reactions that stem directly from the host response.

Immune Complexes

Immune complexes are potent mediators of localized inflammatory processes that form in many parasitic infections presumably as a result of the chronic low-dose antigen release seen in these infections. Circulating immune complexes have been identified in both experimental and human filarial and schistosomal infections. These have been shown to induce lymphatic inflammation and vasculitis in filarial infections as a result of their deposition. It has been demonstrated that FcγR-mediated cell activation is the dominant pathway of immune-complex corneal disease in animal models of onchocerciasis. In addition, a common manifestation of immune-complex–mediated pathology, immune-complex glomerulonephritis (ICGN), has been documented by renal biopsy in patients with schistosomiasis and filarial infections. Other manifestations of immune complex–mediated damage, such as reactive arthritis and dermatitis, have also been described in patients with helminth infections.

Autoantibodies and Molecular Mimicry

Autoantibodies have been implicated as causing disease in a variety of helminth infections, including filarial infections, schistosomiasis, and hookworm infection, and are thought to reflect a polyclonal B-cell expansion that often accompanies these infections.[26] Autoantibodies against nuclear material have been found in a vast majority of patients with chronic schistosomiasis, and antibodies against human calreticulin and defensin have been found in onchocerciasis. New data also suggests that Nodding syndrome (an epileptic disorder) is an autoimmune disorder based on findings of cross-reacting antibodies between neuronal structures and a protein present in *Onchocerca volvulus*, called leiomodin-1. Finally, infection with *Toxocara canis* can trigger systemic vasculitis by anti-DNA, anti-nuclear, and anti-C1q antibodies.

Granulomatous Reactions

Granuloma formation is the mainstay of the protective immune response to certain helminths, but it can also lead to deleterious effects in the form of pathology. Although granulomatous reactions occur in many helminth infections (e.g., toxocariasis, *Angiostrongylus* infections, and lymphatic filariasis), parasitical granulomata have been best studied in *S. mansoni* infections, where granulomatous and fibrosing reactions against tissue-trapped eggs is orchestrated by CD4 T cells, and the fibrosis that results from the cellular response is the principal cause of morbidity in infected individuals. The severity of the inflammatory process markedly varies both in humans and in experimental animal models, with severe pathology associated with Th1 and Th17 responses and milder pathology with Th2-dominant responses. Studies in murine models of granuloma formation have demonstrated the important roles of IL-13 and TNF.

Fibrosis

Fibrosis is commonly associated with chronic helminth infections that result in chronic inflammation and dysregulated wound healing.[27] These infections activate macrophages and fibroblasts, resulting in the production of TGF-β, platelet-derived growth factor, connective tissue growth factor, IL-1β, and other factors. Macrophages also promote inflammation by recruiting and activating monocytes and neutrophils, as well as activating CD4 T cells. Hepatic stellate cells produce copious amounts of collagen, which accumulates and leads to fibrosis. In addition, fibroblasts are stimulated to synthesize matrix metalloproteinases (MMPs) and tissue inhibitors of metalloproteinases (TIMPs), leading to extracellular matrix remodeling and fibrosis. Another consequence of chronic schistosomiasis is pulmonary arterial hypertension, which has been shown to be associated with IL-4– and IL-13–mediated type 2 inflammation resulting in TGF-β–induced pulmonary vascular disease. In the same manner, IL-10 and IL-12 are known to modulate IL-13–mediated fibrosis; in the combined absence of IL-10, IL-12, and IL-13Rα, IL-13–dependent fibrosis in chronic schistosomiasis proceeds rapidly to lethal cirrhosis. One of the primary sources of morbidity and mortality in *S. haematobium*-infected individuals is egg-induced urinary tract fibrosis. Infection with *Wuchereria bancrofti* (one of the causative agents of lymphatic filariasis) is associated with similar fibrotic reactions.

Toll-Like Receptors

Immunopathology in lymphatic filariasis is associated with the presence of an endosymbiotic, *Rickettsia*-like bacteria called *Wolbachia*. *Wolbachia* are known to stimulate immune cells through TLR2 and TLR4 and to release proinflammatory cytokines, as well as vascular endothelial growth factors (VEGFs), which might contribute to lymphatic pathology.[28] *Wolbachia*–TLR4 interaction has also been shown to be the major mechanism of corneal inflammation in onchocerciasis, and a TLR-signaling molecule, IL-1 receptor–associated kinase-2 (IRAK-2), regulates pathogenic Th17-cell development in *S. mansoni* infection. TLR upregulation might also underlie post-treatment adverse events in filarial infections.

Immediate Hypersensitivity Responses

Immediate hypersensitivity responses are associated with the early and/or acute phase of infections with invasive helminth parasites, such as *Ascaris*, hookworm, schistosomes, or filariae. Patients manifest symptoms suggestive of allergic reactivity, such as wheezing or urticaria. Furthermore, in clinical syndromes associated with *Loa loa* infection (with its angioedematous Calabar swellings), with tropical pulmonary eosinophilia, and with larva currens in strongyloidiasis, IgE-mediated reactions are thought to underlie these signs and symptoms. Anaphylaxis

is a severe, life-threatening, generalized or systemic hypersensitivity reaction, and is associated with IgE interaction with high-affinity IgE receptors on basophils and mast cells. The risk of anaphylaxis in individuals with helminth infections can vary, depending on the parasite, tending to occur more frequently with echinococcosis or after *Anisakis* infection, while being extremely rare in most other helminth infections.

Wound Healing

Recent studies have shown a close association of type 2 cytokine responses with many aspects of wound healing and repair.[27] It has been proposed that the type 2 cytokine response has evolved to not only mediate resistance to helminth infection but also to activate the wound healing apparatus to repair and reconstruct tissue, since tissue damage is intricately associated with helminth infections. Thus, AAMs are intimately involved in this process as they produce MMPs, arginase-1, insulin-like growth factor 1, VEGF, and TGF-β, which together promote myofibroblast activation, angiogenesis, epithelial cell turnover, and extracellular matrix deposition. Additional signals from neutrophils, as well as lung surfactant protein A or trefoil factor 2, can also contribute to tissue repair. In addition, an ortholog of the human protein granulin from *Opisthorchis viverrini* accelerates wound repair.

Lymphangiogenesis

The anatomical changes in the architecture of lymphatics, which range from lymphangiectasia and granulomatous responses to the development of collaterals, suggest that active lymphatic remodeling involving endothelial cell growth, migration, and proliferation is an important feature of early lymphatic filarial disease.[2] Live filarial parasites (and their excretory/secretory products) have been shown to induce activation, proliferation, and tube formation in lymphatic endothelial cells (LECs) and their differentiation into tubelike networks. This was found to be associated with significantly increased levels of MMPs and TIMPs. Recent studies have also implicated the VEGF family in lymphangiogenesis, with VEGF-C being associated with lymphedema and VEGF-A with hydrocele and single nucleotide polymorphisms in MMP-2, CEACAM-1, and VEGF-R3 with lymphedema. Finally, TLR-mediated events are considered to be the main drivers of this angiogenic/lymphangiogenic process in filarial disease.

Carcinogenesis

Infection with *O. viverrini, Clonorchis sinensis*, and *S. haematobium* are classified as group 1 biological carcinogens (i.e., definitive causes of cancer).[29] The former (liver fluke) is associated with cancer of the bile duct (cholangiocarcinoma) and cancer of the liver (hepatocarcinoma), and the latter is associated with squamous cell carcinoma, an especially aggressive type of bladder cancer, as well as urothelial carcinoma. The mechanisms of helminth-induced cancer include chronic inflammation, sustained cellular proliferation, modulation of the host immune system, reprogramming of glucose metabolism and redox signaling, induction of genomic instability and destabilization of tumor proteins, stimulation of angiogenesis, resistance to apoptosis, and activation of invasion and metastasis. Seven other helminth infections have been reported to be associated with cancer, although their role remains to be firmly established.

Epileptogenesis

Neurocysticercosis, caused by the larval form of *T. solium*, is the most common preventable risk factor for epilepsy worldwide and accounts for nearly 30% of all epilepsies in some endemic areas. The manifestations are variable, depending on the location, number, and size of the cysts in the central nervous system as well as the degree of accompanying inflammation, provoked by cyst degeneration, calcification, and/or perilesional edema. The development of epileptic symptoms results from a complex interplay between the anatomical location of the cyst, environmental factors, parasite factors, host genetics, and especially host immune responses.

Epidemiological evidence has emerged pointing to a link between epilepsy and areas hyperendemic for *O. volvulus* throughout tropical Africa.[30] A 2018 longitudinal study from Cameroon suggested that the intensity of infection with *O. volvulus* has a crucial part in inducing epilepsy. Case-control studies in onchocerciasis-endemic regions in the Democratic Republic of the Congo suggested that ivermectin intake prevents the development of epilepsy in children. Living close to a river containing breeding sites of vector flies that transmit *O. volvulus* is associated with an increased risk of developing epilepsy. Onchocerciasis has been linked to two poorly understood diseases of tropical Africa: Nakalanga syndrome and nodding syndrome. Nakalanga syndrome is a child-development disorder that causes growth failure, wasting, skeletal deformities, endocrine dysfunction, intellectual disability, and epilepsy. Nodding syndrome is characterized by epileptic seizures with repeated slow dorsoventral head movements. A recent study showed that onchocerciasis eradication could eliminate nodding and Nakalanga syndromes and reduce the burden of epilepsy. Finally, nodding syndrome has been shown to be due to an immune cross-reaction to antigens of *O. volvulus*, thus conclusively providing a link.

MECHANISMS OF EVASION AND IMMUNE REGULATION BY HELMINTH PARASITES

Helminths exert profound immunoregulatory effects on the host immune system with parasite antigen–specific immune suppression as well as more generalized levels of immune suppression. It has been shown that patients with schistosomiasis or filariasis have markedly diminished responses to parasite antigens and to some measurable attenuation in responses to bystander antigens and routine vaccinations. Thus, host immunosuppression is usually antigen-specific, whereas chronic infection can be associated with some spillover effects. Among the mechanisms utilized by parasites to avoid immune-mediated elimination are those of evasion—the use of sequestration, camouflage, and antigenic variation—and suppression, regulation, or blockade of immune effector pathways.

Parasite-Derived Factors

Parasite-derived products play a very important role in host immune evasion.[31,32] Parasite products, such as the schistosome-secreted proteins, alpha-1 and omega-1, promote Th2 differentiation. Alpha-1 (also known as IL-4–inducing principle of schistosome eggs [IPSE]), released by schistosome eggs, induces IL-4 release and degranulation by human and mouse basophils by cross-linking surface IgE. Omega-1 is a ribonuclease abundantly secreted by eggs, shown to condition DCs to drive Th2 polarization. Omega-1 binds to and is internalized by DCs in a mannose receptor–dependent process and then suppresses protein synthesis through degradation of messenger RNA (mRNA).

Phosphorylcholine (PC) is a small hapten-like moiety present in the excretory/secretory products of many helminths, and one particular PC-containing molecule, called ES-62, from filarial worms has been shown to have a wide variety of immunomodulatory properties. Thus ES-62 can inhibit the proliferation of CD4 T cells and conventional B cells, decrease IL-4 and IFN-γ production, promote proliferation and IL-10 production by B1B cells, modulate complement activation, and condition APCs to drive Th2 differentiation with concomitant inhibition of Th1 responses. ES-62 has also been shown to exhibit bystander anti-inflammatory activity in collagen-induced arthritis, rheumatoid arthritis, chemical contact sensitivity, lupus-associated atherosclerosis, ear inflammation, chronic asthma, and airway hyperreactivity. Helminths utilize glycans within glycoproteins and glycolipids, which mimic host glycans, to regulate host immune responses. In addition, these host-like helminth glycans can directly interact with host glycan–binding proteins, such as CLRs and galectins, to shape innate and adaptive immune responses. Similarly, helminth lipids have also been implicated in immune modulation; schistosome phosphatidylserine induces DCs to polarize IL-4–producing T cells, whereas schistosome lysophosphatidylserine induces DCs to induce IL-10–secreting Tregs.

Helminth parasites utilize mechanisms involving cytokine mimicry and interference to establish chronic infection. Thus, parasites produce cytokine- and chemokine-like molecules to interfere with the function of host innate immune products. The first helminth cytokines were found to be homologues of TGF-β expressed by *B. malayi*, and both schistosomes and filarial parasites express members of the TGF-β receptor family. *H. polygyrus* secretes a protein called *H. polygyrus* TGF-β mimic (Hp-TGM) that binds to the mammalian TGF-β complex and drives human and mouse Treg production. Similarly, *Echinococcus granulosus* expresses a TGF-β ligand, and thus all helminth groups might have the potential to exploit TGF-β–mediated immune suppression. Various helminths, including *B. malayi*, produce homologues of macrophage migration inhibitory factors (MIFs), which are known to activate an antiinflammatory pathway through SOCS-1, a molecule involved in cytokine signaling. *T. muris* is known to express a homologue of IFN-γ, which binds to the IFN-γ receptor in vitro and induces signaling. As *T. muris* is expelled by IL-4, secretion of an IFN-γ–like protein can prolong its survival. *H. polygyrus* also secretes a cytokine-binding protein called HpAR1, which inhibits the release of alarmins.

Similarly, helminth parasites utilize chemokine- or chemokine-receptor like proteins to evade protective immunity. *A. suum* is known to express a neutrophil chemoattractant with chemokine-binding properties. *S. mansoni* eggs secrete a protein (*S. mansoni* chemokine-binding protein [smCKBP]) that binds the chemokines CXCL8 and CCL3 and inhibits their interaction with host chemokine receptors and their biological activity, resulting in suppression of inflammation. Similarly, *B. malayi* (and all of the other filariae sequenced to date) have been shown to express galectins that can bind host immune cells in a carbohydrate-dependent manner. *O. viverrini* secretes a granulin (GRN)-like growth factor, Ov-GRN-1, which promotes wound healing and angiogenesis.

Helminths secrete two major classes of protease inhibitors called *cystatins* and *serpins*, each with proposed immunomodulatory roles. Cystatins inhibit cysteine proteases (cathepsins and aspartyl endopeptidases) required for antigen processing and presentation (through the MHC class II pathway) and therefore inhibit T-cell activation as well as secretion of inflammatory cytokines. They also elicit the regulatory cytokine IL-10, leading to direct impairment of T-cell proliferation. The serpins are serine protease inhibitors, which can cause specific inhibition of the neutrophil proteinases, cathepsin G and neutrophil elastase. Aspartic proteases from *Ascaris lumbricoides* have been shown to block efficient antigen processing that is dependent on proteolytic lysosomal enzymes.

Other parasite products mediate their effect by blocking effector functions, including recruitment and activation of inflammatory cells and limiting the destructive potential of activated granulocytes or macrophages in the local extracellular milieu. For example, the host chemoattractant platelet-activating factor (PAF) is inactivated by a complementary enzyme PAF hydrolase secreted by *N. brasiliensis*. Eotaxin-1, a potent eosinophil chemoattractant, is degraded by metalloproteases from hookworms. *A. caninum* secretes a protein called *neutrophil inhibitory factor*, which binds the integrins CD11b/CD18 and blocks adhesion of activated neutrophils to vascular endothelial cells and also the release of hydrogen peroxide (H_2O_2) from activated neutrophils. *N. americanus* ES products also bind to host NK cells and augment the secretion of IFN-γ, which might cross-regulate deleterious Th2 responses. Other modulators, such as prostaglandins, and other arachidonic acid family members, such as PGE_2 and PGD_2, are known to inhibit IL-12 production by DCs. Finally, helminths susceptible to oxidant-mediated killing express both secreted and membrane-associated enzymes, such as superoxide dismutase, glutathione-S-transferase, and glutathione peroxidase, molecules that are thought to play a significant role in assisting parasite survival in inflamed tissues. Recently, a family of helminth defense molecules secreted by parasitic helminths has been shown to exhibit biochemical and functional characteristics similar to human antimicrobial peptides. These molecules can modulate innate cell activation by classic TLR ligands, such as lipopolysaccharide.

The discovery that parasitic helminths secrete extracellular vesicles (EVs) has spurred a new paradigm in the discovery of helminth-derived therapeutics and anthelmintic vaccines.[33] EVs are a heterogeneous group of lipid-enclosed vesicles in the nano- to the micrometer size range. There is increasing evidence that helminth EVs are essential players in regulating host inflammation and immunity, and their application as anti-inflammatory therapeutics has been considered. Helminth EVs are actively internalized by host cells, providing a mechanism by which the parasites transfer genetic material to the host in a bid to actively manipulate host gene expression. Both platyhelminths and nematodes have been found to release EVs from the gastrodermis. In nematodes, EVs released in the intestines may be released into the host via the anterior or posterior openings. EVs of *B. malayi* have also been found to be secreted from the excretory/secretory pore. In the case of platyhelminths, EVs can be shed directly from the tegument itself into the surrounding environment. Recent attention has now turned to identifying and testing specific EV-associated antigens, which can be formulated into a vaccine that recapitulates the protective action of vaccination with whole EVs.

Host-Related Factors

Regulatory T and B Cells

Helminths actively induce Tregs either by directly secreting factors such as TGFβ mimic Hp-TGM or indirectly by interacting

with DCs and macrophages, which in turn induce Tregs.[34] Expansion of Tregs not only enhances parasite survival but also plays a role in preventing helminth-mediated pathology. In addition, Tregs generated during helminth infection have been associated with suppression of bystander immune responses. Helminth-induced Tregs are thought to act mostly via their expression of co-inhibitory receptors, CTLA-4 and PD-1 and to a lesser extent by their production of IL-10. Tregs play a vital role in limiting host pathology by downregulating harmful Th1/Th17 responses in filarial infection and schistosomiasis. Regulatory activity can also be demonstrated in other T-cell subsets, such as the Foxp3-IL-10^{+} Tr1 cells. Moreover, filarial infection is associated with an expansion of T cells expressing the IL-10 superfamily cytokine members (IL-19 and IL-24), and inhibition of these cytokines results in increased Th1 and Th2 responses. Both Tregs and Tr1 cells are associated with an isotype switch from IgE to IgG4 and a greater IgG4 to IgE ratio in helminth infections. IgG4 is a strongly antiinflammatory isotype that does not stimulate FcγR receptor activity.

A number of studies have recently reported that B cells might have an active regulatory role in helminth infections.[20] Helminth infections, and in particular infection with schistosomes such as *S. mansoni,* are well-known to induce regulatory B (Breg) cells. Regulatory B cells suppress proinflammatory immune responses via several mechanisms, of which the ones best described are the expression of the regulatory cytokine IL-10 and induction of regulatory T (Treg) cells. CD19^{+}CD24^{hi}CD38^{hi} regulatory B cells have been described in a variety of helminth infections.

Hyporesponsive T Cells

Effector T-cell responses can be turned off or modulated through a variety of mechanisms, including through cytotoxic T-lymphocyte antigen-4 (CTLA-4) and programmed death-1 (PD-1). Interestingly, increased expression of CTLA-4 and PD-1 has been demonstrated in filarial infections, and blocking of CTLA-4 can partially restore a degree of immunological responsiveness in cells from infected individuals. Moreover, T cells have decreased induction of *T-bet*, the Th1 master regulatory gene indicating a failure at the transcriptional level to differentiate into Th1 cells. Finally, T cells from individuals with filarial infection exhibit classic signs of anergy, including diminished T-cell proliferation to parasite antigens, lack of IL-2 production, and increased expression of E3 ubiquitin ligases. Similarly, anergic T cells are found in both humans and mice with *F. hepatica* infection and schistosomiasis; in the latter case, these T cells express high levels of the anergy molecule GRAIL (gene related to anergy in lymphocytes). Finally, helminth-induced Th2 cells can exhibit PD-1/PDL-2–dependent intrinsically hyporesponsive phenotype, which is characterized by diminished IL-4/IL-5 expression and enhanced IL-21 expression.

Modulation of Antigen-Presenting Cell Function

DCs are the first APCs usually to encounter parasites, and helminth modulation of DC function has been well characterized. Filarial parasites induce downregulation of MHC class I and class II molecules, as well as cytokines and other genes involved in antigen presentation, thereby rendering DCs suboptimal in their ability to activate CD4 T cells. Schistosomes have similar effects on DCs, with subsequent Th2 polarization and inhibited responses to Th1-inducing TLR ligands. In addition, schistosomes modulate the activation of Nlrp3 (NLR family, pyrin domain containing 3) inflammasome and thus IL-β production. Excretory/secretory antigens produced by helminths can inhibit DC synthesis of proinflammatory cytokines, chemokines, and costimulatory molecules and promote DC production of the regulatory cytokines IL-10 and TGF-β. Helminth infection has also been shown to induce in vivo differentiation of a CD103^{-}CD11c^{lo} population of regulatory DCs, which are inefficient in priming effector T cells and instead favor the generation of Tregs. AAMs are able to markedly suppress target cell proliferation, as well as mediate repair of tissue that has been damaged by parasites. In addition, human filarial infection is associated with the expansion of the nonclassical monocyte subset, as well as an immunoregulatory monocyte subset. Helminth antigens can modulate MHC class II and CD80/86 expression on "antigen-presenting" basophils to induce the development of Th2 cells. Myeloid-derived suppressor cells (MDSCs) are a heterogeneous population of immature myeloid cells that share a common property of suppressing immune responses. MDSCs have been demonstrated to be induced by several helminth infections. MDSCs are inflammatory cells that secrete many types of cytokines, including GM-CSF, IL-1α, IL-6, and IL-10, among others. However, the primary suppression function-related cytokines of MDSCs are TGF-β and IL-10.

Apoptosis

Another mechanism of immune evasion is the ability of some helminths to induce host cell apoptosis. Apoptosis has been described as a host regulatory mechanism in various helminth infections, including schistosomiasis, lymphatic filariasis, and onchocerciasis. Helminths and their products have been found to trigger apoptosis pathways and anergy in host immune cells including T cells, APCs, NK cells, and eosinophils. Besides immune cells, nonimmune cells such as intestinal epithelial cells are also targeted by helminths and their products for apoptosis. The two general pathways involved in the apoptosis process are the death receptor pathway and the mitochondrial pathway.

HELMINTHS AND THE MICROBIOTA

Recent work has highlighted the importance of the microbiota in influencing host immunological and metabolic functions. Helminths secrete a variety of products that can directly influence the composition and function of the microbiota, whereas changes in microbiota can have an impact on susceptibility to helminth infection, indicating that helminth–microbiota crosstalk can regulate a variety of host processes.[35] Direct evidence that type-2-cell-derived cytokines can alter the gut microbiota during infection with *T. muris* or *H. polygyrus* has been shown in mouse models, but similar shifts have also been observed in human populations. For *Trichuris* parasites, the microbiota are essential for their developmental process. The expansion of populations of lactobacilli upon helminth colonization is one the most frequently reported observations in helminth infections. However, there have been quite a few discrepancies in the results of human studies evaluating the helminth–microbiota interaction. Helminth modification of the microbiota is also known to directly impact host immunity to other pathogens, including impaired immunity to enteric bacterial infection and enhanced immunity to respiratory viral pathogens. This can be accomplished by helminth antimicrobial peptide production, mucin expression, goblet cell proliferation, and metabolite production. Alteration of short-chain fatty acid production and

affecting the balance of proinflammatory versus antiinflammatory bacterial species play important roles.[35]

REGULATION OF ALLERGY, AUTOIMMUNITY, AND METABOLIC DISEASES IN HELMINTH INFECTION

The hygiene hypothesis postulates that the stimulation of the immune system by microbes or microbial products protects from the development of inflammatory and atopic disorders. Human studies have demonstrated that people living in areas endemic for helminth infections have a decreased reactivity to skin tests for allergens and milder forms of asthma.[36] Experimental animal models have revealed the protective effect of helminth infections against atopy and asthma. Several mechanisms have been proposed for the helminth-induced protection, the chief of which are the induction of Treg activity, regulatory B-cell activity, immunosuppressive cytokines including IL-10 and TGF-β, and the interaction between helminth-derived products and the gut/lung microbiome (barrier regulation of allergy). Similarly, exposure to helminth parasites has been shown to prevent the onset of Th1-mediated diseases, such as multiple sclerosis (MS), diabetes mellitus, and Crohn disease in experimental animal models.[37] Finally, recent studies in mice have shown that type 2 immunity, induced by helminth infection, can maintain adipose tissue homeostasis and promote adipose tissue beiging, protecting against obesity and metabolic dysfunction, and that the immunomodulatory glycan LNFPIII, which is secreted by helminths, can alleviate hepato-steatosis and insulin resistance. Finally, multiple studies have now demonstrated that helminth infections can contribute to tissue-specific and systemic metabolic homeostasis and protection against obesity-associated meta-inflammation. Epidemiological studies conducted in endemic countries have reported an inverse association between infection with the trematode *Schistosoma japonicum* and the nematode *S. stercoralis* and the prevalence of metabolic syndrome and type 2 diabetes in lean and obese subjects, respectively.[38] Improvements of the homeostatic model assessment for insulin resistance (HOMA-IR), hyperinsulinemia, and hyperglucagonemia, hallmarks of whole-body insulin resistance and metabolic dysfunctions, were also observed in rural populations infected with various species of soil-transmitted helminths. These effects were associated with eosinophilia and increased serum levels of total IgE and prototypical type 2 cytokines, IL-4, IL-5, and IL-13, suggesting that the helminth-induced type 2 immune response might play a role. Anthelmintic treatment was shown to reduce circulating markers of type 2 immunity and to impair metabolic homeostasis, as characterized by elevated HOMA-IR and hemoglobin A_{1c} (glycated hemoglobin) (HbA_{1c}).

HELMINTH THERAPY FOR INFLAMMATORY DISEASES

This is based on the principle that helminth-derived factors promote polarized regulatory or Th2 responses, which result in antiinflammatory molecule production and promotion of barrier integrity (often compromised in inflammatory bowel disease and foodborne incompatibilities) and that helminth colonization provides a diverse bacterial environment that protects against intestinal inflammation.[37] To date, two species of helminths have been tested as clinical treatment for therapy of inflammatory diseases: *Trichuris suis* ova (TSO) and infection with *N. americanus*. TSO has been demonstrated to have a minimal or no effect on Crohn disease, ulcerative colitis, and multiple sclerosis and no significant effect on rheumatoid arthritis, allergic rhinitis, or plaque psoriasis. *N. americanus* larvae have been shown to be minimally or not effective against Crohn disease and celiac disease but have no significant effect on allergic rhinoconjunctivitis, asthma, or multiple sclerosis. Further to therapeutic trials of helminth infections in inflammatory disease settings, dose-escalation controlled helminth infections in healthy volunteers are currently ongoing.[31]

ON THE HORIZON

- Identification and synthesis of helminth products that can be useful as immune therapy in a variety of inflammatory disorders.
- Deciphering the three-way cross-talk among host immunity, helminths, and the microbiota.
- Elucidation of the detailed mechanisms by which helminths manipulate immune responses to bystander antigens.
- Development of clues to production of novel vaccine candidates to protect against not only helminth infection but also helminth-induced morbidity.
- Combined approaches involving genomics, transcriptomics, proteomics, and metabolomics for assessment of host–helminth interactions.

VACCINES AGAINST HELMINTH PARASITES

Vaccines against helminth infections are a necessary tool for their elimination and eradication for several different reasons.[31,39] Four *Schistosoma* vaccines are currently in different phases of clinical development. *S. haematobium* Sh28GST did not exhibit high efficiency in a phase III trial and is currently being modified. *S. mansoni* Sm-14 is in phase II/III trials while *S. mansoni* Sm-TSP-2 is in phase II trials. Finally, *S. mansoni* Sm-p80 is now entering phase I/II trials. Three hookworm vaccine candidates, Na-GST-1, Na-APR-1, and Na-Asp-1, are in phase I clinical trials. Moreover, human challenge infection models to study the efficiency of schistosomiasis and hookworm vaccines are now established with the first studies showing proof of concept for safety and tolerability. In addition, the two *O. volvulus* vaccine candidates—Ov-103 and Ov-RAL2—are in preclinical testing. With rapid advances in the parasite genomics and proteomics, as well as the newer, better vaccine delivery systems offering more efficient and quicker assessment, the prospects for newer anthelmintic vaccines are excellent, although the potential lack of commercial markets imposes a significant impediment to their development.

REFERENCES

1. Awasthi S, Bundy DA, Savioli L. Helminthic infections. *BMJ*. 2003;327:431–433.
2. Babu S, Nutman TB. Immunopathogenesis of lymphatic filarial disease. *Semin Immunopathol*. 2012;34:847–861.
3. Allen JE, Maizels RM. Diversity and dialogue in immunity to helminths. *Nat Rev Immunol*. 2011;11:375–388.
4. Mishra PK, Palma M, Bleich D, et al. Systemic impact of intestinal helminth infections. *Mucosal Immunol*. 2014;7:753–762.
5. Harris NL, Loke P. Recent advances in type-2-cell-mediated immunity: insights from helminth infection. *Immunity*. 2017;47:1024–1036.

6. Coakley G, Harris NL. The intestinal epithelium at the forefront of host-helminth interactions. *Trends Parasitol*. 2020;36:761–772.
7. Gause WC, Rothlin C, Loke P. Heterogeneity in the initiation, development and function of type 2 immunity. *Nat Rev Immunol*. 2020;20:603–614.
8. Schneider C, O'Leary CE, Locksley RM. Regulation of immune responses by tuft cells. *Nat Rev Immunol*. 2019;19:584–593.
9. Vivier E, Artis D, Colonna M, et al. Innate lymphoid cells: 10 years on. *Cell*. 2018;174:1054–1066.
10. Bouchery T, Le Gros G, Harris N. ILC2s-trailblazers in the host response against intestinal helminths. *Front Immunol*. 2019;10:623.
11. Motran CC, Ambrosio LF, Volpini X, et al. Dendritic cells and parasites: from recognition and activation to immune response instruction. *Semin Immunopathol*. 2017;39:199–213.
12. Coakley G, Harris NL. Interactions between macrophages and helminths. *Parasite Immunol*. 2020;42:e12717.
13. Finkelman FD, Shea-Donohue T, Morris SC, et al. Interleukin-4- and interleukin-13-mediated host protection against intestinal nematode parasites. *Immunol Rev*. 2004;201:139–155.
14. de Ruiter K, Jochems SP, Tahapary DL, et al. Helminth infections drive heterogeneity in human type 2 and regulatory cells. *Sci Transl Med*. 2020:12.
15. Nakayama T, Hirahara K, Onodera A, et al. Th2 cells in health and disease. *Annu Rev Immunol*. 2017;35:53–84.
16. Bouchery T, Kyle R, Ronchese F, et al. The differentiation of CD4(+) T-helper cell subsets in the context of helminth parasite infection. *Front Immunol*. 2014;5:487.
17. Crotty S. T follicular helper cell biology: a decade of discovery and diseases. *Immunity*. 2019;50:1132–1148.
18. Maizels RM, McSorley HJ. Regulation of the host immune system by helminth parasites. *J Allergy Clin Immunol*. 2016;138:666–675.
19. Mukai K, Tsai M, Starkl P, et al. IgE and mast cells in host defense against parasites and venoms. *Semin Immunopathol*. 2016;38:581–603.
20. Shen P, Fillatreau S. Suppressive functions of B cells in infectious diseases. *Int Immunol*. 2015;27:513–519.
21. Klion AD, Ackerman SJ, Bochner BS. Contributions of eosinophils to human health and disease. *Annu Rev Pathol*. 2020;15:179–209.
22. Eberle JU, Voehringer D. Role of basophils in protective immunity to parasitic infections. *Semin Immunopathol*. 2016;38:605–613.
23. Grencis RK. Immunity to helminths: resistance, regulation, and susceptibility to gastrointestinal nematodes. *Annu Rev Immunol*. 2015;33:201–225.
24. Logan E, Chetty A, Horsnell WG. The role of antibody in parasitic helminth infections. *Adv Exp Med Biol*. 2014;828:1–26.
25. Anthony RM, Rutitzky LI, Urban Jr. JF, et al. Protective immune mechanisms in helminth infection. *Nat Rev Immunol*. 2007;7:975–987.
26. Zandman-Goddard G, Shoenfeld Y. Parasitic infection and autoimmunity. *Lupus*. 2009;18:1144–1148.
27. Gieseck 3rd RL, Wilson MS, Wynn TA. Type 2 immunity in tissue repair and fibrosis. *Nat Rev Immunol*. 2018;18:62–76.
28. Venugopal PG, Nutman TB, Semnani RT. Activation and regulation of toll-like receptors (TLRs) by helminth parasites. *Immunol Res*. 2009;43:252–263.
29. Brindley PJ, Loukas A. Helminth infection-induced malignancy. *PLoS Pathog*. 2017;13:e1006393.
30. Johnson TP, Sejvar J, Nutman TB, et al. The pathogenesis of nodding syndrome. *Annu Rev Pathol*. 2020;15:395–417.
31. Ryan SM, Eichenberger RM, Ruscher R, et al. Harnessing helminth-driven immunoregulation in the search for novel therapeutic modalities. *PLoS Pathog*. 2020;16:e1008508.
32. Maizels RM, Smits HH, McSorley HJ. Modulation of host immunity by helminths: the expanding repertoire of parasite effector molecules. *Immunity*. 2018;49:801–818.
33. Drurey C, Coakley G, Maizels RM. Extracellular vesicles: new targets for vaccines against helminth parasites. *Int J Parasitol*. 2020;50:623–633.
34. White MPJ, McManus CM, Maizels RM. Regulatory T-cells in helminth infection: induction, function and therapeutic potential. *Immunology*. 2020;160:248–260.
35. Moyat M, Coakley G, Harris NL. The interplay of type 2 immunity, helminth infection and the microbiota in regulating metabolism. *Clin Transl Immunology*. 2019;8:e01089.
36. Gazzinelli-Guimaraes PH, Nutman TB. Helminth parasites and immune regulation. *F1000Res*. 2018:7.
37. McSorley HJ, Hewitson JP, Maizels RM. Immunomodulation by helminth parasites: defining mechanisms and mediators. *Int J Parasitol*. 2013;43:301–310.
38. van der Zande HJP, Zawistowska-Deniziak A, Guigas B. Immune regulation of metabolic homeostasis by helminths and their molecules. *Trends Parasitol*. 2019;35(10):795–808.
39. Diemert DJ, Bottazzi ME, Plieskatt J, et al. Lessons along the critical path: developing vaccines against human helminths. *Trends Parasitol*. 2018;34:747–758.

Emerging Pandemic Infectious Disease Threats

David B. Corry, Peter J. Hotez, Maria Bottazzi, Jill Weatherhead, Anthony Maresso, Emmaline Heckmann, and Lynn Bimler

INTRODUCTION

Pandemic infectious diseases are caused by pathogens (usually viruses) that have adapted well to growth and reproduction within the human host and that through unique environmental, socioeconomic, and cultural circumstances are able to rapidly spread across national boundaries and even globally. Although uncommon and caused by relatively few pathogens, the extraordinary human, economic, and societal losses caused by pandemic diseases as exemplified by coronavirus disease 2019 (COVID-19) make pandemic diseases of unique importance to clinicians, immunologists, and many other scientists and health care professionals. The pathogenesis of pandemic diseases is complex and unique to each pathogen, but common to all is widespread immunological naïveté within the host population. In this chapter, we consider the pathogens of greatest concern for their pandemic potential. Most of these organisms are viruses, including betacoronaviruses, alphainfluenzavirus, ebola, and the flaviviruses, but numerous bacteria are also emerging with pandemic disease potential. For each organism, we consider the factors, especially immunological, that lead to pandemic spread and prospects for effective therapy and prevention.

HUMAN *BETACORONAVIRUS*

The human betacoronaviruses are enveloped viruses with positive-strand RNA genomes of approximately 30 kb.[1] Bats represent important natural animal reservoirs for betacoronaviruses, causing serious human infections and disease of pandemic potential including the severe acute respiratory syndrome coronavirus (SARS-CoV), Middle East respiratory syndrome coronavirus (MERS-CoV), and the most recent coronavirus disease (COVID)-19 caused by the SARS-CoV-2 virus. In addition, at least two betacoronaviruses can cause upper respiratory infections in humans—HCoV-OC43 and HCoV-HKU1 (Fig. 31.1).[1] These typically cause common cold symptoms, although rarely they can also cause more serious respiratory infections. There are also two alphacoronaviruses that cause upper respiratory infections in humans, known as HCoV-NL63 and HCoV-229E (Table 31.1).[1]

Microbiology and Clinical Expression

Coronaviruses are spherical, enveloped viruses that contain a single-stranded, positive-sense RNA genome. Their name derives from the prominently expressed surface projections composed of the surface membrane (M), envelope (E), and spike (S) structural proteins that appear as halos (coronas) on electron micrographs. The spikes are the most distinguishing feature of coronaviruses and are typically composed of a trimer of the S protein. The receptor-binding domain (RBD) of the S protein engages a variety of host receptors including angiotensin-converting enzyme 2 (ACE2), aminopeptidase N (APN), and dipeptidyl peptidase 4 (DPP4) to trigger cellular entry and the initiation of replication. Host proteins such the proteinase transmembrane protease serine 2 (TMPRSS2) are usually required to activate the S protein and facilitate cellular entry. Genome replication and virion assembly occur within the endoplasmic reticulum and Golgi apparatus and viral shedding occurs through exocytosis.

SARS-CoV, MERS-CoV, and SARS-CoV-2 are significant causes of serious human illness. These viruses are transmitted directly from person to person via aerosol droplets and readily infect epithelial cells of the upper and lower airways after inhalation; intermediate arthropod vectors are not involved in transmission. The coronavirus genome is relatively plastic and can undergo recombination with related viruses co-infecting the same host. It is thought that combinations of mutations and recombination events involving an ancestral bat coronavirus, likely transmitted through other hosts such as the civet or pangolin, led to the emergence of SARS and MERS viruses with enhanced human transmissibility and thus pandemic potential. SARS and MERS viruses cause primarily a pneumonitis-predominant syndrome marked by fever, cough, congestion, malaise, and shortness of breath. Severe disease is marked by respiratory failure and may be complicated by dysfunction of numerous other systems, including the cardiovascular, hematologic-coagulation, urinary, neurologic, and gastrointestinal systems. SARS-CoV emerged in 2002 in South China before causing epidemics in more than 20 countries, including approximately 8000 cases and 800 deaths, mostly in China, Hong Kong, Taiwan, Canada, and Singapore. MERS-CoV emerged in Saudi Arabia in 2012 causing approximately 1000 cases and more than 400 deaths, in addition to cases and deaths in other Middle Eastern countries and in South Korea. SARS-CoV-2 emerged in Wuhan, China, and is currently causing a pandemic and public health emergency of international concern (PHEIC). SARS-CoV-2 is the cause of COVID-19, that so far has caused more than one million deaths and a global financial and security crisis.

Immunopathogenesis

Evidence for acquired immunity to betacoronaviruses with pandemic potential is primarily based on persistent host antibodies in recovered patients and experimental animals. Patients who

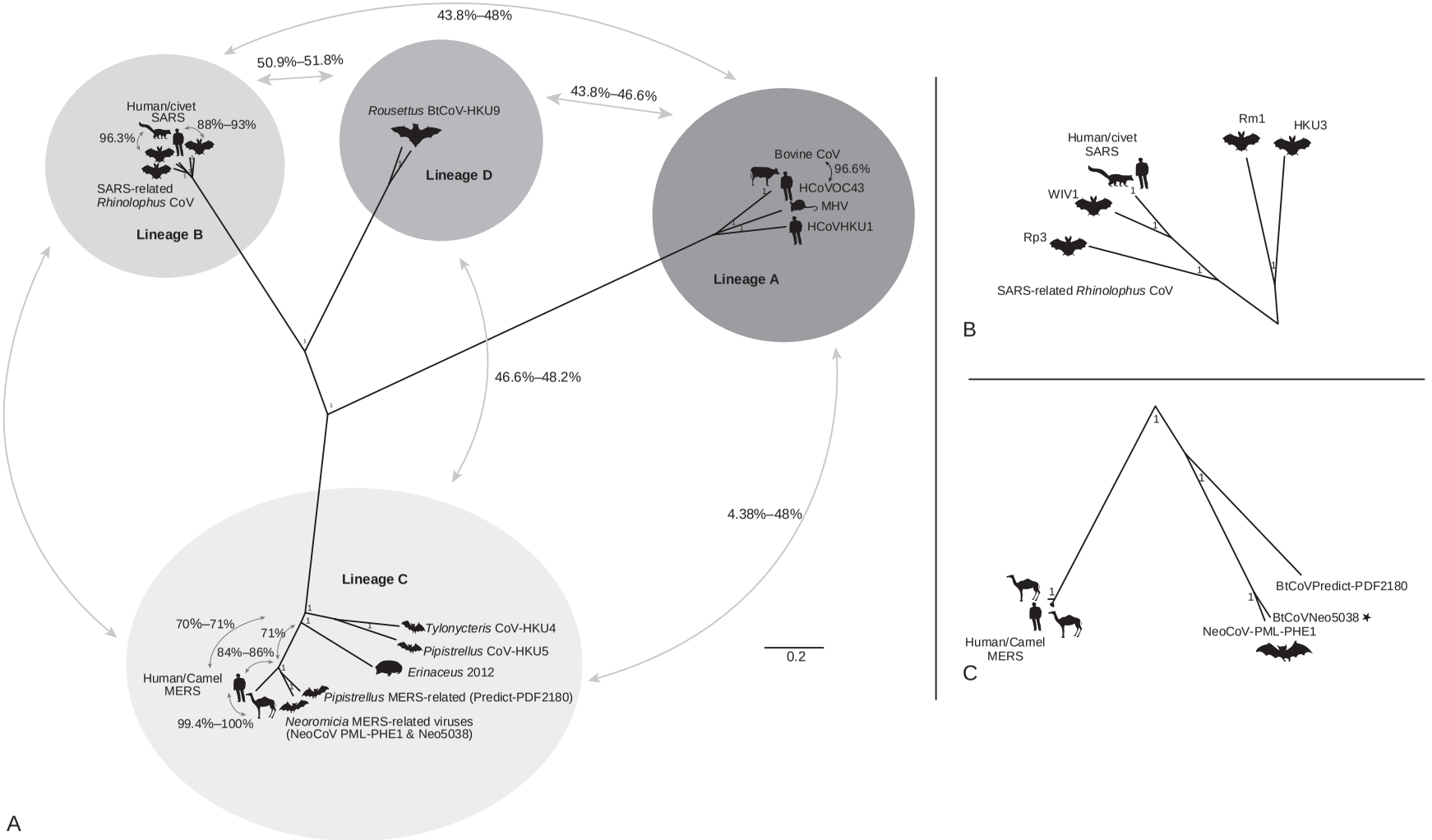


FIG. 31.1 Phylogenetic tree of the lineages of genus *Betacoronavirus* with detail for SARS-CoV and MERS-CoV. (A) The full-genome phylogeny of 4 lineages (A–D) of the genus *Betacoronavirus* constructed using BEAST software with the GTR substitution model using invariant sites and gamma distribution. The MCMC chain was set to 15,000,000 generations sampled every 1500 steps, with a 10% burn-in of the first generated trees and displayed as a radial tree in Figtree. The lineages are indicated with clipart images of host species. Also displayed are the averaged pairwise similarities between lineages as well as highlighted similarities between human coronaviruses and related viruses identified in bats (and other animals). (B) Close-up of the external nodes of the lineage B phylogeny to show relative distances of human and civet SARS-CoV strains and SARS-related *Rhinolophus* strains (WIV1, Rp3, Rm1, and HKU3). (C) Close-up of the lineage C external nodes depicting the human and camel MERS strains with the bat MERS-related viruses (BtCoVNeo5038 from this study is indicated with a star).

TABLE 31.1 Human Beta and Alpha Coronavirus

Betacoronavirus of Serious Pandemic Importance	Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV)	Middle East Respiratory Syndrome Coronavirus (MERS-CoV)	Coronavirus Disease (COVID)-19 Caused by the SARS-CoV-2 Virus
Betacoronavirus of upper respiratory infections	HCoV-OC43	HCoV-HKU1	
Alphacoronaviruses of upper respiratory infections	HCoV-NL63	HCoV-229E	

From: Kearney C, Gleasner JT, Cui H, and Marketter W—https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0194527

recover from SARS-CoV produce persistent antibodies against the coronavirus spike protein, including its RBD. The RBD is required for both SARS-CoV and SARS-CoV-2 to bind to the host ACE2 receptor, which itself is a surface-bound carboxypeptidase found in cells in the lungs, heart, vasculature, and other target organs. For many recovered SARS-CoV patients, antiviral antibody can persist for two years or more.[2] In contrast, it was reported that patients with mild or asymptomatic COVID-19 may exhibit rapidly declining levels of host antibodies,[3] whereas those with severe illness exhibit higher and durable levels of virus neutralizing antibodies, including antibodies directed against the spike protein.[4] So far, there have been a few case reports of COVID-19 patients experiencing reinfection, with at least one individual without virus neutralizing antibodies present at the second clinical presentation.[5] However, rhesus macaques were noted to be resistant to SARS-CoV-2 reinfection after homologous virus challenge, with protection correlating with virus neutralizing antibody titers.[6] In humans, even with declining virus neutralizing antibody titers, memory T and b cells may play an important role in the recall against SARS-CoV-2 antigens, and therefore host immunity.[7,8]

Treatment and Prevention

Immunity to betacoronaviruses has also been shown with passive transfer of antibodies, which confer protection in both humans and animals. In addition, and under development, is the premise that vaccines will induce protection, as seen previously in experimental animals and being evaluated in vaccine clinical trials as of 2020.

Passive Antibody Transfer

In humans, the administration of convalescent plasma was shown in a systematic review and meta-analysis to reduce both viral loads and mortality of patients with SARS-CoV.[9] However, a US National Institutes of Health panel found that for COVID-19, the current data are "insufficient" to endorse the use of convalescent plasma, and instead called for adequately powered and well-controlled randomized clinical trials, even after the US Food and Drug Administration (FDA) had released convalescent plasma under emergency use authorization.[10] Most recently in older adults, high titer convalescent plasma was shown to elicit more convincing therapeutic outcomes.[11] Recently, several monoclonal antibodies directed against the SARS-CoV-2 spike protein RBD were identified and evaluated as promising prophylactic or therapeutic reagents in experimental animal models. They include both neutralizing monoclonal antibodies that block ACE2-RBD binding, as well as those with the ability of inducing subtle or conformational alterations in the RBD protein. Some of the monoclonal antibodies are broadly neutralizing against both SARS-CoV and SARS-CoV-2, while others are virus-specific.

Experimental Vaccines

Some of the best evidence for acquired immunity to the betacoronaviruses is through experimental vaccine trials. An international effort is underway to test experimental vaccines against COVID-19, including a US-based initiative known initially as Operation Warp Speed in addition to national efforts based in Brazil, China, India, Russia, and elsewhere. Almost all of the vaccine concepts rely on inducing strong immunity to the SARS-CoV-2 spike protein, especially its RBD component, with more than 100 candidates in preclinical and clinical development.[13] Several of these vaccines have gone into preclinical testing in nonhuman primates—rhesus macaques—including vaccines using traditional technologies such as whole inactivated virus and recombinant protein vaccines, and new platform approaches such as mRNA, DNA, and adenovirus vectored vaccines. An important observation is that protection against challenge infection in nonhuman primates is associated with high levels of virus neutralizing antibodies, together with T-cell immune responses. Similar observations have been made with experimental vaccines in both rodent and nonhuman primates following SARS-CoV or MERS-CoV challenge infections. For COVID-19, several vaccines have now been released through emergency use authorization processes, based on evidence for protection in limited or expanded phase 3 randomized trials. They include whole inactivated virus, adenovirus-vectored, and nucleic acid vaccine candidates, with results indicating that at least two doses are required to reach virus neutralizing antibody levels (together with T-cell responses) equivalent or exceeding to those found in convalescent sera from recovered patients. At least one study found that protection against COVID-19 exceeds 85% when the level of virus neutralizing antibodies from the vaccine is at least 1.5 fold greater than levels found in recovered patients convalescent sera, although this finding has not yet translated fully into a true correlate of protection.

Immune Enhancement

For some human respiratory infections, it was noted how experimental vaccines for respiratory syncytial virus and measles may have actually exacerbated lung pathology following experimental challenge in preclinical models. For both of these cases, formalin-inactivated whole-virus vaccines were implicated in this phenomenon, known as immune enhancement, which resulted in host cellular infiltrates in the lungs. In some experimental animal models of coronavirus vaccines, especially for SARS-CoV and MERS-CoV, a similar phenomenon was noted, and included eosinophilic or neutrophilic infiltrates in the lungs or liver. Different immunologic mechanisms were ascribed to these observations, including Th2 or Th17 responses, as well as antibody-dependent enhancement (ADE). However, to date immune enhancement has not yet been noted either in

nonhuman primates or in humans following administration of COVID-19 experimental vaccines, as well as virus challenge in the case of the former.

KEY CONCEPTS

- The betacoronavirus SARS-CoV-2 is the causative agent of COVID-19, a worldwide pandemic disease that emerged in late 2019 that has caused devastating human morbidity and mortality in addition to widespread economic disruption.
- The emergence of betacoronaviruses with pandemic potential is related to social determinants including political destabilization, globalization and urbanization.
- Much remains to be learned about betacoronavirus immunity and vaccine development, but rapidly accumulating experience with currently available vaccines approved for emergency use indicate that they are substantially safe and effective.

FLAVIVIRUSES—DENGUE AND ZIKA VIRUS DISEASE

The flavivirus group includes many dozens of viruses, the most medically important of which are the viral causes of dengue, Japanese encephalitis, West Nile fever, yellow fever, and Zika virus disease. Of these five viruses, dengue and Zika viruses have shown the most pandemic potential given their alarming spread from their original home ranges of northern Africa and equatorial Africa and Asia, respectively, to North and South America. Although dengue virus is now endemic in North America, Zika virus is not yet endemic; the many recorded mainland US cases of Zika virus disease are believed to have been acquired abroad. Nonetheless, it is likely that dengue, Zika and other flaviviruses transmitted by Aedes mosquitoes will become endemic to the Gulf Coast of the United States due to climate change, urbanization, and human migrations, among other factors.[14] These viruses most often produce mild, inconsequential disease, but dengue virus is capable of inducing severe and lethal infections and Zika virus can induce great harm to the fetus.

Microbiology

The flaviviruses are enveloped viruses with icosahedral and spherical geometries that contain a single linear positive-sense RNA strand of approximately 11 kb in length. The RNA is transcribed into a polyprotein that is cleaved into separate polypeptides through the combinatorial use of host and viral proteinases that include three structural proteins (capsid, prM, and envelope) and seven nonstructural proteins. Positive viral RNA strands assemble at the host cell membrane where they bud through, becoming encapsulated in host lipids, leading to host cell rupture and release of new virions. Flaviviruses bind to diverse host cell proteins through the viral envelope (E) protein. Host receptors include AXL, Tyro3, DC-SIGN, and TIM-1 for Zika virus[15] and human mannose-binding receptor (MR) and DC-SIGN for dengue virus.[16] Virions enter the cell through clathrin-mediated endocytosis and then commence to replicate.

Like most other flaviviruses, transmission of dengue and Zika virus occur most often through an arthropod, usually the mosquitoes *Aedes aegypti* and *A. albopictus*, but many other arthropods are suitable vectors. These viruses are sufficiently adapted to humans, in which they replicate to a high degree, to permit re-transmission of disease through mosquitoes with no need for other intermediate hosts. However, Zika virus can also be transmitted sexually, by blood transfusion, and vertically to the fetus. It is not known whether specific strains of dengue and Zika virus or whether mutations have facilitated their rapid spread in recent years. Clearly important factors appear to be the spread of suitable mosquito vectors to urban environments where they thrive and the rapid increase in international travel that has repeatedly allowed persons acquiring infections while abroad to transmit the viruses to previously naïve vectors in their native communities, eventually establishing autochthony.[17]

Clinical Expression

Both Zika and dengue viruses typically elicit no symptoms after infection. Approximately 20% of infected subjects will after 3 to 5 days of incubation present with a similar mild syndrome that includes fever, lethargy, joint and muscle aches, joint swelling, maculopapular rash, retro-orbital headache, and conjunctivitis. A more serious form of disease, severe dengue, occurs in approximately 5% of symptomatic subjects and is marked by a systemic vasculitis that results in widespread hemorrhage and resulting organ dysfunction, most often including the liver, brain, and heart. Increasingly, severe dengue in adults has been linked to underlying co-morbid conditions including diabetes mellitus and hypertension.[18] Similar finding have been identified for COVID-19. Severe dengue, which can be fatal even with optimal supportive care, usually occurs soon after defervescence and is preceded by key warning signs. Zika virus has yet to be linked to a severe dengue-like syndrome in adults, but it can produce prostatitis-like symptoms, hematospermia, and neurologic complications including Guillain-Barré syndrome. Maternal infection with Zika virus can lead to fetal loss, intrauterine growth retardation, and CNS complications including microcephaly.[19]

CLINICAL RELEVANCE

WHO Criteria for Severe Dengue

- Severe plasma leakage leading to
  1. Shock and/or
  2. Respiratory distress
- Severe bleeding as assessed clinically
- Severe organ dysfunction
  1. Liver: alanine transaminase or aspartate aminotransferase ≥1000
  2. Central nervous system: impaired consciousness
  3. Heart and other organs

CLINICAL RELEVANCE

WHO Criteria for Dengue Warning Signs

- Abdominal pain or tenderness
- Persistent vomiting
- Fluid accumulation
- Mucosal bleeding
- Lethargy/restlessness
- Liver enlargement >2 cm
- Increase in hematocrit concurrent with thrombocytopenia

Immunopathogenesis

Both viruses are believed to initially infect dermal dendritic cells (DCs), after which they are carried to regional lymph nodes and then systemically via the lymphatic system.

Subsequent disease expression likely depends in part on host age and host and viral genetic factors, but an extremely important determinant of host outcome is immune status. Effective immunity is believed to derive from robust neutralizing antibody responses. Two antibody-related immune phenomena, ADE and original antigenic sin (Hoskins effect), critically influence the effectiveness of humoral responses to flaviviruses, but also influenza. ADE occurs when non-neutralizing antibodies bind to virions and promote their uptake through alternate receptors such as Fcγ receptors present on phagocytic and antigen-presenting cells (APCs). Viral uptake through this aberrant, antibody-dependent mechanism can lead to viral proliferation within $Fc_{\gamma\gamma}$-expressing cells, resulting in prolonged and more severe disease. The closely related Hoskins effect occurs when dengue or Zika viral infection occurs in a semi-immune host that has previously either been infected with a related, but antigenically distinct, viral strain or been vaccinated with a viral strain that is only partially related to the infecting strain. Immunity to the infecting viral strain may in these instances result only in activation of memory, not primary, responses to the previous virus or vaccine, resulting in persistent production of ineffective, non-neutralizing antibodies that may promote ADE. Only primary immunity triggers the process of affinity maturation that results in the high-affinity antibodies needed for virus neutralization.[20,21]

TREATMENT AND PREVENTION

There are currently no antiviral agents approved for use against dengue or Zika viruses, nor is a vaccine against Zika available. While candidate Zika virus vaccines are under evaluation, a major hurdle to their development is the unpredictable nature of Zika virus disease outbreaks, which have abated in recent years. A dengue virus vaccine has recently completed phase III trials and has been approved by the FDA for use in children age 9 to 16 years living in US territories with endemic dengue. While broadly effective, the vaccine can result in severe disease due to ADE in vaccinated persons who subsequently acquire new dengue virus infection.[22] Additional vaccines are under development and completing advanced clinical testing. Control of mosquito populations is an additional promising means of disease control.

KEY CONCEPTS

- Of the many known flaviviruses, dengue and Zika viruses offer the highest pandemic disease risk.
- Both dengue and Zika viruses are transmitted to humans by arthropod, usually mosquito, vectors.
- Rapid spread of vectors to new urban environments and the rise in international travel strongly contribute to the global spread of Zika virus disease and dengue.
- Infection with dengue and Zika viruses is usually inconsequential, but dengue can produce a severe, potentially fatal disease and Zika virus can be harmful or lethal to the fetus.
- Antivirals are unavailable for Zika virus disease and dengue, but a vaccine against dengue is available for children living in endemic regions, with additional vaccines pending.
- Antibody-dependent enhancement and the Hoskins effect contribute to the development of severe disease and complicate efforts to develop effective vaccines.

EBOLA

Ebola virus first emerged in 1976 independently in Zaire, present-day Democratic Republic of Congo (DRC), and Sudan.[23,24] Since 1976, Ebola virus has caused 28 outbreaks, most of which, until recently, have occurred in the Central African countries of DRC, Sudan, Gabon, and Uganda.[25] However, in 2014 to 2016, *Zaire ebolavirus* (EBOV) emerged in West Africa, beginning in Guinea and spreading to neighboring Sierra Leone and Liberia. The 2014 to 2016 EBOV outbreak is the largest Ebola virus outbreak recorded since its emergence. The extent of this outbreak, which caused 28,600 cases and 11,325 deaths, was secondary to involvement of densely populated areas with poor public health infrastructure and increased mobilization across borders leading to spread to seven additional countries including the United States.[25] The global humanitarian and economic impact of the 2014 to 2016 EBOV outbreak led to widespread initiation of public surveillance strategies and a surge in preventive and therapeutic antiviral interventions for future viral outbreaks.

Microbiology

Ebola virus is a nonsegmented, negative-sense, single-stranded RNA virus in the *Filoviridae* family with 5 species: *Zaire (EBOV), Tai Forest (TAFV), Sudan (SUDV), Bundibugyo (EDBV)*, and *Reston (RESTV)*.[23] The virus is made of seven proteins: nucleoprotein, viral protein (VP) 35, VP40, glycoprotein (GP), VP30, VP24, and polymerase L.[23] The RNA resides within the nucleoprotein nucleocapsid alongside polymerase L and transcription cofactor VP30. The viral spike is formed from the transmembrane GPs on the outer envelope and soluble GP (sGP) is a byproduct of viral GP expression.[23,24] The VPs provide structural integrity to the nucleocapsid and modulate the host immune response.[26]

Viral infection occurs as a result of Ebola virus GPs engagement with host cell membrane proteins, promoting cell entry by micropinocytosis. Once engulfed, the viral GP is processed by the host cysteine proteases cathepsin within the endosome, allowing the cleaved protein to interact with the endosomal cellular receptor cholesterol transporter Niemann Pick C1 (NPC1

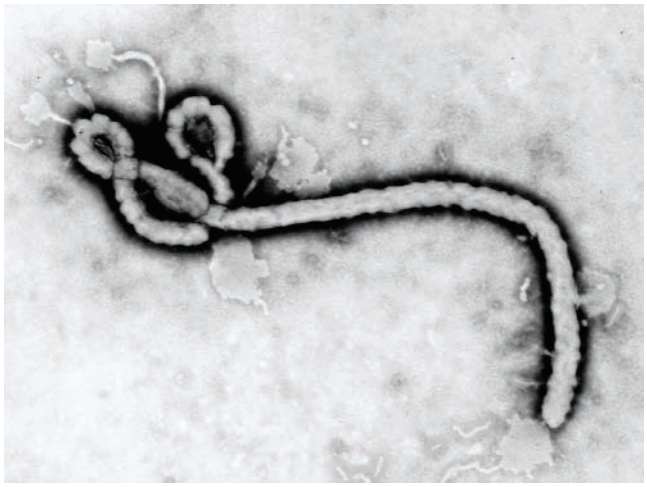


FIG. 31.2 Transmission Electron Microscopy of Ebola virus Demonstrating Filamentous Morphology. (Public Health Image Library #10815. Centers for Disease Control and Prevention and Frederick A. Murphy, 1976.)

protein). Binding of cleaved GP to NPC1 allow the virus to fuse with the endosomal membrane.[23] The viral genome is then released into the cytosol of the host cell and undergoes transcription and replication.[27] The newly replicated virus is assembled into a nucleocapsid and transported to the host cell surface for release of progeny virions.[23] Ebola viruses have the ability to attenuate or shut down host interferon alpha and beta responses leading to progressively higher viral loads. These processes occur through specific viral proteins interfering with interferon signaling cascades.[28]

All five Ebola species can infect humans. While *Zaire*, *Sudan*, and *Bundibugyo* species typically develop similar clinical syndromes, *Reston ebolovirus*, a species identified in a Filipino swine and nonhuman primates, is most commonly associated with asymptomatic infection in humans. Additionally, there is only one recorded case of *Tai Forest ebolavirus*, which occurred in 1994 in Cote d'Ivoire and was nonfatal. Ebola virus is a zoonotic pathogen that typically maintains a sylvatic cycle. It is speculated that the fruit bat, *Pteropodidae* family, is the natural reservoir of Ebola virus and may be capable of transmitting Ebola virus to other intermediate or dead-end hosts.[23]

Initial infections in humans emerge from contact with bats and their excretions or through contact with intermediate or dead hosts such as infected nonhuman primates. Human-to-human transmission can then propagate disease through direct contact with infected human blood, secretions, organs, or other bodily fluids or indirectly through contact with contaminated surfaces. People at high risk of viral transmission include family care providers, healthcare workers, and people involved in traditional burial practices. With an estimated person-to-person transmission ratio of up to 1.34, Ebola virus has the potential to rapidly spread to local and international communities.[29]

Clinical Expression

Ebola virus disease occurs 2–21 days after exposure.[2] Ebola virus disease occurs 2 to 21 days after exposure.[2] Direct viral cytopathogenic effects causing cell death and dysregulated innate and adaptive immune responses lead to a wide range of clinical presentations. Symptoms include fever, fatigue, myalgias, anorexia, nausea, and severe vomiting and diarrhea causing significant fluid losses. Systemic hypoperfusion results in multiorgan dysfunction. Bleeding in the form of microscopic hemorrhage, increased vascular permeability, and impaired coagulation occurs in less than half of cases manifesting clinically as petechiae, bleeding from gums, or bleeding in emesis or stool. Despite the general similarities between Ebola virus species, case fatality rates differ. Although the nature of these differences remain unclear, case-fatality rates range from 40% in *Bundibugyo* species to near 70% to 90% in *Zaire* species. Additionally, evidence from the 2014 to 2016 EBOV outbreak suggest a "post-Ebola syndrome" with chronic, debilitating neurologic, ocular, and musculoskeletal abnormalities in survivors.[24]

Immunopathogenesis

Ebola virus infects humans through the mucosal membranes or breaks in skin. The virus is taken up by APCs such as macrophages and DCs, which permit early intracellular replication. Migration of infected cells to lymph nodes facilitates system-wide spread. Ebola virus leads to disturbances of the innate immune response through inhibition of type I interferons (I-IFN), prolific production of cytokines (interleukin [IL]-1β, IL-1Rα, IL-6, IL-8, IL-15, IL-16, tumor necrosis factor [TNF]-α, nitric oxide [NO]) and chemokines (macrophage inflammatory protein [MIP]-1α, MIP-1β, monocyte chemotactic protein [MCP]-1, migration inhibitory protein [MIF], IP-10), impairment of DC maturation, and natural killer (NK) cell apoptosis leading to overwhelming tissue necrosis. VP35 and VP24 both suppress type-I IFN pathways, VP35 by disrupting retinoic acid-induced gene [RIG]-1 pathway and VP24 by preventing nuclear accumulation of dimerized phosphorylated signal transducer and activator of transcription 1 [STAT-1]. Furthermore, viral sGP act as an immune response accelerant, by binding and activating noninfected DC and macrophages through Toll-like receptor 4 [TLR4], enhancing proinflammatory cytokine and chemokine release. The proinflammatory milieu promotes recruitment of more inflammatory mediators to the site of infection, increasing viral targets. Inhibition of the APC–T-cell synapse due to decreased function of APCs prevents CD4 and CD8 T-cell expansion. The lack of T-cell clonal expansion blocks CD4 T-cell helper functions, including reduction in B-cell release of antigen-specific antibodies. Bystander apoptosis of lymphocytes, which are not directly infected with Ebola virus, occurs potentially due to loss of APC signaling or the overexaggerated cytokine environment. Endothelial dysfunction from proinflammatory cytokines causes increased vascular permeability, fluid extravasation, as well as consumption of coagulation factors.[24]

Treatment and Prevention

There are no approved preventive or curative interventions for Ebola virus infections. Management strategies fall largely on

TABLE 31.2 Ebola Virus Species, Geographic Locations, and Clinical Manifestations

Ebola Species	Emergence	Outbreak Locations	Clinical Manifestations	Case Fatality Rate
Zaire ebolavirus (EBOV)	DRC 1976	Central and West Africa	Fever, fatigue, myalgia, arthralgias, vomiting, diarrhea, maculopapular rash, conjunctival injection, hypovolemia, bleeding or bruising (<50%)	70%–90%
Sudan ebolavirus (SUDV)	Sudan 1976	East Africa	Fever, fatigue, myalgia, arthralgias, vomiting, diarrhea, maculopapular rash, conjunctival injection, hypovolemia, bleeding or bruising (<50%)	50%
Bundibugyo ebolavirus (BDBV)	Uganda 2007	East Africa	Fever, fatigue, myalgia, arthralgias, vomiting, diarrhea, maculopapular rash, conjunctival injection, hypovolemia, bleeding or bruising (<50%)	<40%
Tai Forest ebolavirus (TAFV)	Cote d'Ivoire 1994	Cote d'Ivoire	Unknown (only 1 case reported)	0 (to date)
Reston ebolavirus (RESTV)	Philippines 1989–1889	Philippines, potentially other Asian countries	Subclinical	0 (to date)

the availability of medical resources such as personal protective equipment and on the strength of the public health infrastructure for contact tracing. Detection of Ebola virus is critical for surveillance and initiation of public health measures to prevent outbreaks. Molecular assays including real-time polymerase chain reaction (RT-PCR) are widely used to assess acute infection but are dependent on conserved sequences and may have decreased sensitivity with the emergence of new strain variants. Additionally, RT-PCR may be negative in early disease and requires repeat testing if patient has epidemiologic risk factors and ongoing symptoms. Virus can be detected in many fluids, including blood, breast milk, and semen using RT-PCR.[2] Use of serologic assays, including immunoglobulin (Ig) M and IgG detection are critical for disease surveillance. For persons with Ebola virus disease, standard of care includes aggressive restoration of fluid status, pain control, antiemetics, treatment of co-infections, nutritional supplementation, and correction of electrolyte abnormalities.

Several experimental therapeutics have been evaluated in clinical trials. Early in the course of the 2014 to 2016 outbreak, neutralizing antibodies in the form of convalescent plasma from EBOV survivors was used with varying success. However, development of monoclonal antibodies that target virulent EBOV epitopes have become more readily available. The PREVAIL II trial demonstrated that ZMAPP, triple monoclonal antibody, led to only 22% mortality. However, in the follow-up PALM trial comparing ZMAPP, REGN-EB3 (3 human IgG1 monoclonal antibodies), Mab114 (human monoclonal antibody derived from an Ebola survivor), and remdesivir (an antiviral drug, nucleotide analogue RNA polymerase inhibitor), improved survival was documented in participants who received Mab114 (35.1% mortality) or REGN-EB3 (33.5% mortality) compared to remdesivir or ZMAPP (both with 50% mortality). Differences in effectiveness of monoclonal antibodies may reflect patient population, standard of care practice, and Ebola virus strain variant. The US FDA approved REGN-EB3 (Inmazeb) for treatment of EBOV in October 2020. Synthetic antiviral drugs such as favipiravir, galidesivir, and remdesivir may be more impactful during early viral replication. Trials to evaluate combination therapy with monoclonal antibodies and antivirals are warranted.[30]

Vaccine-driven neutralizing antibodies against Ebola virus proteins may be the key to preventing and controlling outbreaks. A live, attenuated, vectored vaccine using recombinant vesicular stomatitis virus expressing *Zaire ebolavirus* GP known as rVSV-ZEBOV-GP is now approved by the European Medicines Agency and the US FDA and has been valuable for inducing rapid community immunity. This vaccine was highly effective, achieving greater than 90% protective immunity in a devastating outbreak in the Democratic Republic of the Congo, and was instrumental in preventing the spread of Ebola virus infection across Africa. Other vaccines in clinical trial include a chimpanzee adenovirus 3 vaccine (ChAd3-EBO-Z) and the multivalent, modified vaccinia Ankara-based vaccine (MVA-BN). Used in a prime-boost model in clinical trials, MVA-BN with Ad26.EBOV, an adenovirus vector vaccine, may prove to be the most durable and provide cross-protection for several filoviruses, making it an attractive vaccine for high-risk workers and community members. Despite the advances in preventive and therapeutic options, the principles of outbreaks—diagnosis, isolation, contact tracing, quarantining to prevent transmission—remain the staples of viral containment. Building safe workspaces with appropriate biosafety protocols (Biosafety level-4) and supply of personal protective equipment in addition to education of the community on Ebola virus allow for improved outbreak control efforts.[24]

ALPHAINFLUENZAVIRUS—PANDEMIC INFLUENZA

Despite the widespread, albeit imperfect, administration of seasonal vaccines and use of effective antiviral agents, annual influenza epidemics incur high rates of hospitalization and 290,000 to 500,000 deaths worldwide.[31] Pandemic influenza, which is fortunately much less common, can cause many millions of deaths worldwide (Fig. 31.3). The unusually plastic influenza genome that underlies the remarkable tendency of this virus to change its antigenic character; the extended time required for vaccine production, itself an imperfect practice; and the development of antiviral resistance are essential factors driving the risk of pandemic disease. Here we will focus on the paradoxical virulence of influenza as seen even in populations that are widely vaccinated and otherwise immune and future prospects for medical interventions, especially novel vaccines, that may offer the potential of improved long-term protection against this ancient scourge.

Microbiology

Influenza viruses are spherical or filamentous lipid-encapsulated viruses containing 6 to 8 negative sense single RNA strands for a total genome length of 10 to 14 kb. Each RNA strand encodes a functionally distinct polypeptide, the best characterized of which include the hemagglutinin (HA) and neuraminidase (NA) subunits that dominate the viral surface and serve as key antigens recognized by the immune system.

Influenza virions are spread from human vectors by respiratory droplet and through contact with contaminated bodily fluids. Birds, swine, bats, and other mammals are also important vectors for human transmission. Viral HA engages sialic acid residues present on surface proteins of epithelial cells, triggering endocytosis and replication within nuclei. Virions assemble

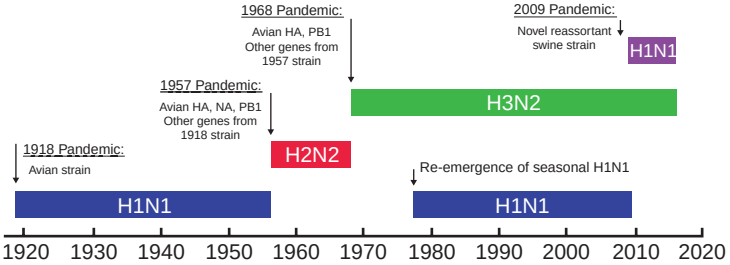


FIG. 31.3 Timeline of Pandemic Influenza Outbreaks and Etiologic Strains Since 1918.

intracellularly and bud from host cells after the RNA genome is enveloped in host phospholipids. Intact virions self-cleave from the host cell surface through the action of NA to continue the infectious cycle.

Clinical Expression

Influenza infection presents as acute respiratory illness with respiratory symptoms (coughing, congestion, etc.) accompanied by the sudden onset of severe systemic symptoms including fever, headache, fatigue, nausea and gastrointestinal symptoms, malaise, and muscle and body aches. Lower respiratory tract infections are more severe and often lead to complications including pneumonia.[32] Pandemic influenza can range in symptoms but generally is associated with more severe disease and often increased mortality in younger populations. However, during the 2009 pandemic, symptoms and mortality were mild, although there was increased mortality in young adults.[32] Avian-derived viruses often cause severe disease with high mortality rates and involve systemic dissemination that results in dysfunction of many organ systems.[32] The defining characteristic of a pandemic influenza strain is its potential to spread quickly through the human population, causing more infections, hospitalizations, and deaths than in a typical influenza season, rather than a specific or more severe clinical presentation. Symptoms, disease severity, and clinical presentation critically depend on the specific virus and immune naïveté within a given population.[32]

Immunopathogenesis

Vaccine efficacy is often measured by the level of neutralizing and HA-inhibiting antibodies produced by the vaccine and a neutralizing humoral response is the primary goal of the influenza vaccine.[31] However, while antibody neutralization of HA or NA can prevent or mitigate infection, the efficacy of these antibodies begins to decrease as the virus mutates targeted regions (i.e., HA globular head group and NA).[33] Other aspects of secondary immunity have been demonstrated to be extremely effective against influenza, including both non-neutralizing antibody responses and T-cell responses.[31,33] Nearly all non-neutralizing antibody-mediated effector functions targeting infected cells have been shown to protect against influenza, including antibody-dependent cellular phagocytosis (ADCP), antibody-dependent cell-mediated cytotoxicity (ADCC), and complement-dependent type III immunity.[33] Cellular immunity directed by T cells is often considered essential in addition to neutralizing antibody responses as it can be directed against more conserved intraviral proteins, targeting viruses that cannot be neutralized because of escape mutations and antigenic drift.[31] Both non-neutralizing antibodies and T cells are sufficient to provide protection from lethal influenza in mice, can target infected cells and highly conserved proteins, and are associated with decreased illness in patients.[31,33]

Alpha and beta influenza viruses (influenza A and B) cause annual epidemics and require annual vaccination to reduce the potential for illness. These viruses mutate rapidly in response to selective pressure from neutralizing antibodies[34] and the most immunogenic regions have a high tolerance for mutation or functional plasticity. Antigenic drift is the accumulation of mutations in viral RNA during infections, thus resulting in altered, antigenically distinct, VP. Antigenic drift is the process that is believed to be responsible for the functionally naïve responses of even heavily vaccinated populations.[34]

Influenza A viruses have high pandemic potential because they infect many species (including birds, swine, and bats) and are capable of antigenic shift, which is the rapid development of antigenically distinct influenza viruses when different influenza A viruses co-infect the same cell and exchange RNA segments (Fig. 31.4). Influenza A viruses are subdivided into serotypes based on expression of distinct HA and NA surface proteins. Each species hosts serotypes with specific HAs and mutations in HA can lead to greater infectivity of humans.[35] Genetic recombination may produce a new HA and NA combination and potentially a pandemic strain when it involves a newly human-adapted (often from birds or swine) or a highly mutated HA.[35]

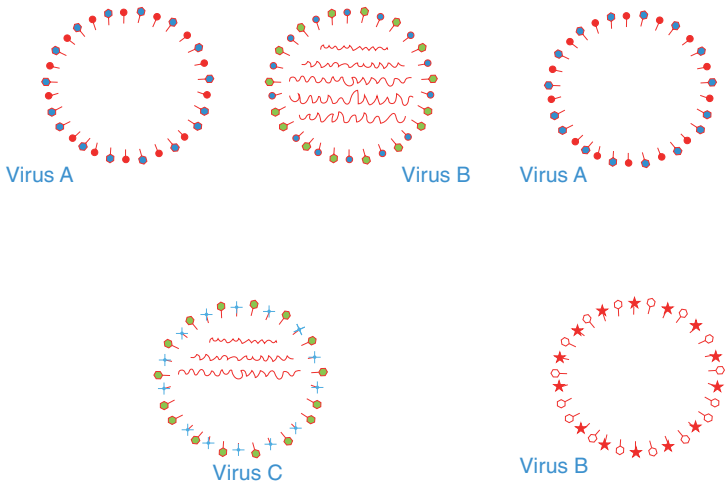


FIG. 31.4 Antigenic Shift and Drift in Influenza. Antigenic shift is the substantial genetic changes that occur during gene recombination in a host cell that is co-infected with two distinct viruses. Antigenic shift occurs rapidly and can result in new subtypes of virus with distinct surface antigens. Antigenic drift reflects the comparatively slow changes in viral surface antigenic character that occur due to naturally accumulating mutations. Antigenic drift can result in new viral strains.

Treatment and Prevention

Vaccines

The WHO leads a global effort to screen viral isolates and select three to four influenza viruses to target in the seasonal vaccine based on analysis of antigenic sites and predictions of population susceptibility.[34] There are three approved types vaccine: inactivated, recombinant, and live attenuated. Utilized in the United States since 1945, inactivated vaccines are most common. Production is relatively slow because expansion of selected viruses in chicken eggs or cell culture, the typical methods of production, are time-consuming processes. The purified and neutralized virus or VPs (HA and NA) are used to create the final vaccine. Another recombinant vaccine uses insect cells modified to express the seasonally predominant HA proteins in vitro without resorting to viral reproduction. This vaccine has the shortest production timeline.

The live attenuated vaccine is a temperature-sensitive influenza virus that expresses the selected HA and NA proteins and is expanded in chicken eggs. It produces an attenuated upper respiratory tract infection that is rarely transmitted and induces strong T cell responses.[35]

Regardless of the vaccine method used, viral strains and mutations are selected 6 months prior to the peak influenza season. Antigenic drift after selection can and does occur, limiting vaccine efficacy (between 10% and 60% since 2005). If antigenic shift occurs after selection, the vaccine will likely provide no protection against a potentially pandemic influenza strain and vaccine production is reinitiated with the novel virus strain.

Egg-based production is limited by the number of quality eggs, which could be even more limited during a pandemic of avian origin.[31] Cell culture–based methods are less limited and faster, but all methods of production are limited by resources[31] and the time required for development, manufacturing, quality control, and distribution.[34]

Developing vaccines for novel strains can have unexpected challenges. Highly virulent H5N1 avian strains initially killed eggs.[35] This and other unanticipated problems occurring during a pandemic could further hamper attempts to release a vaccine. Even optimally designed and produced influenza vaccines provide limited protection to the most at-risk population, people over age 65.[31] Experimental influenza vaccines are showing promise for eventually overcoming these limitations through the inclusion of novel vaccine formulations and adjuvants and the targeting of VPs that are less prone to developing mutations such as the M (matrix) protein.

Antiviral Therapy

There are three classes of approved influenza antivirals: adamantanes, NA inhibitors, and endonuclease inhibitors. All current treatments are approved only within 48 hours of symptom development and some can be administered prophylactically.

Amantadine and rimantadine are adamantanes and block the M2 ion channel, preventing infection. The FDA approved in 1966; adamantanes were the first treatment for influenza A and resistance remained low, with only 0.4% during the 1994 to 1995 influenza season. In 2003, resistance began to rise, reaching 12.3% in the 2003 to 2004 season and 92% in the 2005 to 2006 season. By 2009, all H3N2 and H1N1 isolates tested were resistant.[36]

The first NA inhibitor was FDA approved in 1999. Oseltamivir, zanamivir, and peramivir inhibit NA, preventing viral budding and are currently effective against > 99% of circulating influenza A and B strains. In 2009, oseltamivir-resistant H1N1 began circulating intermittently and some patients develop resistant viruses after treatment. Transmission of NA inhibitor-resistant viruses is limited, but studies report transmission in family clusters and among immunocompromised patients and mutations which may improve transmission. Zanamivir resistance is less common.[36] The Centers for Disease Control and Prevention (CDC) reports that of isolates (H1N1, H3N2, and influenza B) tested during the 2017 to 2018 season, 1% of H1N1 isolates were resistant to both oseltamivir and peramivir, and these were susceptible to zanamivir.

Approved by the FDA in 2018 to treat influenza A and B, baloxavir marboxil is a selective inhibitor that prevents the translation of viral RNA. The CDC reports the isolation of resistant strains from some patients in clinical trials.

Summary

Due to the nature of antigenic shift, pandemic influenza strains usually develop suddenly. With a minimum of six to eight months after detection to develop and mass produce a vaccine, the limited window for treatment with antivirals, the potential for resistance to or the development of resistance to one or more antiviral treatments, an influenza pandemic could cause a world health crisis. Current efforts aim to improve vaccination efficacy in high-risk groups, improve the efficiency of vaccination production, produce a universal vaccination against all serotypes of influenza A, and to improve treatments.

KEY CONCEPTS

Viral Influenza: Pathogenesis, Vaccines and Treatment

- Antigenic drift: the accumulation of mutations during normal infections.
- Antigenic shift: the exchange of RNA segments between two influenza A viruses that have infected the same cell,
- Influenza A viruses have the highest pandemic potential because they:
  - Undergo both antigenic drift and antigenic shift
  - Infect multiple species.
- Pandemic influenza strains usually develop suddenly.
- Vaccine production and limitations:
  - A minimum of 6 to 8 months after detection
  - Seasonally based on predictions
  - Limited by resources
  - Limited efficacy in high-risk groups.
- Antiviral treatments and limitations:
  - Limited window for treatment
  - Potential for resistance to or the development of resistance to one or more antiviral treatments.

EMERGING BACTERIAL THREATS

In the latter half of the 20th century, bacteria have comprised the largest fraction of new emerging infections inclusive of all infectious agents (Fig. 31.5).[37] One can classify emerging bacterial threats of the future into two broad categories, each with a unique natural history. The first are opportunistic infections that result from advances in medicine and civil engineering that otherwise

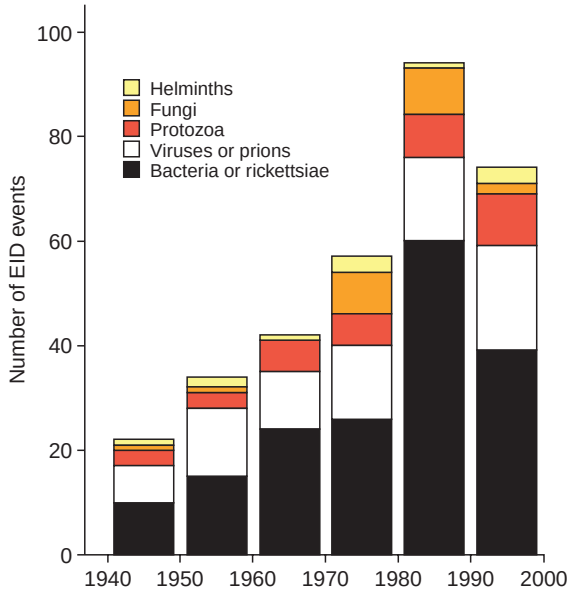


FIG. 31.5 Emerging Bacterial Infections. The number of new emerging infectious diseases *(EIDs)* per decade are indicated for bacteria and rickettsiae relative to other causative microorganisms. (Adapted from Jones KE, Patel NG, Levy MA, et al. Global trends in emerging infectious diseases. *Nature*. 2008;451:990–993.)

prolong life expectancy, temporarily suppress the immune system (such as during organ transplants or due to autoimmunity), or are associated with some type of indwelling device.[38] Such infections are termed healthcare-associated infections (HAIs) or aging-associated infections (AAIs). The human life expectancy has doubled in the last 100 years, resulting in a significantly older population than what would naturally be allowed, thus putting many individuals at a higher risk of life-threatening infections due not just to the longer exposure time to infectious organisms, but also due to senescence of the immune system and other organs. A key difference though as compared to ancient times is that whereas bacterial infections of the past were often acquired through external means (i.e., via contamination of water or food or close human-to-human contact), modern infections originate largely from bacteria that live in and on us—our microbiome. These organisms cause bacteremia and sepsis in the infirm and terminally ill, they infect and grow biofilms on devices (*e.g.*, heart-assist devices, catheters, prosthetics, etc.), and they cause chronic infections in the immunocompromised or those with other underlying illnesses.[39] They are emerging because as medical care becomes more advanced, and human life expectancy continues to push limits, the expectation is that such infections will continue to rise and cause significant morbidity and mortality.

The second classification of emerging bacterial infections are those that arise not through medical advancement, but one in which technology allows human habitation in every corner of the globe.[40] These can be termed anthropogenic-associated infections (AGAIs) because their emergence is fueled by human activity. The ease of habitation of all ecosystems, their eventual encroachment and destruction by human settlement, and the rapid transport of people or goods anywhere in the world, facilitates the emergence of unknown (or previously rare) bacterial species, and strains, into naïve or otherwise susceptible populations. Most of these types of infections can be considered as zoonotic transmissions or diseases. As much remains unknown about these infectious contexts, there is concern that one or more bacterial agents could result in pandemic disease.

Microbiology and Clinical Manifestation

Table 31.3 illustrates some important bacterial species that can be considered as emerging or concerning. The list spans those that are classified as gram-positive or gram-negative upon Gram stain as well as species of spirochetes (vector born) and mycobacteria. Most of the HAIs and AAIs comprise key hallmark species that inhabit our microbiome and include *Staphylococcus, Streptococcus, Escherichia, Pseudomonas, Enterococcus, Klebsiella, Helicobacter*, and *Enterobacter* spp. The AGAIs are more widely varied and include species of *Mycobacteria, Borrelia, Legionella, Bartonella*, and *Rickettsia* spp. The types of infections these bacteria cause and their clinical manifestations are as broad as the species themselves. Generally speaking, HAIs and AAIs are common in those undergoing treatment for cancer or are otherwise immunocompromised, have had surgery, or elderly or infirm, and have indwelling devices including catheters, left ventricular assist devices (LVADs), or implants of other prosthetics that form surfaces by which biofilms can attach. The dominant life-threatening presentation occurs when these bacteria, which are often found at mucosal surfaces, breach this barrier of defense and translocate into the blood and/or distal tissues and organs. AGAIs are often infections of lifestyle, including those that spend much time, or whose lifestyle interfaces, with the rugged outdoors (vector-borne infections from ticks), those involved in animal husbandry and farming (e.g., horses and cattle), those that live in poor sanitary conditions (especially food- or water-borne infections) or whose housing is very dense (slums and towering apartment complexes), and those with free-roaming pets (including dogs and cats).

Vaccine Development and Treatment

There are currently no vaccines for hospital- or aging-associated bacterial infections.[39] There are several reasons for this, including that many bacterial species are a normal part of the human microbiome (and therefore often adapted to pressures placed by the human immune system) and that many bacteria have malleable, pleiotropic genomes where horizontal gene transfer introduces new virulence factors that make identification of vaccine targets more challenging. *Escherichia coli*, for instance, expresses at least 10 distinct pathotypes and each pathotype not only causes a different type of infection but also displays a novel set of virulence factors that drive pathogenesis, sometimes such factors being readily exchanged between strains. In contrast, AGAIs are relatively infrequent and there is a substantial lack of knowledge as to molecular pathogenesis; both factors strongly impair vaccine development.

The most significant concern regarding emerging bacterial threats is the high level and continued rise in resistance to conventional antibiotics. This factor poses the most serious risk for pandemic disease, especially involving key species comprising the human microbiome. Some sources estimate that by 2050, the total loss of lives due to multidrug-resistant bacteria will be 10 million at a cost of $100 trillion to the world economy.[41] Because HAIs and AAIs are often exposed to antibiotics as residents of the human microflora, the very act of treatment generates selective pressures that drive the evolution of clonal groups amongst our microbiome, and these strains then

TABLE 31.3 Emerging Bacterial Pathogens of the Last 50 Years

Year	Bacterial species	Clinical Manifestations	Transmission	Treatment	Comments
N/A	ESKAPE drug-resistant bacteria				
	E*nterococcus* spp.	Endocarditis, UTI, sepsis/blood infection	Microbiome (intestine)	Multidrug due to resistance	---
	S*taphylococcus* spp.	Endocarditis, abscess, sepsis/blood, device and/or skin infection	Microbiome (skin and nares)	Multidrug due to resistance	---
	K*lebsiella* spp.	Device, lung and/or peritoneum infection, UTI	Microbiome (intestine and lung)	Multidrug due to resistance	---
	A*cinetobacter* spp.	Abscess, sepsis/blood infection, lung, and/or wound infection	Microbiome (lungs), environment	Multidrug due to resistance	---
	P*seudomonas* spp.	Lung, wound, eye and/or ear infection, UTI/catheter infection	Microbiome (intestine and lungs)	Multidrug due to resistance	---
	E*nterobacter* spp.	Prosthetics infection, abscess	Microbiome (intestine)	Multidrug due to resistance	---
	E*scherichia* spp.	UTI/catheter, intestine, sepsis/blood, meningitis, and/or peritoneum infection	Microbiome (intestine)	Multidrug due to resistance	---
1973	*Campylobacter* spp.	Diarrhea	Zoonosis	Unnecessary in most cases (macrolides, quinolones)	---
1974	*Clostridium difficile*	Pseudo-membrane colitis; toxic megacolon	Human to human	Vancomycin	Commonly associated with antibiotic use
1974	*Streptococcus bovis* group	Endocarditis	Human to human and/or zoonosis	β-Lactam	Commonly associated with adenocarcinoma of colon and chronic liver diseases
1976	*Legionella pneumophila*	Lung infection	Amoebae in water	Azithromycin, respiratory quinolones	---
1976	*Capnocytophaga canimorsus*	Sepsis	Zoonosis	β-Lactam–β-lactamase combinations	In asplenic patients, hepatic diseases, alcohol abuse
1982	*Escherichia coli* O157:H7	Hemorrhagic colitis, hemolytic uremic syndrome	Zoonosis	Not required	Known as 'hamburger disease'
1982	*Borrelia burgdorferi*	Lyme disease	Zoonosis	Doxycycline, amoxicillin	---
1983	*Chlamydia pneumoniae*	Lung infection	Human to human	Macrolides, doxycycline	First isolated in 1965 in context of trachoma vaccine trial in eye
1983	*Helicobacter pylori*	Gastric ulcers	Human to human	PPI + clarithromycin + amoxicillin/metronidazole	Associated with higher risk of gastric adenocarcinoma and lymphoma
1986	*Rhodococcus equi*	Pneumonia in immunosuppressed	Zoonosis	Multidrug therapy due to resistance	---
1987	*Ehrlichia chaffeensis*	Human ehrlichiosis	Zoonosis	Doxycycline	---
1990s	Non-diphtheria *Corynebacterium* spp.	Endocarditis in immunosuppressed, patients with underlying valve disease or prosthetic valve; other invasive infections	Human to human	β-Lactam + glycopeptides; if resistant, vancomycin	Most important: *C. amycolatum*, initially confounded as *C. xerosis, C. striatum*
1990s	Spotted fever group *Rickettsia* spp.	Spotted fever rickettsiosis	Zoonosis	Doxycycline	Notably *R. africae, R. helveticae, R. slovaca, R. mongolotimonae*
1990	*Anaplasma phagocytophilum*	Human granulocytic anaplasmosis	Zoonosis	Doxycycline	Previously thought to be *Ehrlichia* spp.
1991	*Tropheryma whipplei*	Whipple disease	?	Ceftriaxone followed by trimethoprim–sulfamethoxazole	---
1992	*Vibrio cholerae* O139	Diarrhea	Contaminated water	Not required	---
1992	*Bartonella henselae*	Cat-scratch disease, bacillary angiomatosis	Zoonosis	Generally not required in immunocompetent patients	Initially named *Rochalimaea*
1992	*Aerococcus* spp.	UTI, endocarditis	Human to human	β-Lactam, glycopeptides,	Mainly *A. urinae* and *A. sanguinicola*; especially in elderly or patients predisposing factors such as diabetes, urinary catheters

(Continued)

TABLE 31.3 Emerging Bacterial Pathogens of the Last 50 Years—Cont'd

Year	Bacterial species	Clinical Manifestations	Transmission	Treatment	Comments
1995	*Wolbachia* spp.	Associated with onchocerciasis and lymphatic filariasis	Zoonosis	Doxycycline with or without antifilarial treatment	Indirectly acts as endosymbionts of filarial
1997	*Simkania negevensis*	Lung infection	?	Macrolides, doxycycline	---
1997	*Actinobaculum schaalii*	UTI	?	β-Lactam, glycopeptides,	First considered as a contaminant; especially in elderly or patients with predisposing factors such as diabetes, urinary catheters
1997	*Parachlamydia acanthamoebae*	Lung infection	Amebae in water (?)	Macrolides, doxycycline	Isolated from water of humidifier involved in epidemic of fever in Vermont
2007	*Waddlia chondrophila*	Miscarriages	?	Macrolides, doxycycline	---
2007	*Alloscardovia omnicolens*	UTI	?	β-Lactam, cotrimoxazole, glycopeptides, fluoroquinolones	Especially in elderly or patients with predisposing factors such as diabetes, urinary catheters
2010	*Neoehrlichia mikurensis*	Neoehrlichiosis: systemic inflammatory response; vascular and thromboembolic events	Zoonosis	Doxycycline	More frequent among immunocompromised patients

PPI, proton pump inhibitor; *UTI*, urinary tract infection.
Adapted from Vouga M, Greub G. Emerging bacterial pathogens: the past and beyond. *Clin Microbiol Infect*. 2016;22:12–21.

become endemic to the population. Unlike hospital-associated infections, fortunately, infections emerging due to anthropogenic means are more often treatable with antibiotics unless the infecting species results from strains that circulate in farm animals where high doses of antibiotics (chicken, pork, and cattle) are often used.

ON THE HORIZON

- Many emerging bacterial threats have true pandemic potential, including ESKAPE bacteria.
- There are no current vaccines available against HAIs and AAI-associated organisms.
- Vaccine and new drug development efforts should begin with these species using paradigm-changing approaches that adapt to the evolution of these pathogens in real time.
- As human expansion and travel continue to increase through globalization, so will the emergence of zoonotic bacterial species capable of causing serious infections in humans.
- Future research efforts should focus on surveillance combined with rigorous basic science research programs aimed at understanding the molecular mechanism of disease causation and virulence.

REFERENCES

1. Owusu M, Annan A, Corman VM, et al. Human coronaviruses associated with upper respiratory tract infections in three rural areas of Ghana. *PLoS One*. 2014;9:e99782.
2. Mo H, Zeng G, Ren X, et al. Longitudinal profile of antibodies against SARS-coronavirus in SARS patients and their clinical significance. *Respirology*. 2006;11:49–53.
3. Ibarrondo FJ, Fulcher JA, Goodman-Meza D, et al. Rapid Decay of Anti-SARS-CoV-2 Antibodies in Persons with Mild Covid-19. *N Engl J Med*. 2020;383:1085–1087.
4. Nguyen-Contant P, Embong AK, Kanagaiah P, et al. S Protein-Reactive IgG and Memory B Cell Production after Human SARS-CoV-2 Infection Includes Broad Reactivity to the S2 Subunit. *mBio*. 2020;11
5. To KK, Hung IF, Chan KH, et al. Serum antibody profile of a patient with COVID-19 reinfection. *Clin Infect Dis*. 2020
6. Chandrashekar A, Liu J, Martinot AJ, et al. SARS-CoV-2 infection protects against rechallenge in rhesus macaques *Science*. 369 (2020812–817.
7. Canete PF, Vinuesa CG. COVID-19 Makes B Cells Forget, but T Cells Remember. *Cell*. 2020
8. Hartley GE, Edwards ESJ, Aui PM, et al. Rapid generation of durable B cell memory to SARS-CoV-2 spike and nucleocapsid proteins in COVID-19 and convalescence. *Sci Immunol*. 2020;5:eabf8891.
9. Deng W, Bao L, Liu J, et al. Primary exposure to SARS-CoV-2 protects against reinfection in rhesus macaques. *Science*. 2020;369:818–823.
10. Pau AK, Aberg J, Baker J, et al. Convalescent Plasma for the Treatment of COVID-19: Perspectives of the National Institutes of Health COVID-19 Treatment Guidelines Panel. *Ann Intern Med*. 2021;174:93–95.
11. Libster R, Pérez Marc G, Wappner D, et al. Early High-Titer Plasma Therapy to Prevent Severe Covid-19 in Older Adults. *N Engl J Med*. 2021;384:610–618.
12. Yu F, Xiang R, Deng X, et al. Receptor-binding domain-specific human neutralizing monoclonal antibodies against SARS-CoV and SARS-CoV-2. *Signal Transduct Target Ther*. 2020;5:212.
13. Krammer F. SARS-CoV-2 vaccines in development. *Nature*. 2020;586:516–527.
14. Messina JP, Brady OJ, Golding N, et al. The current and future global distribution and population at risk of dengue. *Nat Microbiology*. 2019;4:1508–1515.
15. Lee I, Bos S, Li G, et al. Probing Molecular Insights into Zika Virus-Host Interactions. *Viruses*. 2018;10:02.
16. Lo YL, Liou GG, Lyu JH, Hsiao M, Hsu TL, Wong CH. Dengue Virus Infection Is through a Cooperative Interaction between a Mannose Receptor and CLEC5A on Macrophage as a Multivalent Hetero-Complex. *PLoS ONE Electron Resour*. 2016;11:e0166474.
17. Tilak R, Ray S, Tilak VW, Mukherji S. Dengue, chikungunya ... and the missing entity - Zika fever: A new emerging threat. *Med*. 2016;72:157–163.
18. Mehta P, Hotez PJ. NTD and NCD Co-morbidities: The Example of Dengue Fever. *PLOS Neglected Tropical Dis*. 2016;10:e0004619.
19. Yacoub S, Mongkolsapaya J, Screaton G. Recent advances in understanding dengue. *F1000Res*. 2016;5
20. Huisman W, Martina BE, Rimmelzwaan GF, Gruters RA, Osterhaus AD. Vaccine-induced enhancement of viral infections. *Vaccine*. 2009;27:505–512.

21. Midgley CM, Bajwa-Joseph M, Vasanawathana S, et al. An In-Depth Analysis of Original Antigenic Sin in Dengue Virus Infection. *J Virology*. 2011;85:410–421.
22. World Health Organization. Dengue vaccine: WHO position paper – September 2018, Weekly Epidemiological Record. 2018;93:457–476.
23. Baseler L, Chertow DS, Johnson KM, Feldmann H, Morens DM. The Pathogenesis of Ebola Virus Disease. *Annu Rev Pathol*. 2017;12:387–418.
24. Feldmann H, Sprecher A, Geisbert TW. Ebola. *N Engl J Med*. 2020;382:1832–1842.
25. Centers for Disease Control and Prevention, Ebola (Ebola Virus Disease). https://www.cdc.gov/vhf/ebola/index.html (2020).
26. Falasca L, Agrati C, Petrosillo N, et al. Molecular mechanisms of Ebola virus pathogenesis: focus on cell death. *Cell Death Differ*. 2015;22: 1250–1259.
27. Feldmann H. Ebola--a growing threat? *N Engl J Med*. 2014;371:1375–1378.
28. Kühl A, Pöhlmann S. How Ebola Virus Counters the Interferon System. *Zoonoses Public Health*. 2012;59:116–131.
29. Camacho A, Kucharski AJ, Funk S, Breman J, Piot P, Edmunds WJ. Potential for large outbreaks of Ebola virus disease. *Epidemics*. 2014;9:70–78.
30. Mulangu S, Dodd LE, Davey Jr. RT, et al. A Randomized, Controlled Trial of Ebola Virus Disease Therapeutics. *N Engl J Med*. 2019;381:2293–2303.
31. Soema PC, Kompier R, Amorij J-P, Kersten GFA. Current and next generation influenza vaccines: Formulation and production strategies. *Eur J Pharmaceutics Biopharm*. 2015;94:251–263.
32. Peteranderl C, Herold S, Schmoldt C. Human Influenza Virus Infections. *Semin Respir Crit Care Med*. 2016;37:487–500.
33. Sicca F, Neppelenbroek S, Huckriede A. Effector mechanisms of influenza-specific antibodies: neutralization and beyond. *Expert Rev Vaccines*. 2018;17:785–795.
34. Webster RG, Govorkova EA. Continuing challenges in influenza. *Ann N Y Acad Sci*. 2014;1323:115–139.
35. Doherty PC, Turner SJ, Webby RG, Thomas PG. Influenza and the challenge for immunology. *Nat Immunol*. 2006;7:449–455.
36. Fiore AE, Fry A, Shay D, Gubareva L, Bresee JS, Uyeki TM. Antiviral agents for the treatment and chemoprophylaxis of influenza --- recommendations of the Advisory Committee on Immunization Practices (ACIP). *MMWR Recomm Rep*. 2011;60:1–24.
37. Jones KE, Patel NG, Levy MA, et al. Global trends in emerging infectious diseases. *Nature*. 2008;451:990–993.
38. Price LB, Hungate BA, Koch BJ, Davis GS, Liu CM. Colonizing opportunistic pathogens (COPs): The beasts in all of us. *PLoS Pathog*. 2017;13:e1006369.
39. Maresso A. *Bacterial Virulence - A Conceptual Primer*. 1 ed. Switzerland AG: Springer International; 2019.
40. Vouga M, Greub G. Emerging bacterial pathogens: the past and beyond. *Clin Microbiol Infect*. 2016;22:12–21.
41. R. on, A. Resistance, Review on Antimicrobial Resistance. Antimicrobial Resistance: Tackling a Crisis for theHealth and Wealth of Nations. 2014; in: J. O'Neill (Ed.), UK Government, Wellcome Trust, UK, 2014.
42. Geldenhuys M, Mortlock M, Weyer J, et al. A metagenomic viral discovery approach identifies potential zoonotic and novel mammalian viruses in Neoromicia bats within South Africa. *PLOS ONE*. 2018;13:e0194527.

Approach to the Evaluation of the Patient With Suspected Immunodeficiency

Jennifer M. Puck and Thomas A. Fleisher

The number of characterized primary immune deficiency (PID) disorders has expanded exponentially over the past three decades, with the underlying genetic basis for the majority of previously known PIDs being identified as well as many new disorders reported in the past decade.[1] PIDs are individually rare, but accurate and prompt specific diagnosis is critical for the early initiation of appropriate therapy in an attempt to prevent or diminish irreversible disease–associated complications and mortality. Hallmark features in most patients suspected of having a PID include a history of increased susceptibility to infection, often together with manifestations of immune dysregulation and/or malignancy. The types of infections experienced by a patient are often clues pointing to which immune functions may be compromised (Table 32.1). The combination of accurate clinical and laboratory characterization of patients with PID combined with mutation analysis (Chapter 18) has also revealed that many PID genotypes can manifest a wider range of clinical phenotypes than previously appreciated.

EVALUATING SUSPECTED DEFECTS IN ANTIBODY RESPONSE

The clinical circumstances that would typically lead to an evaluation for a defect involving antibody production include a history of recurrent sinopulmonary infections with encapsulated bacteria (e.g., *Streptococcus pneumoniae, Haemophilus influenzae*). This group of disorders involves both primary B-cell defects and some complement defects as well as combined defects affecting both humoral and cellular immunity.[2,3] The medical history is essential before proceeding with any directed diagnostic testing and should at a minimum include recurrent infections of the upper and/or lower respiratory tract. Additional infectious sites may involve recurrent/chronic conjunctivitis and gastrointestinal infections. Depending on the underlying defect, these patients may also develop infections with intracellular pathogens as well as evidence of immune dysregulation (e.g., arthritis, gastrointestinal disease). In many cases there will be a family history of recurrent infections such that carefully ascertaining the family pedigree is important, focusing on any relatives' history of recurrent infections, immune dysregulation, and/or malignant disease. The physical examination can also provide important information; in addition to noting evidence of cutaneous infections, easy bruising/bleeding, etc., one of the most telling findings can be whether tonsils are visible. Although a healthy child with recurrent sinusitis or otitis would be expected to have large tonsils, the absence of tonsils, a tissue containing primarily B cells, in this clinical setting strongly suggests a humoral immune defect involving an abnormality in B-cell development. The specific laboratory approach used to further evaluate the patient will depend on what is learned up to this point.

Initial screening for humoral immunity is typically accomplished by measuring serum levels of the major circulating immunoglobulin classes (isotypes), IgG, IgA, IgM, and IgE (see Table 32.1).[4] Antibody levels must be compared with age-matched reference intervals (normal ranges) typically provided as 95% confidence intervals (CIs). The need to compare results to age-specific data relates to the fact that the immune system is undergoing important maturation during infancy and childhood; immunoglobulin reference intervals change substantially until mid to late adolescence. There are no rigid standards regarding the diagnosis of immunoglobulin deficiency, although an IgG value less than 3 g/L (300 mg/dL) in an adolescent or adult or values clearly less than the age-appropriate reference interval (95% CI) in infancy or childhood should trigger further evaluation.[4] When immunoglobulin levels are decreased, it is important to rule out protein loss as the underlying cause by evaluating the serum albumin level. An additional and readily available test is to measure levels of IgG subclasses, most useful in diagnosing selective IgA deficiency with IgG subclass deficiency that is often associated with recurrent sinopulmonary infections.

Measurement of specific antibody responses is necessary to confirm defective antibody production in vivo and is now a criterion for making the diagnosis of many PIDs. In the setting of a medical history that includes recurrent bacterial infections with normal or modestly decreased immunoglobulin levels, confirming a defect in specific antibody production establishes a diagnosis of a rare condition, specific antibody deficiency.[5] One method to evaluate specific antibody responses is testing for spontaneous specific antibodies (e.g., anti-blood group antibodies [isohemagglutinins]) as well as for antibodies to documented previous immunizations or infections (Table 32.2). However, this fails to provide definitive proof of the capacity to respond to a specific antigen at the time of evaluation, which requires evaluation of in vivo antibody production following immunization using one or more protein antigens (e.g., tetanus toxoid, diphtheria toxoid) as well as polysaccharide antigens (e.g., antigens present in the 23-valent polysaccharide vaccine, Pneumovax) (see Table 32.2). Guidelines for normal responses are usually provided by the testing laboratory and typically consist of finding at least a fourfold increase in antibody titer and/or antibody levels considered to be protective in a sample obtained 3 to 4 weeks post immunization (Chapter 94).

TABLE 32.1 Screening Tests According to Immune Deficit

Affected Immunity Arm	Typical Site of Infection	Common Pathogens	Screening Tests
B cells/antibody	Sinopulmonary tract, GI tract, joints, CNS	Pyogenic bacteria: *Streptococcus, Streptococcus, Haemophilus influenzae* Enteroviruses: ECHO, polio *Mycoplasma, Ureaplasma*	IgG IgA IgM IgE Vaccine responses (titers)
T cells	Sepsis, lung, GI tract, skin	Viruses: CMV, adenovirus, measles, *Molluscum contagiosum*, Fungi: *Candida, Aspergillus Pneumocystis jirovecii* Pyogenic bacteria Protozoa: *Cryptosporidium*	CBC with differential Flow cytometry for T cells and T-cell subsets T-cell proliferation to mitogens and antigens (typically *Candida*, tetanus)
NK cells	Skin, lung, GI tract, disseminated infections	Viruses: EBV, CMV, VZV, HSV, HPV	Flow cytometry for NK cells CD107a surface expression NK cytotoxicity assays
Phagocytes	Skin infections, lymphadenitis, liver, lung, bone, GI tract, gingivitis/periodontitis	Bacteria: *Staphylococcus, Serratia marcescens, Burkholderia cepacia, Klebsiella, Escherichia coli, Salmonella, Proteus* Fungi: *Candida, Aspergillus, Nocardia*	Absolute neutrophil count Flow cytometry for expression of CD11/CD18 Dihydrorodamine 123 flow cytometry (DHR) test
Complement	Systemic infections, meningitis	Pyogenic bacteria: *Streptococcus, Haemophilus influenzae, Neisseria*	CH50 AP50

ECHO, enteric cytopathic human orphan virus

Adapted from Rosenzweig SD, Kobrynski K, Fleisher TA. Laboratory evaluation of primary immunodeficiency disorders. In Sullivan KE, Stiehm ER editors, Stiehm's Immune Deficiencies: Inborn Errors of Immunity 2nd ed. Academic Press. Cambridge, MA, 2020. page 117.

An additional approach for assessing the humoral immune response in patients receiving immunoglobulin replacement therapy involves vaccination with a neoantigen (an antigen to which the patient has not previously been exposed) such as the bacteriophage Phi X 174 (not readily available), rabies vaccine, or *Salmonella typhi* capsular polysaccharide vaccine (see Table 32.2). *Salmonella* capsular polysaccharide vaccine, generally given only for foreign travel, has been validated as a tool to assess defective antibody production in individuals receiving IgG products, which generally lack anti-*Salmonella* antibodies.[6]

Additional testing of humoral immunity focuses on characterizing B cells by flow cytometry (Chapter 93). This is useful in the evaluation of congenital forms of agammaglobulinemia that are characterized by absent or extremely low numbers of circulating $CD19^+/CD20^+$ B cells associated with genetic defects that block B-cell development (see Table 32.2).[3] More recently, characterization of B-cell subsets has focused on memory ($CD27^+$) and immature B cells ($CD27^-$) as well as non–class-switched (IgM^+) or class-switched (IgM^-) memory B cells to stratify patients with common variable immune deficiency (CVID).[7] Advanced studies testing in vitro B-cell signaling, apoptosis, and immunoglobulin biosynthesis/secretion are performed only in research centers.

EVALUATING SUSPECTED T-CELL DEFECTS

The typical clinical scenario that suggests a possible T-cell defect is a history of recurrent or prolonged viral or other opportunistic infections. Defects presenting in infancy and early childhood are often associated with a history of failure to thrive, and the clinical phenotype of these disorders may also include autoimmunity. In addition, early onset of a T-cell defect generally heralds progressively severe symptoms that can be fatal. As noted earlier, a family history is important as part of these evaluations because most primary T-cell defects are single gene disorders, and a family member may have suffered from similar events, including in the most serious T-cell disorders resulting in early childhood death. During any evaluation for an underlying T-cell defect, it is important to ensure that the patient is not exposed to harm, either by being administered blood products that have not been filtered and irradiated to eliminate leukocytes responsible for transfusion-related graft-versus-host disease (GvHD) or through infection. In T-cell deficient children, avoiding all live vaccines is necessary to prevent severe infections caused by vaccine strain organisms. The physical examination is another critical part of the evaluation in this group of patients and may reveal cutaneous or oral findings (e.g., petechiae, telangiectasias, thrush) as well as changes in the lymphoreticular system (e.g., absent or enlarged lymph nodes).

KEY CONCEPTS

Medical/Family History in Primary Immune Deficiency Diseases

- Infections
  - Frequent and/or recurrent
  - Severe
  - Persistent, despite standard therapy
  - Caused by opportunistic pathogens
- Failure to thrive in infants and children, weight loss in adults
- Dysregulated immune reactions
  - Autoimmunity
- Family history
  - Similarly affected family member(s)
  - Family member(s) with any recognized immune disorder, recurrent infections, or undiagnosed infant death
  - Parents with common ancestry, consanguinity (for autosomal recessive primary immune deficiency diseases)
  - Membership in a group carrying recognized founder mutations (Amish/Mennonite; Navajo or Apache Native American, etc.)
- Primary Versus Secondary Immune Deficiency

TABLE 32.2 Overview of Laboratory Testing to Evaluate Primary Immunodeficiencies

	Initial Tests	Secondary Tests	Advanced Tests
Suspected B-cell defects	– Quantitative immunoglobulins (IgG, IgA, IgM, IgE) – Specific antibody titers – Natural antibodies (e.g., isohemagglitinins anti-A and anti-B against blood group antigens) – Random, pre/post immunization titers to protein (e.g., tetanus toxoid, diphtheria toxoid) and polysaccharide antigens (pneumococcal vaccine)	– IgG subclasses (restricted utility) – B-cell immunophenotyping (CD19, CD20, CD10, CD21, CD23, CD27, CD38, CD40, CD81, CD138, surface Igs, κ chain, λ chain) – Antibody response to vaccination with a neoantigen (Phi X 174, rabies, *Salmonella typhi*)	– Class switching – In vitro immunoglobulin production (antibody secreting cell generation, ELISPOT for specific Ig production) – Mutation analysis (e.g., *BTK, AID, IGHM*)
Suspected T-cell defects	– CBC and differential – T-cell immunophenotyping (CD3, CD4, CD8, CD45RA/RO, TCRαβ/γδ) – TRECs (population-based newborn screening)	– Extended T-cell immunophenotyping (CD3 chains, CD62L, CD31, CCR7, CXCR5, CD40L, CD127, CD132; MHC-I, MHC-II) – Lymphoproliferation in response to mitogens (PHA, ConA, PWM, PMA+I), CD3/CD28, and antigens (including alloantigens and recall antigens) – Assessment for interstitial chromosome deletion, most commonly ch22q11.2 deletion DiGeorge syndrome	– Vβ TCR repertoire (by immunophenotyped, spectratyping or deep sequencing) – In vitro cytokine production – ADA/PNP enzyme activity and accumulation of toxic purine nucleotides – Radiosensitivity testing – Mutation analysis (e.g., *IL2RG, IL7R, JAK3, RAG1/2, DCLRE1C*)
Suspected NK and NKT cell defects	– CBC and differential – NK/NKT cell immunophenotyping (CD3, CD16, CD56)	– Expanded NK/NKT cell immunophenotyping (KIRs, CDG2/CD94, NKp46, CD117, Vα24, Vβ11) – NK cytotoxic activity on K562 cells	– NK cytotoxic activity on other cells (Raji, 721.221, SKBR3) – NK ADCC – NK cytokine production (ELISPOT) – NK redirected lysis assays
Suspected phagocyte cell defects	– CBC and differential – Morphology: smear evaluation – DHR flow cytometry assay (alternative NBT test)	– Adhesion molecule evaluation: β_2 integrins (CD18, CD11a, b, c), CD15 – Phagocyte cell evaluation, *i.e.*, APC, monocytes – Immunophenotyping (CD14, CD68, CD86, HLA-DR, IFNGR1, IL12Rβ1)	– Chemotaxis – Bactericidal activity – STAT1/STAT4 phosphorylation in response to IFN-γ/IL-12 – IL-12 production in response to IFNγ; – Mutation analysis (e.g., *CYBB, CYBA, NCF1, NCF2, NCF4, IFNGR1, IL-12RB1*)
Suspected complement defects	– CH50 – AP50	– C3, C4 – MBL	– Immunoassay for individual complement components – Functional component testing
Suspected TLR signaling defects	– CD62L shedding	– Specific TLR ligand stimulation assessed by measuring cytokine secretion	– Mutation analysis (e.g., *IRAK4, MYD88, NEMO, TLR3*)

ADA, Adenosine deaminase; *ADCC*, antibody-dependent cellular cytotoxicity; *CBC*, complete blood count; *DHR*, dihydrorhodamine 123; *Ig*, immunoglobulin; *IL*, interleukin; *IFN*, interferon; *MHC*, major histocompatibility complex; *NBT*, nitroblue tetrazolium; *NK*, natural killer; *NKT*, natural killer T; *TCR*, T-cell receptor; *TLR*, Toll-like receptor; *TREC*, T-cell receptor excision circles.

Adapted from Rosenzweig SD, Kobrynski K, Fleisher TA. Laboratory evaluation of primary immunodeficiency disorders. In Sullivan KE, Stiehm ER editors, Stiehm's Immune Deficiencies: Inborn Errors of Immunity 2nd ed. Academic Press. Cambridge, MA, 2020. page 117.

Finally, there may also be syndromic findings consistent with specific diagnoses (e.g., congenital anomalies, microcephaly, unusual facies, short limbed dwarfism).

Initial laboratory evaluation of individuals suspected to have a T-cell disorder should include a white blood cell count and differential focusing on the absolute lymphocyte count compared with age-matched control ranges for proper interpretation (see Table 32.1). Because 50% to 75% of circulating lymphocytes are $CD3^+$ T cells, any process that interferes with T-cell development or increases T-cell loss will result in absolute lymphopenia. It is important to recognize that the total lymphocyte count associated with lymphopenia differs between an infant (<2500/mm^3) and an adult (<1000/mm^3). Profound lymphopenia, particularly in an infant, should prompt immediate immunologic evaluation because it suggests severe combined immunodeficiency (SCID) or complete DiGeorge syndrome, both life-threatening conditions. However, a low number of T cells during infancy is not always found in SCID, because infants with spontaneous engraftment of maternal cells or with "leaky SCID" (including Omenn syndrome) may have normal or elevated total lymphocytes due to oligoclonal expansion.[8,9] In either of these circumstances, the T cells will consist primarily of memory ($CD45RO^+$ T cells) as compared with the T cells found normally in infants, which are primarily naïve ($CD45RA^+$ T cells) (see Table 32.2).[10]

Because of the importance of identifying lymphopenia associated with PID, it is also critical to consider alternative explanations. Lymphopenia associated with human immunodeficiency virus (HIV) infection should be ruled out, and this typically requires testing for the presence of virus nucleic acid or protein (*i.e.*, HIV viral load assay) rather than serologic testing for anti-HIV antibody, which is not reliably produced by immunodeficient individuals. In addition, mechanical loss of lymphocytes, as seen in cases of intestinal lymphangiectasis, should also be considered, especially in patients with chronic diarrhea in whom T-cell proliferation testing is normal. Chylous loss and circulatory alterations can also lead to T-cell lymphopenia.

Direct assessment of the T-cell compartment using flow cytometry provides critical information (Chapter 93).[4,11]

A standard initial clinical flow cytometry approach involves immunophenotyping of T cells with particular focus on enumerating the $CD3^+/CD4^+$ helper and $CD3^+/CD8^+$ cytotoxic T-cell subsets and determining the contribution of naïve, newly produced versus expanded memory T cells. Naïve versus memory T cells can be assessed using a combination of antibodies directed at the CD45 isoforms, CD45RA and CD45RO, respectively. This often is accompanied by in vitro functional testing (e.g., cell proliferation in response to mitogen or antigen, cytokine production, intracellular signaling) (Chapter 94; see Table 32.2).

Other useful tests in selected circumstances include consideration of syndromes due to heterozygous interstitial chromosome deletions, the most common of which is ch22q11.2 deletion associated with DiGeorge syndrome.[11] Although formerly assessed with karyotype analysis or fluorescent in situ hybridization (FISH), current methods including chromosomal microarrays or other tests of copy number, including copy number of the *TBX1* gene in the DiGeorge syndrome critical region, are preferred. Finally, T-cell defects due to adenosine deaminase (ADA) or purine nucleoside phosphorylase (PNP) deficiency require enzyme assays and measurement of the accumulation of toxic purine byproducts in erythrocytes.

NEWBORN SCREENING FOR SEVERE COMBINED IMMUNE DEFICIENCY

Screening an entire population of largely healthy individuals to flag a rare group for follow-up is radically different from the established paradigms described elsewhere in this chapter for diagnosis of immune system disorders. This approach using universal newborn screening for SCID has been adopted throughout the United States and in many other countries as a method to identify infants affected with SCID before they develop life-threatening infections. The great majority of infants with SCID have no family history or outward signs of immune system impairment at birth. However, the heel-stick blood spotted onto filters to screen infants for rare metabolic disorders can be used to isolate DNA for detection of T-cell receptor excision circles (TRECs), formed as byproducts of normal T-cell receptor (TCR) gene rearrangement as T cells mature in the thymus (Chapter 9).[12,13] This assay takes advantage of the fact that the T cell receptor delta (TCRD) locus (located in an intron of the T cell receptor alpha [TCRA] locus) is excised *en bloc*, forming a circle called the δREC-ψJα TREC in all thymocytes destined to express αβTCR antigen receptors. The number of copies of this TREC in a standard punch from a dried blood sample is detected by a quantitative polymerase chain reaction (PCR) amplifying its joined ends. Insufficient or undetectable TRECs constitute a positive SCID newborn screen result, leading to further investigation by flow cytometry as described earlier.

In addition to detecting infants with typical and leaky SCID, regardless of genotype, an abnormal TREC result also flags for further evaluation infants with insufficient T cells due to other causes.[14,15] Non-SCID syndromes, including DiGeorge syndrome, trisomy 21, ataxia telangiectasia, and others, may have neonatal T-cell lymphopenia profound enough to demonstrate an abnormal TREC screen. Secondary non-SCID T-cell lymphopenia also may be caused by T-cell loss or destruction as mentioned earlier, by extreme preterm birth, or by in utero exposure to maternal immunosuppressive medications prescribed for autoimmune disease. A particularly interesting and unanticipated category of infants with abnormal newborn screens for SCID have previously undescribed single-gene disorders, some of which, including *BCL11B* and *EXTL3* deficiencies, have been identified through deep sequencing and molecular studies.[15]

EVALUATING SUSPECTED NATURAL KILLER DEFECTS

Testing of natural killer (NK) cell function is indicated in patients experiencing recurrent viral infections, particularly infections involving the herpes virus family and papilloma virus, as well as patients with hemophagocytic syndromes known as hemophagocytic lymphohistiocytosis (HLH).[16,17] Evaluation includes enumerating NK cells by flow cytometry ($CD3^-CD16^+/CD56^+$ cells) and assaying cytotoxicity using specific in vitro assays (see Table 32.2). Human NK cells can be divided into different subpopulations with the two most characterized defined by the levels of CD56 expression. Among these two NK cell subpopulations are those with high levels of perforin and granzymes in their cytolytic granules with low CD56 expression ($CD56^{dim}$); these cells mediate cytotoxicity. In contrast, NK cells with high levels of CD56 ($CD56^{bright}$) are the most efficient cytokine producers, releasing large amounts of interferon gamma (IFN-γ), enabling this NK cell subpopulation to serve as an immune enhancer.[16]

NK cell function is tested using cytotoxicity assays that measure target cell lysis using either K562 erythroleukemia cells (NK cell cytotoxicity) or B-cell lines (e.g., Raji 721.221) coated with antibody (usually humanized anti-CD20) to assess antibody-dependent cellular cytotoxicity (ADCC). An additional approach as a surrogate in evaluating cytotoxicity depends on the generation of the immunologic synapse associated with binding of the cytotoxic effector (NK) cell to the target cell. This process results in the surface mobilization of CD107a (lysosomal-associated membrane protein [LAMP]-1) that can be evaluated using flow cytometry.[18] Perforin deficiency is unique in that the two different approaches demonstrate contrasting results with abnormal cytotoxicity but normal CD107a expression.[19] This disparity reflects normal cytolytic degranulation, but defective cytotoxicity due to the absence of the critical protein perforin.[19]

Another subset of lymphocytes that is useful to evaluate in certain PIDs (e.g., X-linked lymphoproliferative syndromes [XLP], ITK deficiency) is the natural killer T (NKT) cell. These cells constitute a distinctive subpopulation of mature lymphocytes (Chapter 3) that express the pan–T-cell marker CD3 as well as the NK cell marker CD56 and function as part of the innate immune system. NKT cells can be further subdivided into invariant NKT (iNKT) cells as well as NKT cells with more variable antigen receptors.

EVALUATING IMMUNE DEFECTS INVOLVING MACROPHAGE ACTIVATION

An emerging concept in the field of primary immunodeficiencies is that monogenic disorders can cause recurrent severe infections involving one or a very restricted range of pathogens. Patients with severe, invasive infections involving low virulence or environmental *Mycobacteria* and *Salmonella* species (and other intracellular pathogens) have been found to harbor defects in genes encoding different components of the

interleukin (IL)-12/23-IFN-γ pathway—*IFNGR1, IFNGR2, IL12RB1, IL12RB2, IL23R, IL12B*, *STAT1, TYK2, JAK1, NEMO, IRF8, RORC, ISG15*—and mutations in *SPPL2A, GATA2*, and *CYBB* have also been associated with Mendelian susceptibility to mycobacterial disease (MSMD) (see Table 32.2).[20,21] The family history is a very important part of evaluating these patients because in the majority of cases there will be a similar history of other affected family members.

The two most prevalent genetic defects among this group of patients involve *IL12RB1* and *IFNGR1*, typically resulting in either absent or diminished cell surface protein expression. These surface proteins can be readily assessed using flow cytometry with specific monoclonal reagents. In addition, there is an autosomal dominant defect affecting the *IFNGR1* gene that results in increased surface expression of this protein chain (the receptor is composed of two different proteins) that is most commonly associated with osteomyelitis, which can be detected with flow cytometry (Chapter 35).

Functional response to cytokine stimulation represents a definitive screening test for possible defects in the IL-12/23-IFN-γ axis that generally would be followed by genetic testing to identify the specific defect. For some of the defects associated with MSMD, including defects in *SPPL2a*, *RORC*, *IL23R*, and *IL12RB2*, more specialized or research laboratories are likely to be needed to confirm both protein and functional defects (Chapter 35).

EVALUATING SUSPECTED TOLL-LIKE RECEPTOR DEFECTS

Recurrent infections involving *S. pneumoniae* and *Staphylococcus* species have been associated with defects involving molecules of the Toll-like receptor (TLR) pathway, including IRAK4, MYD88, and NEMO.[22,23] One of the distinctive features of patients with autosomal recessive *IRAK4* and *MYD88* mutations is the diminished inflammatory response to systemic infection, including little or no fever as well as minimal increase in acute phase reactants.[22] NEMO deficiency is a more complex X-linked disorder with a wide range in clinical phenotypes together with varied degrees of immunologic abnormalities.[23] Finally, susceptibility to herpes simplex encephalitis has been linked to mutations in the genes encoding TLR3, the accessory protein of the TLR pathway, UNC93B, and more recently identified intracellular proteins including TBK1 and IRF3 among others.[24] The family history is a very important part of evaluating these patients as in many of the cases there may be a similar history of other affected family members. Additional defects in TLR function associated with specific clinical phenotypes are likely to be identified and represent an evolving field in clinical immunology.

Currently, the evaluation of TLR function is confined to a limited number of centers that usually screen for response using peripheral blood mononuclear cells stimulated with the various TLR-specific ligands (see earlier) followed by measuring cytokine production (intracellular expression or secretion into the culture supernatant) (see Table 32.2). This may then be followed by direct sequencing for the suspected mutant gene(s) involved in the specific TLR signaling process. It is important to recognize that screening TLR function focused on UNC93B and TLR3 defects requires evaluating the response of fibroblasts after exposure to a TLR3 ligand because testing peripheral blood mononuclear cells is ineffective in detecting the functional abnormality. Von Bernuth and colleagues described a simplified assay for screening TLR function that is reported to detect functional defects in the signaling process using whole blood samples (see Table 32.2).[25]

EVALUATING SUSPECTED PHAGOCYTE DYSFUNCTION SYNDROMES

The neutrophil is capable of phagocytosis, bactericidal and fungicidal activity, as well as removal of damaged tissue. A clinical history suggestive of either a numerical or functional abnormality involving neutrophils includes recurrent bacterial and/or fungal infections typically involving the skin and deep organs. As is the case with most PID evaluations, a careful family history should be part of the initial approach to establish if any other members of the extended family have a history of recurrent bacterial and/or fungal infections and/or early death. The physical exam is also an important part of the initial evaluation, looking for evidence of past cutaneous infections that required incision and drainage, wound healing abnormalities, periodontal disease, and other physical findings. The laboratory evaluation should start with a leukocyte count, differential, and morphologic review (see Table 32.1) as neutropenia is the most common category of neutrophil defect (Chapter 39).[26]

Once neutropenia and morphologic abnormalities have been ruled out, the evaluation should be directed at assays that provide functional information about neutrophils, and these primarily focus on two different disorders, one that affects neutrophil migration to sites of infection and the other that impacts neutrophil killing of certain bacteria and fungi (see Table 32.2). The former condition is seen in leukocyte adhesion deficiency (LAD), with the most common type being LAD-1 associated with bacterial infections involving the skin, periodontal tissue, and lungs.[27] The latter defect of neutrophil function is associated with chronic granulomatous disease (CGD), a condition that has a predisposition to bacterial and fungal infections involving the skin and deep organs as well as manifestations of hyperinflammation.[28]

Patients with a functional neutrophil defect involving adhesion abnormalities typically present with increased bacterial infections, leukocytosis (even at times of no infection), and absence of pus at sites of infection. In addition, there is often the finding of severe periodontitis and there may be short stature, distinctive facies and the Bombay (hh) blood type. Laboratory evaluation involves flow cytometric assessment of neutrophils for a defect in β_2 integrin (CD11a/CD18 [LFA1], CD11b/CD18 [Mac-1 or CR3], CD11c/CD18 [p150/95 or CR4]) characteristic of LAD-1 due to mutations in the gene encoding CD18 (see Table 32.2). The actual level of β_2 integrin expression correlates with disease severity in that patients with less than 1% expression have very severe disease, whereas those with 1% to 30% of normal expression have a milder course correlating with the actual expression level.[11] LAD-2, a very rare condition, is associated with a defect in glycosylation type IIc (defect in the GDP fucose transporter). The evaluation of LAD-2 focuses on CD15 (sialyl-Lewis-X) expression, which is absent on neutrophils from LAD-2 patients.

The functional neutrophil defect found in CGD results in recurrent and chronic infections with bacteria and fungi that typically involve the skin, lung, liver, and bone. The five most

prominent microorganisms seen in CGD patients are *Staphylococcus aureus*, *Serratia marcescens, Burkholderia* species, *Nocardia* species, and *Aspergillus* species. The metabolic abnormality associated with serious infectious risk in these patients affects the neutrophil oxidative burst pathway due to defects in the nicotinamide adenine dinucleotide phosphate (NADPH) oxidase multiprotein enzymatic pathway. In addition, these patients typically have hyperinflammation involving the genitourinary (GU) tract and inflammatory bowel disease that is similar to Crohn disease. The most common form of CGD is X-linked recessive, but there are also four autosomal recessive forms. Screening for CGD can be accomplished using the dihydrorhodamine 123 (DHR) flow cytometry assay or the nitroblue tetrazolium (NBT) test, both of which are abnormal in patients with CGD (see Table 32.2).[29] The DHR test is a quantitative assay, meaning that the level of fluorescence is a measure of nicotinamide adenine dinucleotide phosphate (NADPH) oxidase activity. Accordingly, the level of response in the DHR assay has been linked to prognosis and survival in CGD patients.[30] Furthermore, the DHR assay can usually distinguish X-linked CGD female carriers by demonstrating the presence of two distinct populations of neutrophils, normal and abnormal. A recent report established that the proportion of mutant neutrophils in female carriers can be linked to infectious risk with carriers demonstrating less than 20% normal neutrophils; this demonstrates an increased risk of infections involving microorganisms similar to those affecting CGD patients.[31]

Finally, evaluation of neutrophil-directed movement (chemotaxis) can be performed in vivo using the Rebuck skin window technique as well as in vitro with a Boyden chamber or a soft agar system. Abnormalities of chemotaxis have been observed secondary to certain pharmacologic agents as well as in LAD, Chédiak-Higashi syndrome, Pelger-Huet anomaly, juvenile periodontitis, and more recently described diseases such as DOCK2 deficiency. However, chemotaxis tests are difficult to perform and standardize with availability limited to very few laboratories.

EVALUATING SUSPECTED COMPLEMENT DISORDERS

Common phenotypes in complement deficiency are susceptibility to encapsulated bacteria (C1, C4, C2, C3 deficiencies) and neisserial susceptibility (C5, C6, C7, C8, C9 [membrane attack complex] deficiencies) (see Table 32.1).

The complement system involves three activation pathways: the classical pathway (CP), the alternative pathway (AP), and the lectin pathway (LP), all of which converge at C3 and activate a common final pathway (Chapter 40). The three complement activation pathways (the CP, AP, and LP) can be individually evaluated by CH50, AP50 (also known as AH50 or APH50), and the MBL test, respectively (see Table 32.2).[32] The CH50 test assays total complement activity via the CP and is the best single screen for complement abnormalities in that absence or decreased activity in the CH50 implies that at least one of the necessary components is missing or low. The analogous assay for AP activity, the AP50, is not as widely available as the CH50, but it is useful as a screen for complement deficiency, especially when used in conjunction with the CH50. If both CH50 and AP50 are used to screen for complement deficiency, the number of additional tests required to pinpoint the defect can be minimized. Because both assays include the same 6 terminal components (C3, C5, C6, C7, C8, and C9), results will be low or absent for both tests if one or more of these common components is missing. If a unique CP component (C1q, r, s; C4; C2) is missing, the CH50 will be low or absent, but the AP50 will be normal, whereas if a unique AP component is low or missing, the reverse will be true. Functional tests of individual components of the CP and AP are available only from a very limited number of laboratories.

CONCLUSIONS

The expanding identification of PIDs with specific clinical features makes familiarity with the clinical presentation of these disorders of increasing importance. The starting point for an evaluation of a patient for a potential PID is a detailed/directed medical and family history followed by a careful physical examination. These steps, when performed appropriately, should provide sufficient information to proceed with a focused immunologic laboratory evaluation. The appropriate and directed use of immune function testing provides not only critical diagnostic information but also directs decisions regarding the most appropriate therapy with the objective of limiting disease-associated morbidity and mortality. Genetic testing has assumed a greater role in the complete evaluation of individual patients that may also contribute to the decision regarding therapeutic options.

KEY CONCEPTS

Secondary Immunodeficiencies

Immunodeficiency is often secondary or transient, caused by nonimmune factors, including:

- Previous use of high-dose steroids, or other immunosuppressive medications
- Previous use of monoclonal antibodies (mAbs), such as rituximab (anti-CD20)
- Immunoglobulin losses via the gastrointestinal or urinary tract
- Severe illness requiring critical care
- Malnutrition
- Human immunodeficiency virus infection

REFERENCES

1. Tangye SG, Al-Herz W, Bousfiha A, et al. Human inborn errors of immunity: 2019 update on the classification from the International Union of Immunological Societies Expert Committee. *J Clin Immunol*. 2020;40(1):24–64.
2. Conley ME, Dobbs AK, Farmer DM, et al. Primary B cell immunodeficiencies: comparisons and contrasts. *Annu Rev Immunol*. 2009;27:199–227.
3. Smith T, Cunningham-Rundles C. Primary B-cell immunodeficiencies. *Hum Immunol*. 2019;80(6):351–362.
4. Bonilla FA, Khan DA, Ballas ZK, et al. Practice parameter for the diagnosis and management of primary immunodeficiency. *J Allergy Clin Immunol*. 2015;136(5):1186–205 e1–78.
5. Orange JS, Ballow M, Stiehm ER, et al. Use and interpretation of diagnostic vaccination in primary immunodeficiency: a working group report of the Basic and Clinical Immunology Interest Section of the American Academy of Allergy, Asthma & Immunology. *J Allergy Clin Immunol*. 2012;130(3 suppl):S1–S24.
6. Bausch-Jurken MT, Verbsky JW, Gonzaga KA, et al. The use of *Salmonella Typhim* vaccine to diagnose antibody deficiency. *J Clin Immunol*. 2017;37(5):427–433.

7. Wehr C, Kivioja T, Schmitt C, et al. The EUROclass trial: defining subgroups in common variable immunodeficiency. *Blood*. 2008;111(1):77–85.
8. Villa A, Notarangelo LD, Roifman CM. Omenn syndrome: inflammation in leaky severe combined immunodeficiency. *J Allergy Clin Immunol*. 2008;122(6):1082–1086.
9. Muller SM, Ege M, Pottharst A, et al. Transplacentally acquired maternal T lymphocytes in severe combined immunodeficiency: a study of 121 patients. *Blood*. 2001;98(6):1847–1851.
10. Delmonte OM, Fleisher TA. Flow cytometry: surface markers and beyond. *J Allergy Clin Immunol*. 2019;143(2):528–537.
11. Sullivan KE. Chromosome 22q11.2 deletion syndrome: DiGeorge syndrome/velocardiofacial syndrome. *Immunol Allergy Clin North Am*. 2008;28(2):353–366.
12. van Zelm MC, van der Burg M, Langerak AW, et al. PID comes full circle: applications of V(D)J recombination excision circles in research, diagnostics and newborn screening of primary immunodeficiency disorders. *Front Immunol*. 2011;2:12.
13. Chan K, Puck JM. Development of population-based newborn screening for severe combined immunodeficiency. *J Allergy Clin Immunol*. 2005;115(2):391–398.
14. Kwan A, Abraham RS, Currier R, et al. Newborn screening for severe combined immunodeficiency in 11 screening programs in the United States. *JAMA*. 2014;312(7):729–738.
15. Amatuni GS, Currier RJ, Church JA, et al. Newborn screening for severe combined immunodeficiency and T-cell lymphopenia in California, 2010–2017. *Pediatrics*. 2019;143(2):e20182300.
16. Mace EM, Orange JS. Emerging insights into human health and NK cell biology from the study of NK cell deficiencies. *Immunol Rev*. 2019;287(1):202–225.
17. Filipovich AH. Hemophagocytic lymphohistiocytosis and other hemophagocytic disorders. *Immunol Allergy Clin North Am*. 2008;28(2):293–313, viii.
18. Marcenaro S, Gallo F, Martini S, et al. Analysis of natural killer-cell function in familial hemophagocytic lymphohistiocytosis (FHL): defective CD107a surface expression heralds Munc13-4 defect and discriminates between genetic subtypes of the disease. *Blood*. 2006;108(7):2316–2323.
19. Kogawa K, Lee SM, Villanueva J, et al. Perforin expression in cytotoxic lymphocytes from patients with hemophagocytic lymphohistiocytosis and their family members. *Blood*. 2002;99(1):61–66.
20. Rosain J, Kong XF, Martinez-Barricarte R, et al. Mendelian susceptibility to mycobacterial disease: 2014–2018 update. *Immunol Cell Biol*. 2019;97(4):360–367.
21. Filipe-Santos O, Bustamante J, Chapgier A, et al. Inborn errors of IL-12/23- and IFN-gamma-mediated immunity: molecular, cellular, and clinical features. *Semin Immunol*. 2006;18(6):347–361.
22. von Bernuth H, Puel A, Ku CL, et al. Septicemia without sepsis: inherited disorders of nuclear factor-kappa B-mediated inflammation. *Clin Infect Dis*. 2005;41(suppl 7):S436–S439.
23. Hanson EP, Monaco-Shawver L, Solt LA, et al. Hypomorphic nuclear factor-kappaB essential modulator mutation database and reconstitution system identifies phenotypic and immunologic diversity. *J Allergy Clin Immunol*. 2008;122(6):1169–1177 e16.
24. Zhang SY, Casanova JL. Inborn errors underlying herpes simplex encephalitis: From TLR3 to IRF3. *J Exp Med*. 2015;212(9):1342–1343.
25. von Bernuth H, Ku CL, Rodriguez-Gallego C, et al. A fast procedure for the detection of defects in Toll-like receptor signaling. *Pediatrics*. 2006;118(6):2498–2503.
26. Boztug K, Welte K, Zeidler C, Klein C. Congenital neutropenia syndromes. *Immunol Allergy Clin North Am*. 2008;28(2):259–275, vii-viii.
27. Etzioni A. Leukocyte adhesion deficiencies: molecular basis, clinical findings, and therapeutic options. *Adv Exp Med Biol*. 2007;601:51–60.
28. Holland SM. Chronic granulomatous disease. *Hematol Oncol Clin North Am*. 2013;27(1):89–99, viii.
29. Jirapongsananuruk O, Malech HL, Kuhns DB, et al. Diagnostic paradigm for evaluation of male patients with chronic granulomatous disease, based on the dihydrorhodamine 123 assay. *J Allergy Clin Immunol*. 2003;111(2):374–379.
30. Kuhns DB, Alvord WG, Heller T, et al. Residual NADPH oxidase and survival in chronic granulomatous disease. *N Engl J Med*. 2010;363(27):2600–2610.
31. Marciano BE, Zerbe CS, Falcone EL, et al. X-linked carriers of chronic granulomatous disease: illness, lyonization, and stability. *J Allergy Clin Immunol*. 2018;141(1):365–371.
32. Wen L, Atkinson JP, Giclas PC. Clinical and laboratory evaluation of complement deficiency. *J Allergy Clin Immunol*. 2004;113(4):585–593, quiz 94.

33

Primary Antibody Deficiencies

Tracy Hwangpo and Harry W. Schroeder Jr.

Primary antibody deficiency diseases are characterized by an inability to produce sufficient quantities of protective antibodies to protect the host from hazardous antigens.[1] This may be evident from birth or manifest at a later age. In some cases, the deficiency may either resolve or worsen with time.[2] Many of these diseases are caused by mutations that alter the function of genes that regulate B-cell development or homeostasis. Others reflect mutations in immunoglobulin genes themselves. A genetic predisposition marks some of the most common conditions, although the underlying defect may remain unclear. The typical patient is referred for evaluation due to recurrent upper or lower respiratory infections and reduced serum concentrations of one or more classes of immunoglobulin (IgM, IgG, or IgA). Patients with recurrent infections may also present with normal serum immunoglobulin levels but demonstrate an inability to mount a protective response against one or more specific pathogens. Conversely, some virtually agammaglobulinemic patients can be remarkably asymptomatic. Table 33.1 provides a classification of primary antibody deficiency diseases.

Primary immune deficiency disorders are the consequence of specific defects in B-cell development (Fig. 33.1). B-cell production begins in fetal liver and shifts to bone marrow later in fetal life (Chapter 7). As B cells mature, they leave the liver and the bone marrow and migrate via the blood into secondary lymphoid organs (e.g., spleen, lymph nodes, and other peripheral and mucosal tissues). Contact with a polymeric cognate antigen (Chapter 6), such as a polysaccharide, can promote B-cell activation and differentiation into an antibody-producing plasma cell. Responses to protein antigens (e.g., toxins and viral proteins) require T-cell help. In the germinal centers, B cells can replace an upstream heavy (H) chain constant domain with a downstream one, for example, μ to γ1, altering effector function (Chapter 8). They can also introduce a large number of somatic mutations into the variable (V) domains, generating populations of B cells with increased diversity from which those with optimal tailoring of antibody to the antigen are expanded, a process termed *affinity maturation*.

CLINICAL MANIFESTATIONS

Antibody-deficient patients commonly present with recurrent sinusitis, bronchitis, and pneumonia. Patients may also manifest cellulitis, boils, gastrointestinal discomfort, myalgias, arthralgias, fatigue, and depression. Infections typically involve encapsulated bacteria (e.g., *Streptococcus pneumoniae* and *Haemophilus influenzae*). Protection against these bacteria requires production of anti-polysaccharide antibodies, a process that does not require T-cell help. Because similar bacterial infections occur among patients deficient in neutrophil function or in complement, all three of these host defense mechanisms should be evaluated in patients who present with repeated bacterial infections.

CLINICAL PEARLS

Clinical Manifestations of Antibody Deficiency

- Recurrent bacterial infections
  - Early in untreated disease, infections are primarily due to encapsulated pyogenic bacteria (e.g., *Streptococcus pneumoniae* and *Haemophilus influenzae* type b).
  - Later in untreated disease, damage to mucosal surfaces engenders susceptibility to a wider array of pathogens (e.g., staphylococci, nontypeable *H. influenzae*, and gram-negative rods).
- Recurrent viral infections
  - Typically, viral infections clear normally, but protective immunity against re-infection fails (e.g., recurrent shingles).
  - Occasionally, patients may continue to excrete virus for prolonged periods after resolution of clinical symptoms.
- Increased prevalence of other immunologic disorders
  - Paradoxical increased risk of antibody-mediated autoimmune disorders (e.g., idiopathic thrombocytopenia, autoimmune thyroiditis, systemic lupus erythematosus, pernicious anemia, and celiac disease).
  - Lymphoid hypertrophy.
  - Increased risk of allergic disorders (especially in selective IgA deficiency).

The clinical course of uncomplicated primary infections with viruses such as Varicella zoster or mumps does not differ significantly from that of the normal host. However, antibody deficient patients have difficulty generating long-lasting immunity; thus, chickenpox may repeatedly recur as shingles. The general rule is that T cells typically control established viral infections, whereas antibodies limit initial viral dissemination and cell entry, thereby preventing re-infection. As with all rules, there are exceptions. Hypogammaglobulinemic patients can have difficulty clearing hepatitis B virus from the circulation, poliovirus from the gut, and enterovirus from the brain, leading to progressive and sometimes fatal outcomes.

Because sinopulmonary infections are also common in normal infants and children, in allergic individuals, in smokers, and in patients with other pulmonary diseases (e.g., cystic fibrosis), the threshold for an extensive evaluation for immunodeficiency can be a matter of clinical judgment. However, two or more episodes of bacterial pneumonia within a 5-year period, unexplained bronchiectasis, *H. influenzae* meningitis in an older child or adult, chronic otitis media in an adult, recurrent intestinal infections and diarrhea due to *Giardia lamblia*, or a family history of immunodeficiency all warrant evaluation by an immunologist.

TABLE 33.1 Primary Antibody Deficiencies

Disorder	Gene or Locus	Chromosome
IgAD1: IgA deficiency/common variable immunodeficiency	MHC, KIR	6p21.3, 19p13.3
CVID1: ICOS deficiency (AR)	ICOS	2q33.2
CVID2/IgAD2: TACI deficiency (AD/AR)	TNFRSF13B	17p11.2
CVID3: CD19 deficiency (AR)	CD19	16p11.2
CVID4: BAFF-R (AR)	TNFRSF13C	22q13.2
CVID5: CD20 deficiency (AR)	CD20	11q12.2
CVID6: CD81 deficiency (AR)	CD81	11p15.5
CVID7: CD21 (AR)	CD21	1q32.2
CVID8: LRBA (AR)	LRBA	4q31.3
Formerly CVID9: PKCδ (AR), nowALPS III	PRKCD	3p21.1
CVID10: NF-κB2 (AD)	NFκB2	10q24.32
CVID11: IL-21 (AR)	IL21	4q27
CVID12: NF-κB1 (AD)	NFκB1	4q24
CVID13: IKAROS (AD)	IKZF1	7p12.2
CVID14: IRF2BP2 (AD)	IRF2BP2	1q42.3
CTLA-4 (AD)	CTLA-4	2q33
TWEAK (AD)	TNFSF12	17p13
P13K GOF mutations (AD), p110delta	PIK3CD	1p36.2
PI3K regulatory subunit (AD)	PIK3R1	5q13.1
BLK (AD)	BLK	8p23.1
PTEN deficiency LOF (AD)	PTEN	10q23.31
TRNT1 deficiency (AR)	TRNT1	3p26.2
ATP6AP1 deficiency (XR)	ATP6AP1	Xq28
ARHGEF1 deficiency (AR)	ARHGEF1	19q13.2
SH3KBP1 (CIN85 def) (XR)	SH3KBP1	Xp22.12
SEC61A1 deficiency (AD)	SEC61A	3q21.3
RAC2 deficiency (AR)	RAC2	22q13.1
Mannosyl-oligosaccharide glucosidase deficiency (AR)	MOGS	2p13.1
PLCG2 (AD)	PLCG2	16q23.3
Kabuki syndrome (AD)	KMT2D, KDM6A	12q13.12, Xp11.2
Transient hypogammaglobulinemia of infancy (THI)		
X-linked agammaglobulinemia (XR)	BTK	Xq21.3-q22
X-linked agammaglobulinemia with growth hormone deficiency (XR)	BTK	Xq22.1
Hyper-IgM syndrome		
HIGM1: X-linked hyper-IgM syndrome (XHM) (XR)	CD154 or CD40L	Xq26.3
HIGM2: Activation-induced cytidine deaminase deficiency (AR/AD)	AID	12p13.31
HIGM3: CD40 deficiency (AR)	CD40	20q12.12
HIGM5: Uracil-DNA glycosylase (UNG) deficiency (AR)	UNG	12q24.11
XHM with ectodermal dysplasia (XHM-ED) (XR)	NEMO	Xq28
IKBA/IκBα (AD)	IκBα	14q13.2
INO80 deficiency (AR)	INO80	15q15.1
MSH6 deficiency (AR)	MSH6	2p16.3
Autosomal agammaglobulinemia (AGM)		
AGM1: Immunoglobulin μ H chain deficiency (AR)	IGHG1	14q32.33
AGM2: Surrogate light chain deficiency (AR)	IGLL1/CD179B	22q11.23
AGM3: Ig-associated alpha (Igα) deficiency (AR)	CD79A	19q13.2
AGM4: BLNK deficiency (AR)	BLNK	10q24.1
AGM5: LRRC8 truncation (AD)	LRRC8	9q34.11
AGM6: Ig-associated beta (Igβ) deficiency (AR)	CD79B	17q23.2
AGM7: PI3K regulatory subunit (AR)	PIK3R1	5q13.1
AGM8: E47 transcription factor deficiency (AD)	TCF3	19p13.3
Myelodysplasia with hypogammaglobulinemia		Monosomy 7
Thymoma with immunodeficiency (Good syndrome)		
Selective κ light chain deficiency (AR)	IGKC	2p11.12
Selective IgG subclass deficiencies		
Specific antibody deficiency with normal serum immunoglobulin levels		

AD, Autosomal dominant; *AR*, autosomal recessive; *XR*, X-linked recessive.

KEY CONCEPTS

Hypogammaglobulinemia and Antibody Deficiency

- The genetic mutations that underlie most primary antibody deficiencies tend to affect molecular pathways involved in the regulation of lymphocyte development and homeostasis.
- Clinically significant antibody deficiency is not synonymous with hypogammaglobulinemia.
  - Serum immunoglobulin concentrations vary widely with age.
  - Serum immunoglobulin levels vary with exposure to drugs (e.g., steroids and biologics), infectious agents, and other environmental stressors.
- A complete absence of IgG subclasses resulting from homozygous deletions of heavy chain genes has been observed in healthy individuals.
- Absence of a specific Vκ gene segment has been associated with an increased risk of *Haemophilus influenzae* infection in spite of normal serum Ig levels.

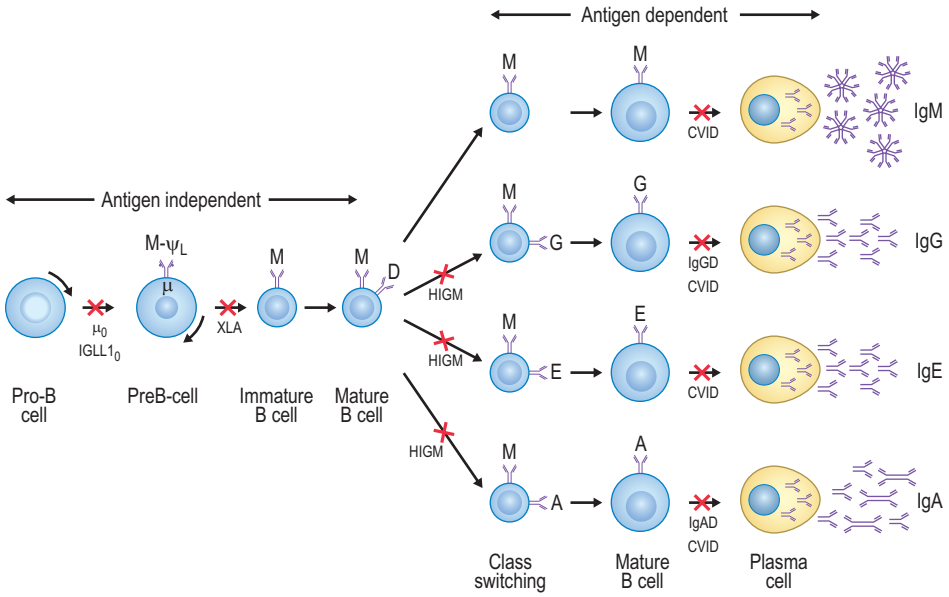


FIG. 33.1 Defects in B-Cell Development Can Lead to Humoral Immune Deficiency. Examples of autosomal agammaglobulinemias (AGM) include the homozygous absence of an immunoglobulin μ H chain (μ_0) or of the λ5 component ($IGLL1_0$) of the surrogate light chain (ψ_L). In the case of μ_0 mutations, a preB-cell receptor ($M\text{-}\psi_L$) cannot be generated, resulting in a block at the proB- to preB-cell stage (illustrated by a *red X*). Signaling through the preB-cell receptor is needed to advance the development process. Patients with X-linked agammaglobulinemia *(XLA)* have loss of function mutations in the kinase Bruton tyrosine kinase (BTK), which impairs signaling through the preB-cell receptor and the B-cell receptor (BCR), resulting in a block in B-cell development between the preB- and immature B-cell stages. Hyper-IgM syndromes (HIGM) result from either a failure to engage in proper cognate interactions with T cells or disruptions in the genes that permit class switch recombination. These failures inhibit class switching to immunoglobulins *IgA, IgG,* and *IgE*. Selective IgG deficiency (*IgGD* or hypogammaglobulinemia), selective IgA deficiency (*IgAD*), and common variable immune deficiency *(CVID)* reflect a selective or generalized impairment in the ability to progress from the mature B-cell stage to the plasma-cell stage.

The purest forms of antibody deficiency result from mutations that prevent the expression or function of the preB-cell receptor for antigen (Chapter 4). Only the B-cell lineage is affected with a loss-of-function mutation of μ heavy chain or components of the surrogate light chain [VpreB, γ14.1 (γ5)]. However, most of the diseases associated with primary antibody deficiency involve more than one cell lineage. For example, X-linked agammaglobulinemia (XLA) is the product of loss-of-function mutations in Bruton tyrosine kinase (*BTK*). Patients with X-linked hyper-IgM syndrome (HIGM1) may exhibit T-cell as well as B-cell dysfunction, placing them at risk for infection with *Pneumocystis jiroveci*. Immune deficiency also appears to place patients at risk for autoimmunity, which is increased among patients with IgA deficiency (IgAD), common variable immunodeficiency (CVID), and hyper-IgM syndrome.

Clinical manifestations of the primary immunodeficiency may also be heavily influenced by the patient's past medical history. A delayed diagnosis and failure to aggressively treat infections can lead to permanent damage to the respiratory or gastrointestinal mucosa, creating susceptibility to nontypeable *H. influenzae*, staphylococci, *Pseudomonas*, and enteric bacteria.

PRINCIPLES OF DIAGNOSIS AND TREATMENT

Diagnostic Tests and Their Interpretation

Testing for immune deficiency should be done for patients (i) with a history of repeated infections that exceeds expectations for normal individuals, (ii) who experience an infection with an opportunistic or low-virulence pathogen, (iii) who are affected with a disorder frequently associated with immunodeficiency, or (iv) who have a family history of primary immunodeficiency. Table 33.2 illustrates four levels of testing complexity.

Level I testing is both revealing and cost-effective. It includes measuring serum immunoglobulins (IgM, IgG, and IgA), complement (50% hemolytic power of serum [CH_{50}] and complement components C3, C4, and mannose-binding lectin protein [MBL]), a complete blood count with differential (CBC/diff), and an erythrocyte sedimentation rate (ESR). Lymphopenia is seen most often in disorders that affect the production or function of T cells (Chapter 34) but can also occur in patients with CVID. Congenital absence of an individual complement component will result in total absence of measurable complement-mediated hemolysis (Chapter 40). MBL deficiency increases susceptibility to respiratory infections.[3] The ESR is elevated in many, although not all, individuals with inflammatory disorders and thus useful in patients with a questionable or unclear history of recurrent or chronic infection.

KEY CONCEPTS

Tests for Immune Function Should Be Performed

- When the patient's history suggests a frequency or severity of infection that exceeds expectations for normal individuals.
- When the organism responsible for infection is of low virulence or is considered to be an opportunistic pathogen (e.g., *Pneumocystis jiroveci*, BCG).
- When there is a diagnosis of a genetic immune deficiency in the patient's family or a multisystem syndrome that can include immune deficiency (e.g., DiGeorge syndrome).

TABLE 33.2 Laboratory Diagnosis of Primary Antibody Deficiency

Level	Test	Application (s)
I	CBC with differential Complement (CH_{50}, C3, C4, mannose-binding lectin protein) ESR Quantitative serum IgM, IgG, and IgA levels	Primary screening tests
Ia	Urinalysis, 24-h urine for protein Stool for alpha-1-antitrypsin Serum albumin	Symptoms suggest protein loss through kidneys or GI tract
II	B-cell functional evaluation Quantitative IgG subclasses, IgE Natural or commonly acquired antibodies (isohemagglutinins, rubella, rubeola, tetanus) Response to immunization T-cell dependent antigens (tetanus) T-cell independent antigens (unconjugated pneumococcal vaccine, unconjugated *Haemophilus influenzae* B vaccine)	Level I normal but history suggests antibody deficiency Better definition of a Level I defect
III	Quantification of blood T- and B-cell subpopulations by immunofluorescence assays using monoclonal antibody markers T cells: CD3, CD4, CD8 B cells: CD19, CD20, CD21, Ig (μ, δ, κ, λ),	Panhypogammaglobulinemia or severely low IgM and IgA
IV	Disease-specific analysis Gene expression Gene sequencing	Gene-specific diagnosis Genetic counseling

CBC, Complete blood count; *ESR*, erythrocyte sedimentation rate; *GI*, gastrointestinal; *Ig*, immunoglobulin.

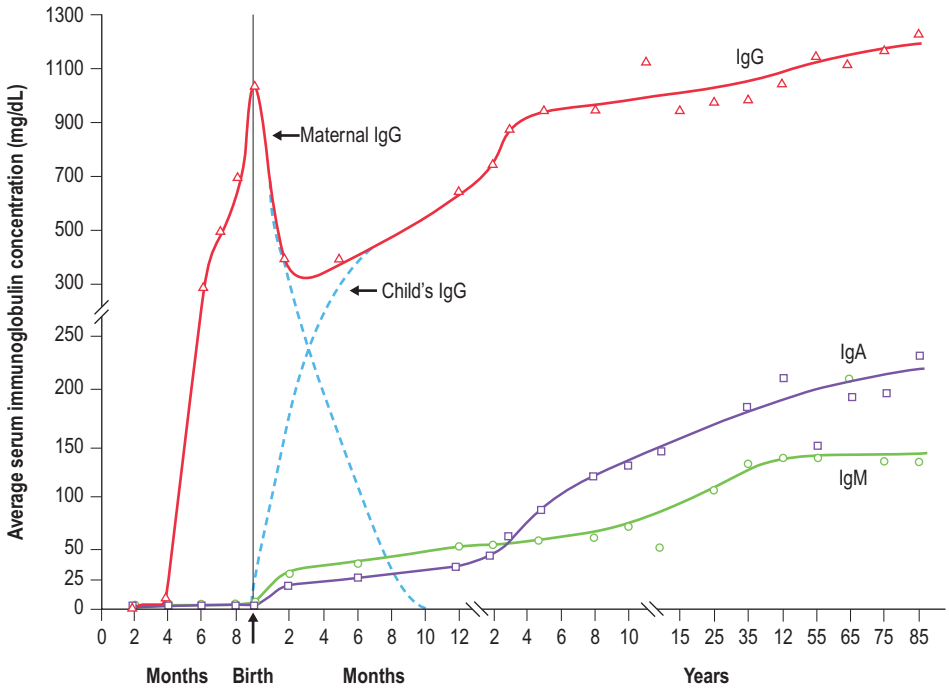


FIG. 33.2 Age-Related Changes in the Serum Concentration of Immunoglobulins. Shown are average serum immunoglobulin (*IgG, IgA*, and *IgM*) concentrations of the major isotypes as a function of age. (From De Greef GE, van Tol MJ, Van Den Berg JW, et al. Serum immunoglobulin class and IgG subclass levels and the occurrence of homogeneous immunoglobulins during the course of ageing in humans. *Mech Ageing Dev*. 1992;66[1]:29–44. Stiehm ER, Fudenberg HH. Serum levels of immune globulins in health and disease: a survey. *Pediatrics*. 1966;37:715–727.)

Interpretation of the significance of the quantitative immunoglobulin determinations requires appreciation of age-related changes in immunoglobulin concentrations (Fig. 33.2).[4,5] At the end of the second trimester of pregnancy, active transport of IgG across the placental barrier begins. By the time of birth at term, the infant's serum IgG concentration is typically 20% to 25% higher than that of the mother. Catabolism of maternal IgG coupled with the slow development of endogenous antibody production leads to a physiologic nadir of serum IgG at 4 to 6 months of age. This loss of maternally derived protection is often associated with the first appearance of otitis media or bronchitis. Thus, the onset of sinopulmonary infections within

the first 3 months of age should also raise the index of suspicion for immunodeficiency in the mother. After age 6 months, maternally derived IgG has largely been lost and IgG antibodies specific for diphtheria or tetanus become useful measures of B-cell function.

CLINICAL PEARLS

Interpreting Quantitative Immunoglobulins

- Normal ranges of serum immunoglobulin levels vary with age; hence evaluation should take the age of the patient into account.
  - As the mother's transplacental contribution of IgG is catabolized, total serum IgG concentrations reach a nadir at 4 to 6 months of age.
  - IgG2 and IgG4 subclass levels rise more slowly than IgG1 and IgG3; hence reference to adult controls can lead to the false diagnosis of IgG subclass deficiency in young children.
  - Serum IgA concentrations typically do not achieve adult values until puberty. They are often the first to decline in many primary immunodeficiencies.

KEY CONCEPTS

Functional Tests of Specific Antibody Production

- IgM T-independent responses may be assessed by measurement of serum isohemagglutinins (anti-A and anti-B titers) in patients who are not blood type AB.
- IgG T-independent responses may be assessed by measurement of antibodies produced in response to immunization with unconjugated purified pneumococcal polysaccharide vaccine.
- IgG T-dependent recall responses may be assessed by measurement of a fourfold or greater rise in titer of antibodies to diphtheria or tetanus toxoid alter booster immunization.

The common laboratory practice of defining the lower range of normal for serum immunoglobulin levels as two standard deviations below the age-adjusted mean carries with it the risk of falsely labeling otherwise normal individuals as immunodeficient. Immunoglobulin levels vary widely with environmental exposure, and normal biologic variation is much broader than that defined by the mean of the population. Patients with a combined deficiency of IgA and an IgG subclass may benefit from the aggressive use of antibiotics and/or immunoglobulin replacement therapy (IGRT); thus, quantitative measurements of all four IgG subclasses (IgG1, IgG2, IgG3, and IgG4) can be useful in fully defining the extent of humoral immune deficiency.

Among patients with borderline serum IgG levels, tests to evaluate the host's ability to produce functional specific antibody should be performed prior to making a decision to institute a more aggressive therapy, especially among patients receiving medications such as corticosteroids, which can lower total serum IgG while preserving function. The most commonly employed tests include measurement of isohemagglutinins (naturally occurring IgM antibodies to the polysaccharide antigens that define the ABO blood type system on red blood cells) and post-immunization responses to polysaccharide antigens (e.g., Pneumovax 23 or unconjugated *H. influenzae* B vaccine) and protein antigens (e.g., tetanus or diphtheria toxoids). IgM is made by the newborn, and most infants can generate isohemagglutinins, making determination of anti-A and anti-B titers a useful measurement of B-cell function even in infants. In older children and adults, isohemagglutinin titers of less than 1:8 are considered significant.[6]

Serum for specific antibody titers should be obtained before and 4 to 6 weeks after immunization. Optimally, the paired sera should be assayed simultaneously to avoid confusion that may result from single-tube dilution differences at the time the assay is performed. As a general rule, a high baseline titer or a fourfold or greater rise in a specific titer in individuals with a low baseline titer confirms that a specific humoral response is intact. The development of conjugated polysaccharide vaccines (e.g., Prevnar 13) complicates analysis for those who have received such a vaccine, although information can still be gleaned from study of responses to polysaccharide antigens present only in multivalent unconjugated vaccines (e.g., Pneumovax 23).

Protein loss through the kidneys or gastrointestinal tract can result in the selective loss of IgG because of its relatively low molecular weight and slow turnover; with severe protein loss, serum albumin levels may also be low. Should symptoms warrant, 24-hour urine for protein or a stool sample for alpha-1-antitrypsin level can document protein loss.

An elevated Inge level may support a suspicion of allergy as an underlying explanation of sinopulmonary symptoms but can also reflect parasite infestation. Serum IgE concentrations are often elevated in patients with IgAD. Extreme elevations of IgE suggest the *hyper-IgE syndrome*.

Enumeration of B cells and of T cells should be performed for any individual who has severe panhypogammaglobulinemia. The most widely used method of demonstrating B cells relies on immunofluorescent labeling of surface CD19, which is restricted in expression to mature B cells (Chapter 7). Because an infant may have serum IgG of maternal origin for the first several months of life, the determination of the number of circulating B cells is the single most useful test in making the presumptive diagnosis of XLA, a disorder in which preB-cells in the bone marrow fail to develop to cells of mature phenotype. Absence of circulating B cells also characterizes the *immunodeficiency associated with thymoma* in adults, whereas adults with *chronic lymphocytic leukemia* (Chapter 77) may have hypogammaglobulinemia with an overabundance of circulating $CD5^+$ B cells.

HIGM1 represents the product of a loss-of-function mutation of the *CD154* (*CD40L*) gene. CD154, a surface antigen found on activated T cells, binds CD40 on B cells to facilitate class switching, survival, and proliferation (Chapter 7). A fluorescent-labeled CD40 fusion protein can be used to evaluate the expression of functional CD154 on T cells by flow cytometry. Confirmation of the diagnosis, carrier detection, and prenatal diagnosis often depend on molecular or sequence analysis of the gene.

Replacement Therapy With Human Immunoglobulin

A number of commercial preparations of human immunoglobulin are US Food and Drug Administration (FDA) approved and available in the United States (Chapter 82). No commercial preparations in this country are available to supplement IgM or IgA exclusively, but some of the preparations contain minute amounts of these isotypes. Since the IgG given in IGRT comes from donors who have produced immunoglobulin in response to the antigens to which they have been exposed, the distribution of antigen specificities can vary between lots, and thus there is no generic form of commercial IgG. Clinically relevant differences between the preparations relate to the route

of administration, which can be by intravenous or subcutaneous infusion; to the method of stabilization and storage; to the donor pool; and to quantities of contaminating IgA. All commercial human immunoglobulin preparations are effective in treating patients with immunodeficiency disorders, although some are better tolerated than others. Depending on the country of origin, some preparations may provide better protection against local endemic organisms. Low IgA content is a concern for those rare individuals with immunodeficiency and absent IgA who manufacture IgG or IgE antibodies directed against IgA and have a history of anaphylactic reactions upon infusion of IgA-containing blood products.[7]

IGRT is not indicated for patients whose immune deficiency is limited to the selective absence of IgA, in part because of the risk of anaphylaxis upon receipt of IgA-containing products, even though such reactions are rare.[7] However, IGRT has been found to be beneficial in patients with a combined deficit of both IgA and an IgG subclass and who exhibit impaired antibody responses to carbohydrate antigens.[8,9]

The goal of immunoglobulin replacement is not to achieve a target IgG level in the serum, but to provide sufficient concentrations of functional antibodies to prevent disease. Specific approaches to and protocols for IGRT are described in detail in Chapter 82. **The only indication for immunoglobulin replacement in a patient with immunodeficiency is severe impairment of the ability to produce functional antibody**. Such impairment exists in primary immunodeficiency diseases associated with low levels of all five isotypes of immunoglobulin, such as XLA, CVID, HIGM, and *severe combined immunodeficiency* (SCID) (Chapter 34). Patients with a documented inability to produce specific antibodies after immunization with a history of significant morbidity from infections are also candidates for intravenous immunoglobulin (IVIG) therapy even though they may present with normal or near-normal levels of IgG.[10] This includes certain cases of IgG subclass deficiency such as those associated with compound IgA, IgG2, and IgG4 deficiency; boys with Wiskott-Aldrich syndrome; and patients with ataxia-telangiectasia. Since most patients with transient hypogammaglobulinemia of infancy (THI) can produce normal amounts of specific antibodies after immunization despite having a low total serum IgG, they usually are not candidates for immunoglobulin replacement. However, patients with a history of significant infections may benefit.

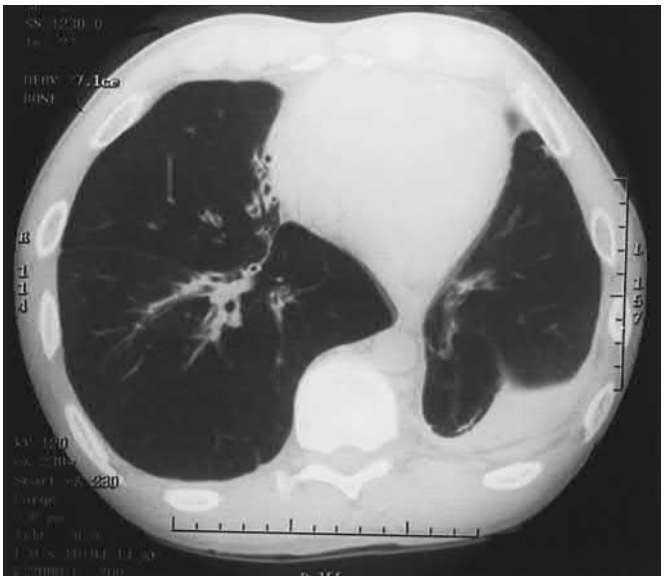


FIG. 33.3 A CT Scan, With Contrast, Demonstrates Bronchiectasis, Bronchitis, and Emphysema in the Lungs of a 36-Year-Old Man With X-linked Agammaglobulinemia. He was referred to immunology and started on immunoglobulin replacement therapy at age 16. Due to a left lower lobe lobectomy, the mediastinum has shifted to the left. As a result of bronchiectatic scarring, the diameter of the bronchi in the right lung are greater than the diameter of the corresponding blood vessels, and the bronchi remain dilated in the periphery. Bronchial plugs can be seen filling some of the bronchi on the right. Finally, due to emphysema, the right upper lobe demonstrates greater radiolucency. In addition to suffering from XLA, this patient has a 30 pack-year history of smoking, which has exacerbated his clinical condition.

X-LINKED AGAMMAGLOBULINEMIA

Diagnosis

XLA (Bruton agammaglobulinemia) is the prototypic humoral immunodeficiency.[11] Due to the transplacental transfer of maternal immunoglobulin, onset of recurrent pyogenic infections generally begins after age 5 to 6 months. Loss-of-function mutations in the *BTK* gene lead to a block in B-cell maturation, near total absence of B cells in the periphery, and panhypogammaglobulinemia (see Fig. 33.1). Testing of infants known or suspected to have XLA should begin with examination of the number of B cells in the blood. Deficient expression of BTK protein in monocytes can be detected by flow cytometry. Analysis of the *BTK* gene at the nucleotide level remains the definitive diagnostic procedure. As with most X-linked lethal diseases, approximately one-third of sporadic cases are due to de novo mutations, so diagnosis may require individual mutation analysis. There can be significant variation in the manifestations of the disease in any given family member; thus a paucity of symptoms should not prevent diagnostic evaluation even in adults.[12] Some patients with "mild" mutations of *BTK* may present with infections late in life.

Clinical Manifestations

Easy access to antibiotics and good hygiene may delay suspicion of the diagnosis well into mid-childhood or beyond.[12] Recurrent upper and lower respiratory tract infections are common. Untreated, these infections may lead to bronchiectasis (Fig. 33.3), pulmonary failure, and death at an early age. The infections are typically due to pyogenic encapsulated bacteria. Diarrhea due to *G. lamblia* is also common, although less so than in CVID. Systemic infections include bacterial sepsis, meningitis, osteomyelitis, and septic arthritis. Mycoplasma and chlamydia infections of the urogenital tract may lead to epididymitis, prostatitis, and urethral strictures. Skin infections include cellulitis, boils, and impetigo.[13]

Although patients with XLA can resolve most viral infections, they are unusually sensitive to infections with enteroviruses. Patients with XLA can develop paralytic poliomyelitis after vaccination with live virus. Echovirus and Coxsackie virus infections may involve multiple organs with the patients going on to develop chronic meningoencephalitis, dermatomyositis, and/or hepatitis.[13] Untreated patients often complain of arthritis affecting the large joints. Enterovirus and mycoplasma have been identified in affected joints. The arthritis typically resolves with IGRT.

Infections with opportunistic organisms (e.g., tuberculosis, histoplasmosis, and *P. jiroveci*) and malignancies are rare, likely reflecting intact cell-mediated immunity.

Origin and Pathogenesis

BTK belongs to a subfamily of the Src cytoplasmic protein-tyrosine kinases. BTK is phosphorylated following activation of the B-cell receptor (BCR). It plays a critical role in the proliferation, development, differentiation, survival, and apoptosis of B-lineage cells. Individuals with XLA begin with normal numbers of early B-lineage progenitors in their bone marrow. In the bone marrow, B-cell progenitors express the expected markers of B-cell differentiation (e.g., terminal deoxynucleotidyl transferase [TdT], CD19, and CD10), but there is a relative deficiency of cytoplasmic μ^+ preB-cells and development beyond the preB stage is severely impaired. Those cells that make it through the gauntlet can produce antigen-specific antibodies. Although in low numbers, these surviving B cells can produce endogenous antigen-specific immunoglobulin, class switch, and even mediate allergic or autoantibody-mediated reactions.

An X-linked recessive form of agammaglobulinemia associated with growth hormone deficiency has been reported. Genetic analysis in one such patient identified a frameshift mutation creating a premature stop codon with the loss of carboxy terminal amino acids in BTK.[14]

Treatment and Prognosis

The primary goal of therapy for B-cell deficiency is to prevent damage to the lungs. Human IGRT should be started as soon as the diagnosis is made. Patients treated with sufficient quantities (0.4 to 0.6 g/kg every 3 to 4 weeks for IVIG or 100 to 150 mg/kg every week for subcutaneous gammaglobulin [SQIG]) suffer few lower respiratory tract infections. However, these patients remain at risk for viral infections, including enteroviral meningoencephalitis. Since mucosal immunoglobulin cannot be replaced, the patients also remain at risk for recurrent upper respiratory infections, which may require prophylactic antibiotic therapy. Immunoglobulin-treated patients may lead normal lives without concern about exposure to infectious agents in childcare settings or classrooms.[15] Immunizations designed to elicit protective antibodies are unnecessary because the monthly replacement therapy will provide passive protection. Since patients are unable to mount antibody responses and vaccines, especially live vaccines, carry some risk of untoward side effects, they are relatively contraindicated.

An XLA patient who develops symptoms of enteroviral central nervous system or neuromuscular infection should have appropriate cultures of the involved organ system. For agammaglobulinemic patients with chronic enteroviral infections, immunoglobulin therapy should be given at higher doses and maintained until symptoms cease and the virus can no longer be detected. Intrathecal immunoglobulin as well as compassionate use of new antienteroviral drugs have also been used.[16]

AUTOSOMAL AGAMMAGLOBULINEMIA

Origin and Pathogenesis

The PreB-Cell Receptor and Signal Transduction Axis

Loss-of-function mutations in any one of the genes that code for components of the pre-BCR and its associated signaling complex can inhibit preB-cell development, leading to an absence of mature B cells.[17] This phenotype is seen in patients with biallelic loss-of-function mutations of mu heavy chain region (μ_0, AGM1), λ-like surrogate light chain (IGLL1, AGM2), immunoglobulin-associated alpha (Igα, CD79A, AGM3) and beta (Igβ, CD79B, AGM6) chains, and the adaptor B-cell linker protein (BLNK, AGM4), a key component of the pre-BCR signaling pathway.

E47

The *TCF3* gene, also called *E2A*, encodes two basic helix-loop-helix (bHLH) transcription factors, E12 and E47, through alternative splicing. E12 and E47 are involved in regulation of immunoglobulin gene expression. A dominant negative mutation in the E47 DNA-binding region resulted an early block in B-cell development and agammaglobulinemia.

LRRC8

Truncation of leucine-rich repeat containing 8 (*LRRC8, AGM5*), a gene of unknown function expressed in progenitor B cells, led to an absence of B cells.

PIK3R1

A homozygous nonsense mutation in *P1K3R1*, which encodes the p85a subunit of phosphoinositide 3-kinase (PI3K), resulted in decreased pro-B cell numbers. One 19-year-old female presented with agammaglobulinemia, absent B cells, and inflammatory bowel disease.

Other Agammaglobulinemia Conditions

Two pediatric patients with monosomy 7 presented with myelodysplastic syndrome of refractory anemia, hypogammaglobinemia, and low B cells. They were treated with bone marrow transplant.

Older adult patients with thymoma may present with low B-cell numbers and hypogammaglobinemia (*i.e.*, Good syndrome). They often suffer invasive bacterial infections and require immunoglobulin replacement even after thymectomy.[18]

Diagnosis and Treatment

Diagnosis requires gene mutation analysis. Treatment follows the guidelines given for XLA.

HYPER-IMMUNOGLOBULIN SYNDROME

Diagnosis

Patients with the hyper-IgM syndrome exhibit markedly reduced serum concentrations of IgG, IgA, and IgE with normal to elevated levels of IgM and normal numbers of circulating B cells.[19] This phenotype reflects polyclonal expansion of IgM synthesis in response to infection as a result of a block in the ability of B lymphocytes to switch from IgM to other isotypes. HIGM patients suffer the same infections with encapsulated bacteria common to all patients with antibody deficiency. HIGM can present as an X-linked recessive, autosomal recessive, or autosomal dominant trait. The phenotype can also result from congenital rubella, medications (e.g., calcineurin inhibitors and phenytoin), and neoplasia.[20]

Hyper-Immunoglobulin Syndrome Type 1: CD40L (CD154) Deficiency

Class switch recombination (CSR) is a multistep process that requires exquisite coordination between the B cell and its cognate helper T cell (Chapter 7). The binding of constitutively expressed CD40 on the B cell to CD40L (CD154) expressed on activated T cells helps initiate CSR. HIGM1 reflects loss-of-function mutations in *CD154* (Xq26).

Hyper-Immunoglobulin Syndrome Type 2: Activation-Induced Cytidine Deaminase dysfunction

Activation-induced cytidine deaminase (AID, 12p13), a member of the cytidine deaminase family, is required for CSR and for somatic hypermutation of immunoglobulin V domains (Chapters 4 and 7). The hyper-IgM phenotype can reflect either biallelic *AID* loss-of-function mutations or a dominant negative mutation on only one *AID* allele.

Hyper-Immunoglobulin Syndrome Type 3: CD40 Deficiency

CD40 is the cognate receptor for CD40L and is located on chromosome 20q12-q13.2. HIGM3 patients with biallelic loss-of-function mutations of *CD40* present with a phenotype indistinguishable from HIGM1.

Hyper-Immunoglobulin Syndrome Type 4: As Yet Unknown Causes

Patients presenting with a HIGM-like phenotype but lacking demonstrable mutations in genes previously associated with HIGM have been grouped into a category termed HIGM4. CSR is defective in HIGM4 patients, but somatic hypermutation appears to occur without hindrance. A genetic cause has yet to be identified and spontaneous recovery has been reported.

Hyper-Immunoglobulin Syndrome Type 5: Uracil-DNA Glycosylase Deficiency

Activation-induced deaminase (AID) acts by deaminating cytidine in DNA, leaving uracil in its place. Uracil-DNA glycosylase (UNG, 12q23–24.1) can remove the uracil, permitting normal or error-prone repair. Patients with biallelic loss-of-function mutations of *UNG* can present with recurrent bacterial infections, hyperplasia, increased serum IgM levels, and low IgG and IgA.

NEMO and IKBA Mutations

The nuclear factor kappaB (NF-κB) essential modulator (NEMO) plays a key role in the NF-kB pathway and consequently in the CD40 signal transduction pathway. NEMO acts as a scaffold for two kinases important for NF-κB activation. *IKBKG* encodes NEMO and is located on the X-chromosome (Xq28). Hypomorphic mutations in *IKBKG* lead to an ectodermal dysplasia syndrome manifesting with conical teeth, absent eccrine sweat glands, and a paucity of hair follicles. NEMO patients demonstrate defective innate and cell-mediated immunity due to defective NF-kB activation, which is important in Toll-like receptor (TLR) signaling. Laboratory abnormalities include impaired NK cell function, impaired pneumococcal responses, hypogammaglobulinemia, antigen-specific T-cell proliferative abnormalities, and increased serum IgA levels. Although fewer than 20% of NEMO patients have a hyper-IgM phenotype, they are typically grouped with the hyper-IgM phenotype.

Autosomal dominant mutations of *IKBA/IκBα* have been reported in some patients with a HIGM phenotype. Mutations in these genes block NF-κB release and thus downstream CD40 signaling.[21]

INO80 and MHSH6

INO80 and MSH6 are components of the mismatch repair machinery involved in CSR-induced generation of double-strand DNA breaks. INO80 deficiency has been described in patients with elevated IgM and low IgG and IgA who suffer with severe bacterial infections. Mutator S homolog 6 (MSH6) deficiency has been associated with elevated IgM, impaired class switching, and reduced numbers of class-switched memory B cells. Patents with MSH6 deficiency have described a personal or family history of cancer.[20]

Clinical Manifestations

CD40-CD154 Axis (HIGM1, HIGM3, NEMO)

Patients with inherited disruptions in the CD40-CD154 axis have difficulty generating germinal centers in their lymph nodes and spleen. Recurrent upper and lower respiratory tract infections are common. Patients may also exhibit recurrent neutropenia with oral ulcers, perirectal abscesses, and opportunistic infections (e.g., *P. jiroveci, Toxoplasma gondii*, or *Cryptosporidium* cholangitis). One in five patients demonstrate autoimmunity.

Without prophylaxis, one third of patients develop *P. jiroveci* pneumonia, which can be the presenting problem in affected infants. Patients are also at risk for cytomegalovirus (CMV), adenovirus, *Cryptococcus neoformans*, or mycobacterial infections. This indicates both cell-mediated and humoral immune deficiency, placing these patients within the spectrum of a combined immunodeficiency.

Chronic diarrhea occurs in more than half of the patients. Organisms include *Cryptosporidium, G. lamblia, Salmonella*, and *Entamoeba histolytica*. One quarter may require total parenteral nutrition due to diarrhea or to perirectal abscesses. Oral ulcers, gingivitis, and perirectal ulcers are associated with neutropenia, which may occur chronically or intermittently in up to two-thirds of the patients. One-fifth of patients develop sclerosing cholangitis that can lead to hepatic failure. Cryptosporidiosis is present in half of these patients.

Approximately one quarter of NEMO patients have autoimmunity or autoinflammation. An intestinal inflammatory disorder may be the presenting problem with chronic diarrhea and abdominal pain, with a few having steroid dependence.

Although originally distinguished by the high level of serum IgM, IgM levels are often normal in affected individuals. IgG is low. IgA and IgE are usually low, but they can be normal or even elevated. B- and T-cell counts are within the normal range in more than 90% of the patients and depressed in the rest.

Lymphoid hyperplasia is a common finding in patients with active infections. Individual nodes may become extremely large, and splenomegaly can develop. Hilar adenopathy causes a diagnostic dilemma, as the risk of lymphoma is increased. Although the lymphoid tissue is usually histologically abnormal, reactive processes are far more common than malignancy. Plasma cells may be abundant or sparse. Primary follicles are poorly developed. The most characteristic abnormality is the absence of germinal centers.

AID-UNG Axis (HIGM2 and HIGM5)

Infected AID-UNG deficient patients may present with giant germinal centers filled with highly proliferating B cells, presumably due to intense antigen stimulation. Approximately one-fourth of HIGM2 patients, but not HIGM5, present with evidence of autoimmunity (e.g., hemolytic anemia, thrombocytopenia, and autoimmune hepatitis). Autoantibodies are IgM.

Origin and Pathogenesis

CD40-CD154 Axis (HIGM1, HIGM3, and NEMO)

CD154, a member of the tumor necrosis factor (TNF) family, is a type II transmembrane protein. It is predominantly expressed on mature, activated CD4 T cells. Expression peaks at 6 to 8 hours post-activation and falls to resting levels by 24 to 48 hours. CD154 is also expressed on CD4 thymocytes, activated

CD8 T cells, NK cells, monocytes, basophils, mast cells, activated eosinophils, and activated platelets. Strongly stimulated neonatal T cells can also express CD154. CD40, a member of the TNF receptor superfamily, is constitutively expressed by pro-B, pre-B, and mature B cells, as well as on interdigitating cells, follicular dendritic cells, thymic epithelial cells, monocytes, platelets, and some carcinomas.

Engagement of B-cell CD40 with CD154 on an activated T cell that also expresses Fas ligand (FasL or CD95L) leads to the upregulation of Fas (CD95) on the B cell. NEMO is a part of the signaling pathway. If the B cell has concomitantly bound its cognate antigen and engaged the BCR-signaling pathway, it becomes resistant to Fas-mediated apoptosis and expresses CD80/CD86 on the cell surface. The activated B cell can then engage CD28 on the T-cell surface and trigger the T cell to secrete its cytokines. If the B cell fails to engage its BCR, the Fas pathway predominates and the B cell is eliminated. With proper activation of the CD40-CD154 pathway, exposure to interleukin-2 (IL-2) and IL-10 induces production of IgM, IgG1, and IgA, and exposure to IL-4 induces production of IgG4 and IgE. This change in immunoglobulin isotype reflects both induction of switching and the enhanced survival and proliferation of the B cell. Absent CD154, B cells can express IgM, but have difficulty switching and are likely to undergo apoptosis rather than proliferation in response to antigen.

Interactions between $CD154^+$ T cells and $CD40^+$ macrophages lead to enhanced production of IL-12, which then stimulates T cells to release interferon-γ. Activation of this pathway appears necessary for the defense against *P. jiroveci* and other opportunistic organisms.

Treatment and Prognosis

IGRT has improved the quality of life in HIGM. Adequate immunoglobulin replacement promotes reduction of serum IgM levels, prevention of infections with encapsulated bacteria, resumption of growth, and the gradual resolution of splenomegaly and lymphoid hyperplasia. Autoimmune and lymphoproliferative complications may respond to anti-CD20 therapy (e.g., rituximab).

Unfortunately, despite the improvement granted by IGRT, prognosis in patients with CD40-CD154 axis defects remains guarded. Deaths at a young age remain common, primarily the result of opportunistic infections, including pneumocystis pneumonia, cholangitis, CMV, mycobacterial infections, and cirrhosis secondary to hepatitis. Prophylaxis with trimethoprim-sulfamethoxazole can significantly reduce the risk of pneumocystis pneumonia and are indicated in those with CD40L and CD40 deficiency. Regular monitoring of gastrointestinal manifestations and management of neutropenia is mandatory. Neutropenia should be treated with a trial of GM-CSF since some have responded to this therapy. Bone marrow transplantation is a viable option for patients who fail to respond to supportive therapy.

SELECTIVE IMMUNOGLOBULIN A DEFICIENCY

Selective IgAD, selective IgG subclass deficiencies, CVID, and a syndrome of recurrent sinopulmonary infections (RESPIs) with normal serum immunoglobulin levels appear to share an overlapping set of gene defects.[22] Clinically, these disorders are marked by an increased susceptibility to upper and lower respiratory infections with encapsulated bacteria. IgAD and CVID feature similar B-cell differentiation arrests but differ in the extent of immunoglobulin deficits. The correlation between serum immunoglobulin levels and severity of infection is not absolute.[23]

Diagnosis

Approximately 1 in 600 individuals of European ancestry are unable to produce detectable quantities of IgA1 and IgA2, making selective IgAD the most frequently recognized primary immunodeficiency in the Americas, Australia, and Europe. The diagnosis is dependent on the sensitivity of the laboratory measurement. Nephelometry, for example, becomes unreliable for serum IgA levels of less than 7 mg/dL.

Uncomplicated patients with IgAD have normal serum levels of IgM, normal or elevated levels of IgG, and demonstrate normal cell-mediated immunity. A minority of patients may demonstrate additional evidence of immune dysfunction, with inability to generate appropriate IgG2 anti-carbohydrate antibodies, frank IgG subclass deficiencies, or evidence of impairment of T-cell function. Individuals with IgA serum levels that fall more than two standard deviations below the mean serum level for their age are considered to have partial IgAD. These individuals are usually healthy but can suffer from recurrent infections.

Clinical Manifestations

The likelihood that an IgA-deficient individual who was identified serendipitously will require medical attention is difficult to assess because most studies in the literature reflect patients who were ascertained due to clinical symptoms. Among IgAD patients referred to immunology clinics, more than 85% present with recurrent infections, typically with encapsulated bacteria. Among affected children, symptoms may begin in the first year of life, although the physiologic lag in serum IgA may delay the diagnosis until after the age of 2. In some patients, respiratory infections disappear with maturity. In others, infections may persist throughout adult life. Rarely, IgAD patients may experience recurrent bronchitis, pneumonia, and even bronchiectasis. These more severely afflicted patients often exhibit concurrent IgG2 and IgG4 subclass deficiencies. Some symptomatic patients have elevated IgE levels and manifest allergic or asthmatic components to respiratory dysfunction. The rise in IgE has been explained as compensatory to the absence of IgA. This appears to be a double-edged sword, because up to 20% of patients complain of allergic rhinitis, conjunctivitis, urticaria, and atopic eczema. Allergic reactions may be enhanced due to the lack of IgA-blocking antibodies in the serum, and unusually severe asthma has also been reported.

Among those less common IgA-deficient patients that are truly devoid of IgA, as many as three-fifths produce IgG or IgE anti-IgA antibodies.[24] These uncommon patients are at an uncertain risk for adverse reactions following transfusion with blood products, plasma from normal donors, or from some preparations of human immunoglobulin, which, of course, contain IgA. Patients with high anti-IgA levels (greater than 1:1000) typically have potent antibodies directed against all IgA. These patients are at risk for severe anaphylaxis. Patients with low anti-IgA antibody titers (less than 1:256) are often multiparous or multi-transfused patients. These patients rarely demonstrate severe anaphylaxis after infusion with plasma or blood products, but do present with hives and rashes.

IgAD patients often develop autoimmune diseases. Gastrointestinal disorders include pernicious anemia, inflammatory bowel disease, intestinal disaccharidase deficiency, lactase

deficiency, pancreatic insufficiency, and celiac disease. The latter in particular can be difficult to diagnose without biopsy since serologic diagnosis often relies on detection of anti-tissue transglutaminase, anti-endomysial or anti-gliadin IgA antibodies.[25] Hepatobiliary disorders include chronic active hepatitis, cholelithiasis, lupoid hepatitis, and primary biliary cirrhosis. Skin disorders include pyoderma gangrenosum, paronychia, and vitiligo. It is unclear whether this autoimmune diathesis is the end result of recurrent infections, the product of recurrent insult by antigens that would otherwise be cleared by IgA, or whether the underlying deficit that leads to IgAD also increases the risk of developing an autoimmune disorder. For example, autoimmune disorders such as insulin-dependent diabetes mellitus and celiac disease are associated with the same major histocompatibility complex (MHC) haplotypes (Chapter 5) as IgAD and CVID.

IgAD is associated with an increased risk for the development of malignancies, including epithelial tumors (e.g., gastric and colonic adenocarcinoma) and lymphoproliferative disorders (e.g., Hodgkin disease and acute lymphoblastic leukemia). Patients with chronic gastrointestinal infections may demonstrate a nodular, small intestine lymphoid hyperplasia that can lead to intestinal obstruction. Active B lymphocyte proliferation in the germinal centers of the Peyer patches is seen. These "constipated" lymph nodes have been mistaken for lymphoma. In others, the simultaneous presence of IgAD and malignancy may simply reflect the high prevalence of IgAD in the Caucasian population.

Origin and Pathogenesis

IgAD, selective IgG subclass deficiencies, and CVID are diseases that are defined by a quantitative phenotype, a paucity of serum immunoglobulins of a given isotype in spite of the presence in the blood of B lymphocytes bearing the missing isotypes. By definition, the fundamental defect involves the failure of B lymphocytes bearing a given isotype to differentiate into plasma cells. These diseases appear to represent a common endpoint for multiple pathogenic processes. All three phenotypes may be acquired and many of the recognized precipitating causes, such as phenytoin, are the same (Table 33.3).

IgAD is associated with MHC haplotypes (6p21.3) that are more common in European populations than in the peoples of sub-Saharan Africa and East Asia. In the United States, the prevalence of IgAD among African Americans is one-twentieth of that observed among Americans of European descent and in Japan the incidence is approximately 1 in 18,500. IgAD has also been observed in family members of CVID patients with altered function of the transmembrane activator and CAML interactor (TACI, 17p11.2), which is a receptor for B-cell activating factor (BAFF).

Treatment and Prognosis

Most individuals with IgAD suffer respiratory infections no more frequently than the average individual, and thus require no special treatment. All individuals with IgAD should be warned of the risk of serious transfusion reactions caused by antibodies to IgA. Wearing a medical alert bracelet is recommended. Should transfusion be necessary, the ideal donors are other individuals with IgAD. Washed erythrocytes are safer than whole blood.

Patients with selective IgAD who suffer from clinically significant, recurrent upper respiratory infections often respond to prophylactic antibiotics with potency against encapsulated bacteria. Treatment of allergy in those patients with a compensatory increase in IgE is helpful. Patients who present with combined IgA and IgG subclass deficiencies and have a poor pneumococcal antibody response may require IGRT.

TABLE 33.3 Other Conditions Associated With Humoral Immunodeficiency

Genetic disorders	
Monogenic diseases	Ataxia-telangiectasia
	Autosomal forms of SCID
	Transcobalamin II deficiency and hypogammaglobulinemia
	Wiskott-Aldrich syndrome
	X-linked lymphoproliferative disorder (EBV associated)
	X-linked SCID
Chromosomal anomalies	Chromosome 18q- syndrome
	Monosomy 22
	Monosomy 7
	Trisomy 8
	Trisomy 21
Systemic disorders	
Malignancy	Chronic lymphocytic leukemia
	Immunodeficiency with thymoma
	T-cell lymphoma
Metabolic or physical loss	Immunodeficiency caused by hypercatabolism of immunoglobulin
	Immunodeficiency caused by excessive loss of immunoglobulins and lymphocytes
Environmental exposures	
Drug-induced	Antimalarial agents
	Captopril
	Carbamazepine
	Glucocorticoids
	Fenclofenac
	Gold salts
	Imatinib
	Levetiracetam
	Penicillamine
	Phenytoin
	Sulfasalazine
	Zonisamide
Infectious diseases	Congenital rubella
	Congenital infection with CMV
	Congenital infection with *Toxoplasma gondii*
	Epstein-Barr virus
	Human immunodeficiency virus

CMV, Cytomegalovirus; *EBV*, Epstein-Barr virus; *SCID*, severe combined immunodeficiency.

COMMON VARIABLE IMMUNODEFICIENCY AND COMMON VARIABLE IMMUNODEFICIENCY-LIKE DISORDERS

Diagnosis

The diagnostic category of CVID includes a heterogeneous group of patients older than age 4 who exhibit deficient production of more than one major antibody class and whose antibody response to vaccination is significantly depressed or absent. Patients tend to have normal numbers of clonally diverse B lymphocytes in their blood. These B cells can recognize antigens and respond with proliferation, but their ability to develop into memory B cells or mature plasma cells appears quantitatively impaired. In the presence of infection, abortive differentiation can lead to massive B-lymphocyte hyperplasia, splenomegaly, and intestinal lymphoid hyperplasia.

With an estimated prevalence of 1 in 25,000, CVID is the most prevalent human primary immunodeficiency requiring medical attention.[26] Men and women are equally affected. As with IgAD, the prevalence among African Americans is one-twentieth that of Americans of European descent. Some patients present during childhood, but most are diagnosed after the third decade of life. The typical patient reports a normal pattern of recurrent otitis media as an infant and toddler that resolved in childhood. During adolescence, respiratory infections appear and steadily increase in frequency and duration. Recurrent pneumonia as a young or middle-aged adult is often the precipitating complaint that brings the patient to the attention of the clinical immunologist. Although CVID appears to be an acquired disorder, family studies have clearly documented that susceptibility for the disease can be inherited and the manifestations of the disorder may change with time. Transitions within the spectrum of normal serum immunoglobulin concentrations to IgAD to IgAD with IgG subclass deficits to frank CVID have been documented in both sporadic and familial cases.[2]

CVID is a diagnostic category of primary immunodeficiencies that includes a number of immune disorders. Most CVID patients of Northern European descent exhibit a distinctive phenotype characterized by a broad deficiency of immunoglobulin isotypes in spite of the presence of normal numbers of surface immunoglobulin bearing B-cell precursors in the peripheral blood. Almost all of these patients are IgA deficient and, by definition, demonstrate total serum IgG levels of less than 500 mg/dL. Some IgG subclasses are more affected than others, with the sequential order of involvement being IgG4 > IgG2 > IgG1 > IgG3. Most patients are also deficient in IgM and IgE.

Uncomplicated patients demonstrate normal cell-mediated immunity, although a minority of patients may have evidence of T-cell dysfunction as well as other hematopoietic cell types. In some cases, B-cell numbers are reduced, although not to the extent exhibited by disorders of pre-BCR formation or BTK signaling.

IgAD and CVID have been associated with congenital infection with rubella virus, CMV, and *T. gondii*. The administration of certain drugs has also been linked to a depression in serum immunoglobulin levels (see Table 33.3). Several medications used to treat epilepsy have been associated with the development of antibody deficiencies. For example, up to 20% of patients treated with phenytoin suffer a mild decrease in serum IgA levels, and a minority may progress to a CVID-like phenotype. Medications used for the treatment of rheumatoid arthritis, inflammatory bowel disease, and chronic myelogenous leukemia can also decrease production of antibody. Persistence of antibody deficiency usually requires continued administration of the drug or continued infection with the virus or parasite. However, recovery of immunoglobulin production may take months to years.

Clinical Manifestations

Although some CVID patients have reduced numbers of circulating B cells, the majority have normal quantities of IgA, IgG, and IgM-bearing B-cell precursors in the blood. Defects in B-cell survival, number of circulating $CD27^+$ memory B cells (including IgM^+CD27^+ B cells), B-cell activation after antigen receptor cross-linking, T-cell signaling, and cytokine expression have been observed. Both increases and decreases in the relative numbers of CD4 to CD8 T cells are common, and cutaneous anergy is a frequent finding.

The clinical manifestations of CVID are similar but more severe than the ones seen in IgAD. Respiratory symptoms often begin with recurrent sinusitis, otitis media, and mild bronchitis. The frequency and severity of the upper respiratory infections worsen in the young adult and lower respiratory infections such as pneumonia become common. Apparently asymptomatic, untreated patients may suffer recurrent subclinical pulmonary infections that can lead to irreversible chronic lung damage with bronchiectasis, unilateral hyperlucent lung, emphysema, and cor pulmonale. Recurrent cellulitis, boils, folliculitis, impetigo, or erythroderma can be presenting complaints.[27]

Intermittent or chronic diarrhea due to *G. lamblia* is a common complaint. Patients can develop a malabsorption syndrome that resembles celiac sprue but is unresponsive to gluten avoidance (Fig. 33.4). Untreated patients often complain of an asymmetrical, oligoarticular arthralgia or frank arthritis, which in some cases reflect infections with encapsulated organisms or with *Mycoplasma* species and thus require antibiotic therapy. Paradoxically, antigen-specific IgE can be produced in sufficient quantities to enable anaphylactic reactions.

CVID patients are often anergic, but only a minority suffer infections characteristic of cell-mediated immune dysfunction (e.g., mycobacteria, *P. jiroveci*, or fungi). CD8 T-cell numbers may be depressed in such patients. Most viral infections are cleared normally. Exceptions include enteroviral infections, including meningoencephalitis, as well as hepatitis B and C, which can progress to chronic active hepatitis. Lack of humoral immunity enhances susceptibility to viral reactivation. Untreated patients often complain of recurrent herpes zoster (shingles).

Autoimmune diseases are common in CVID. Coombs-positive hemolytic anemia with idiopathic thrombocytopenic purpura, a combination known as Evans syndrome, may predate diagnosis.

Non-caseating granulomas in the lung, lymph nodes, skin, bone marrow, and liver reminiscent or indistinguishable from sarcoid-like syndrome are seen in up to one-fifth of all patients and are more common in African Americans. While the granulomas can result from mycobacterial or fungal infections, in most cases the cause remains unclear and the granulomas resolve spontaneously.

There is an increased risk for the development of gastrointestinal (1.5%) and lymphoid malignancies (4.1%), especially non-Hodgkin lymphomas.[28] A confounding factor is a propensity to develop benign lymphoproliferative disorders. Lymphadenopathy, splenomegaly, or both are common in untreated patients.

Origin and Pathogenesis

The typical presenting manifestation of CVID is hypogammaglobulinemia, not agammaglobulinemia, suggesting a partial or varying block in B-cell maturation. Careful analysis of B cells in patients has also revealed a spectrum of immune deficiency ranging from the nearly complete absence of memory B cells to a less severe disorder. All of these findings serve to underline the complex etiology for the disorder, and many details remain to be elucidated. The MHC represents the most common genetic susceptibility locus for CVID.[22] Due to linkage disequilibrium, the gene, or genes, within this locus that create susceptibility have yet to be identified with certainty. However, a co-association between MHC class I and KIR alleles has been reported.[29]

Although they represent only a minority of patients, non-MHC associated single-gene defects have been identified.

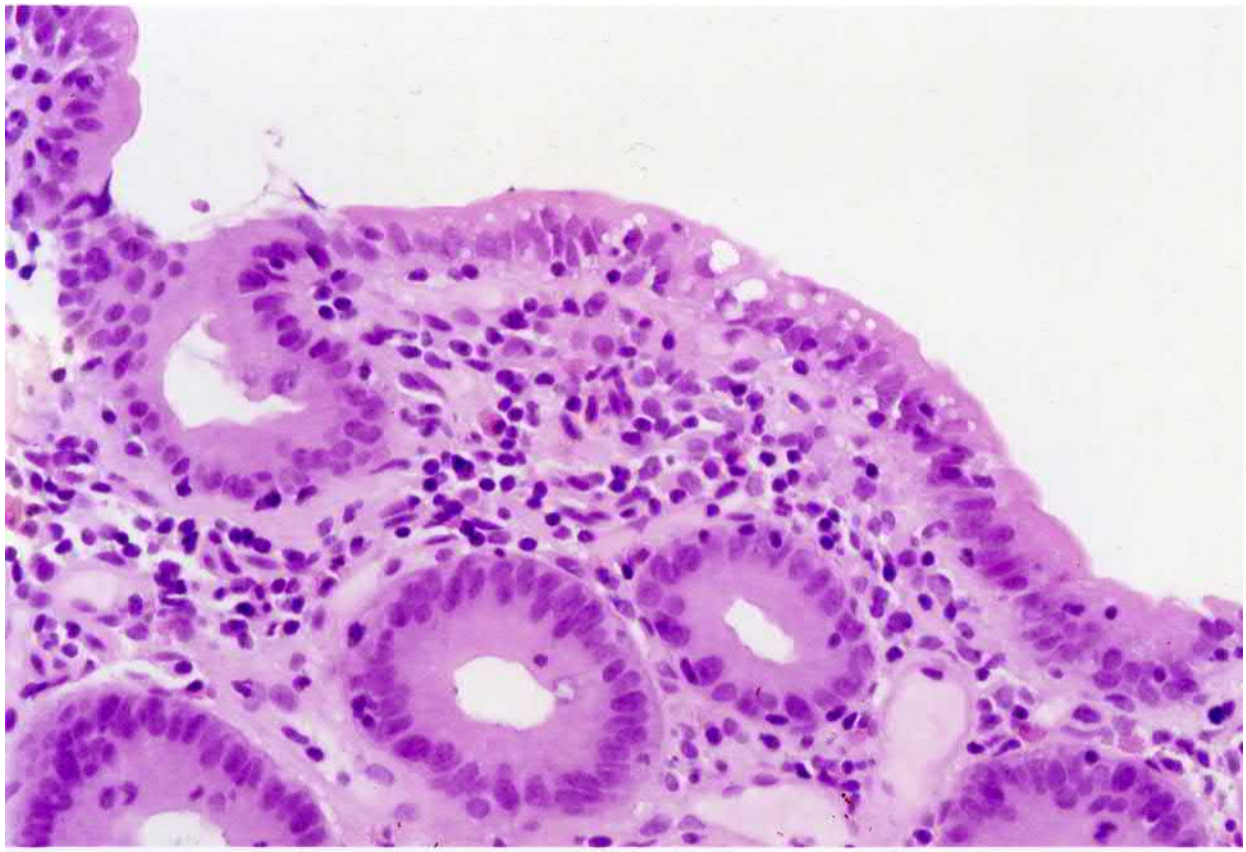


FIG. 33.4 Hypogammaglobulinemic Sprue in a 41-Year-Old White Male With Common Variable Immunodeficiency and Insulin-Dependent Diabetes Mellitus. The patient suffered from intractable diarrhea. Shown is a hematoxylin and eosin stain of a duodenal biopsy obtained by endoscopy. The villi are blunted and there is a marked increase in intraepithelial lymphocytes. However, unlike typical celiac disease, the villi are not completely blunted and few plasma cells are seen. The patient is homozygous for the HLA-DQ2, -DR17 (3), -B8 haplotype. Although the patient failed to respond to a gluten-free diet, the diarrhea resolved with corticosteroid therapy.

Included are loss-of-function mutations of genes involved in late-stage B-cell:T-cell communication, late-stage B-cell growth factors, and B-cell and T-cell signaling and activation pathways (see Table 33.1). Among these are genes for ICOS (CVID1), an immune costimulatory molecule used by T cells to activate B cells in germinal centers; BAFFR (CVID4) and TACI (CVID2), the receptors for B-cell activating factor (BAFF-R also known as CVID4), CD19 (CVID3), CD21 (CVID7) and CD81 (CVID6), components of the B-cell costimulatory receptor; CD20 (CVID5), an important marker of B-cell differentiation; and LRBA(CVID8), CTLA-4, PKCδ (CVID9), TWEAK, PIK3CD, PIK3R1, NF-κB2 (CVID10), IL-21 (CVID11), NF-κB1 (CVID12), IKAROS (CVID11), IRF2BP2 (CVID14), BLK, PTEN, TRNT1, ATP6AP1, ARHGEF1, SH3KBP1, SEC61A1, RAC2, MOGS (mannosyl-oligosaccharide glucosidase), and KMT2D which are involved in B-cell and T-cell signaling pathways.[30] TLR7 and TLR9 activation can be deficient in these patients, although the genes are intact.[31]

The Major Histocompatibility Complex

A large array of genes that play important roles in the control of the immune response are located in the MHC on chromosome 6 (Chapter 5). Many IgAD and CVID patients share parts or all of one of two extended MHC haplotypes marked by either HLA-DQ2, -DR17(3), -B8 and HLA-DQ2, -DR7,-B44. One in seven individuals homozygous for HLA-DQ2, -DR17 (3), -B8 demonstrate IgAD. These MHC alleles are also observed in patients with diabetes mellitus, pernicious anemia, celiac disease, autoimmune thyroid disease, and myasthenia gravis. Some individuals with TACI mutations inherited MHC haplotypes associated with the disease, and the combination of specific MHC and KIR alleles further increases susceptibility.[29] This suggests that epistatic interactions between different susceptibility alleles may influence development of the disorder.

The CD19 (CVID3), CD81 (CVID6), CD21 (CVID7) B-Cell Co-Receptor Complex

CD21 (complement component C3d/Epstein Barr virus receptor 2) binds to membrane IgM-bound antigen when complement C3d is also attached to that antigen (Chapter 40). In association with CD81 and CD19, this co-receptor complex enhances the antigen-binding signal, promoting B-cell activation. Patients with mutations in CD19, CD21, and CD81 have been reported.[32]

The BAFF, BAFFR (CVID4), and TACI (CVID2) axis

The TNF family members B-cell activating factor of the TNF family (BAFF) and a proliferation-inducing ligand (APRIL) bind to two receptors, B-cell maturation antigen (BCMA) and transmembrane activator and calcium-modulator and cyclophilin

ligand interactor (TACI). BCMA is expressed exclusively on B cells, whereas TACI is expressed on activated T cells as well. A third receptor unique for BAFF, BAFF-R, is expressed on B cells and on resting T cells. The BAFF/APRIL system plays a key role in mature B-cell homeostasis and development. BAFF and APRIL can also induce class switching in naive human B cells. Loss-of-function (autosomal recessive) or altered-function (autosomal dominant) TACI alleles have been reported in up to 10% of CVID patients.[33] However, two polymorphic TACI alleles in CVID, A181E, and C104R, are also present in approximately 2% of the normal population. Family members may have IgAD or may have no evidence of immune dysfunction, suggesting reduced penetrance. CVID patients with these altered alleles have a higher prevalence of complications from CVID, including lymphoproliferation, splenomegaly, and autoimmune phenomena.[34] BAFF-R deficiency has also been detected in CVID.[32]

CD20 (CVID5)

CD20 encodes a B-cell membrane-spanning molecule important in B-cell proliferation and differentiation. One female consanguineous patient with CD20 deficiency has been reported with a low IgG and normal IgA and IgM levels, and with impaired antibody responses to pneumococcal polysaccharides.[32]

ICOS (CVID1)

ICOS is a T-cell surface receptor that is important for germinal center formation, terminal B-cell differentiation, effector T-cell responses, and immune tolerance. ICOS-deficient patients have low to absent B cells and some have varying degree of defective T-cell signaling. They present with recurrent respiratory tract infections and autoimmune complications.[32,35]

LRBA (CVID8) and CTLA-4 axis

Lipopolysaccharide-responsive beige-like anchor protein (LRBA) is a cytosolic protein that functions in vesicle trafficking, autophagy, and cell survival. CTLA-4 is an inhibitory T-cell receptor that competes with the costimulatory protein CD28 for binding CD80/86, thereby preventing excessive T-cell activation and maintaining immune tolerance. LRBA plays a role in CTLA-4 surface expression.

LRBA-deficient patients display early-onset hypogammaglobulinemia with autoimmunity and inflammatory bowel disease. They show reduced levels of at least two immunoglobulin isotypes (IgM, IgG, or IgA) and suffer from recurrent infections, autoimmunity, and chronic pulmonary and gastrointestinal disorders.

Patients with haploinsufficiency of CTLA-4 demonstrate autoimmunity, recurrent infections, benign lymphoproliferation, and varying levels of immunoglobulins, B-cell, and T-cell defects.[32]

Interleukin-21

Interleukin-21 (IL-21) deficiency was found in a consanguineous family where three of eight children had inflammatory bowel disease. Hypogammaglobinemia, poor specific antibody responses to polysaccharide antigens, increased IgE, and decreased memory and switched B-cell numbers were observed.[36]

PKCδ Deficiency

PKCδ plays a key role in BCR-mediated signaling downstream of BTK and is important in B-cell proliferation, apoptosis, and tolerance. PKCδ deficiency presents with a variable phenotype with one affected patient having CVID-like characteristics (hypogammaglobulinemia and severe infections) whereas the others had lupus or ALPS-like disease.[32]

TWEAK Deficiency

A patient with an autosomal dominant mutation in TNF superfamily member 12 (TNFSF12), which encodes TNF-like weak inducer of apoptosis (TWEAK), displayed low to normal IgG, low IgM, and low IgA. There was a history of pneumococcal meningitis, osteomyelitis, thrombocytopenia, and neutropenia.[32]

NF-κB1 (CVID12) and NF-κB2 (CVID10) Deficiency

The NF-κB1 and NF-κB2 (non-canonical) pathways are important in B-cell signaling, with NF-κB2 having a more limited set of involved receptors (e.g., ICOS, TACI, BAFR-R, and BCMA), whereas NF-κB1 affects T-cell and TLR signaling as well.

Heterozygous mutations in NFK-B2 have been identified in patients who presented with early-onset panhypogammaglobulinemia, autoimmunity, and RESPI. These patients display aberrant B-cell subsets, some degree of T-cell and NK-cell dysfunction, and pituitary hormone deficiencies.

Patients with NK-κB1 autosomal dominant mutations that create unstable protein have recurrent infections, autoimmunity, benign lymphoproliferative disease, and lymphoma.[32]

PI3K Mutations

Heterozygous mutations in *PIK3CD*, which encodes the PI3K catalytic subunit p110δ, have been reported in patients with respiratory tract infections, skin infections, autoimmunity, and lymphoma. The mutation leads to overactive PI3K signaling. The phenotype associated with dominant gain-of function *PIK3CD* mutations is known as activated phosphatidylinositol 3-kinase δ syndrome (APDS).

PIK3R1 encodes the P13K regulatory subunit p85α. A dominant gain-of-function mutation of p85a also results in autosomal dominant overactive PI3K signaling. A patient with a complete loss of p85a expression has presented absence of B cells and agammaglobulinemia.

BLK, IRF2BP2 (CVID13), and IKAROS (CVID14)

A heterozygous loss-of function mutation in BLK has been noted in related patients with CVID. These patients have respiratory tract infections and bacterial skin infection with pan hypogammaglobulinemia. A gain of function mutation in IRF2BP2 has been found in members of a family with CVID. These patients have autoimmune disease and respiratory tract infections. Heterozygous mutations in *IKZF1*, which encodes the transcription factor IKAROS, has been observed in patients with a progressive loss of B cells and serum immunoglobulins.[32]

Other Genes

Among autosomal dominant conditions, PTEN deficiency has been reported in one young boy with hypogammaglobulinemia and defective specific antibody responses. Mutations in *SEC61A1* have been reported in hypogammaglobinemia with RESPI. Defects in Ras-related C3 botulinum toxin substrate 2 (RAC2) genes are associated with reduced antibody levels and decreased antibody responses and recurrent respiratory infections with normal or low B-cell numbers. PLCγ2-associated antibody deficiency and immune dysregulation (PLAID) results from gain-of-function mutations and presents with cold urticaria, hypogammaglobulinemia, impaired humoral immunity, and autoinflammation.

Among autosomal recessive conditions, TRNT1 deficiency has been reported as the cause of B-cell deficiency with hypogammaglobinemia, as well as deafness and developmental delay. *ARHGEF1* mutations have been associated with hypogammaglobinemia, recurrent infections, and bronchiectasis. Mutations in *MOGS* lead to mannosyl-oligosaccharide glucosidase deficiency, also known as congenital disorder of glycosylation type IIb (CDG-IIb), with panhypogammaglobulinemia, subpar specific antibody responses, and increased susceptibility to bacterial and viral infections.

Among X-linked conditions, mutations in *ATP6AP1* and *SH3KBP1* have been associated with immunoglobulin deficiency.

Kabuki syndrome

Patients with Kabuki syndrome (KS) present with a characteristic face, short stature, cardiac anomalies, a variable degree of intellectual disability, recurrent infections, reduced immunoglobulin levels, and autoimmunity. Mutations in *KMT2D* and *KDM6A* have been identified as the main cause.

Treatment and Prognosis

Therapy in CVID begins with the aggressive treatment of ongoing infections and the institution of prophylactic measures to prevent or ameliorate future infections. Patients suffering from moderate upper respiratory tract infections and bronchitis typically benefit from empiric therapy with agents effective against encapsulated organisms. The concomitant inheritance of mannose-binding lectin protein deficiency further predisposes to the development of bronchopulmonary complications such as bronchiectasis, lung fibrosis, and respiratory insufficiency,[3] as well as urinary tract infections. Prolonged and intravenous administration of antibiotics may be required.

The most effective therapy for hypogammaglobulinemic patients is immunoglobulin replacement. A number of studies have demonstrated a steadily decreasing incidence of infection and improvement in lung function with increasing doses of immunoglobulin administration.[37] Each patient may demonstrate his or her own individual response to therapy, exhibiting dramatic differences in the frequency and severity of infections with moderate changes in the replacement dose. Ultimately, replacement dosage must be individualized based upon the response of the patient. Patients suffering from a serious acute infection often benefit from one-time booster doses of immunoglobulin. Adverse reactions occur most frequently at the time of the first administration of immunoglobulin, likely because of concurrent infection increasing the potential for generation of immune complexes.

Some patients with CVID can sustain severe anaphylaxis when given IVIG or other blood products that contain serum or plasma. These patients may possess anti-IgA antibodies, including IgE anti-IgA antibodies.[7] For patients with a history of severe adverse reactions, it is advisable to try lots of IVIG with the lowest IgA possible and to test the patient with the different lots in an intensive care unit. Once having identified a lot that can be tolerated, the patient may receive therapy under more relaxed conditions.

Serum immunoglobulin concentrations in patients with CVID may change over time,[2] with rare patients regaining normal serum IgG levels and no longer requiring immunoglobulin therapy. Careful review of the clinical history of these patients may reveal evidence of exposure to pharmacologic agents associated with the development of hypogammaglobulinemia (e.g., phenytoin). However, the overwhelming majority of patients require replacement therapy for life.

Although IgG may be replaced, at present IgM and IgA cannot be provided to the patient. The absence of these multimeric proteins may help explain why even patients on high-dose replacement therapy may continue to suffer from recurrent sinusitis or gastrointestinal discomfort.[38] Recurrent sinusitis can be ameliorated with continued prophylactic therapy with antibiotics. Patients with CVID also are at risk from *G. lamblia*, as well as other enteric pathogens. Some patients develop lactose intolerance or gluten-sensitive enteropathy. Gluten avoidance ameliorates symptoms in only a minority of cases. A majority responds to corticosteroids or anti-TNF agents. The use of these agents can be a double-edged sword, since resistance to infection will decrease in a patient who is already immune deficient. Other patients develop a malabsorption syndrome that can lead to hypoalbuminemia and hypocalcemia (due to malabsorption of vitamin D), and decreased levels of vitamin A and carotene.[39] The cause of diarrhea and malabsorption in this latter patient subset remains unclear, and treatment is limited to supportive measures, with vitamin and mineral replacement as indicated.

Patients with bronchiectasis should be treated aggressively with replacement therapy. In severe cases, aggressive pulmonary toilet will benefit the patient, including bronchodilator therapy, position and postural drainage, or other physical therapies.

IgA-deficient mothers may fail to secrete IgA in their colostrum. Although colostral IgM levels may be elevated in an attempt to compensate for the lack of maternal IgA, the newborn remains relatively unprotected against intestinal pathogens. Of greater concern are the children of mothers with untreated CVID who are born in a state of humoral immunodeficiency and thus are at risk for life-threatening sinopulmonary infection. In order to compensate for the loss of IgG across the placenta and to provide the infant with the passive immunity it will require, the dose of replacement gammaglobulin therapy should be increased by 50% by the third trimester of pregnancy.

Splenomegaly is common in untreated patients. Hypersplenism in most patients responds to aggressive therapy with antibiotics and intravenous immunoglobulin. The presumption is that the hypersplenism is secondary to reactive hyperplasia of lymphoid follicles within the spleen attempting to respond to infection. Development of esophageal varices or other hematologic manifestations of hypersplenism (refractory thrombocytopenia, anemia, neutropenia, and lymphopenia) may require splenectomy as the therapy of last resort. The outcome for most such patients has been good, with resolution of symptoms, although patients with altered TACI alleles tend to do less well.

The development of constellation of pulmonary abnormalities that include granulomatous and lymphoproliferative (lymphocytic interstitial pneumonia [LIP], follicular bronchiolitis, and lymphoid hyperplasia) histopathologic patterns, termed granulomatous-lymphocytic interstitial lung disease (GLILD), can be an ominous sign. These patients appear more likely to develop granulomatous liver disease, autoimmune hemolytic anemia, lymphoproliferative disease, and progressive pulmonary disease.[40]

THERAPEUTIC PRINCIPLES

- The primary goal of treatment is to keep the patient infection free and maintain lung function.
- In patients whose respiratory mucosa is intact, intravenous or subcutaneous replacement IgG therapy is generally effective in protecting them from pulmonary infections.
- For those patients who have developed bronchiectasis or who continue to subject themselves to environmental toxins (e.g., smoking), replacement IgG will ameliorate but may not prevent all such infections.
- Because mucosal immunoglobulin cannot be replaced, even patients on adequate IgG replacement therapy remain at risk for sinus or gastrointestinal infections.
- Prophylactic antibiotics that are effective against encapsulated organisms can significantly reduce the frequency of upper respiratory tract infections in patients who continue to suffer despite replacement therapy with intravenous gammaglobulin.
- Prolonged diarrhea in hypogammaglobulinemic patients may be caused by *Giardia lamblia* and responds well to metronidazole therapy.
- Patients with primary antibody deficiencies should not receive live vaccines.

SELECTIVE IMMUNOGLOBULIN G SUBCLASS DEFICIENCIES

Diagnosis

Most individuals with modest reductions in serum IgG subclass levels are functionally normal. Indeed, individuals with deletions of the heavy chain immunoglobulin gene locus, some of whom completely lack IgG1, IgG2, IgG4, and IgA,[41] have been reported to be asymptomatic. The diagnosis of a functional IgG subclass deficiency can only be made with confidence when there is both a significant decrease in the serum concentration of a specific isotype and clear evidence of reduced specific antibody production (e.g., failure to respond to Pneumovax [R] 23).

Up to 10% of normal males and 1% of normal females are IgG4 deficient, which makes a diagnosis of isolated IgG4 subclass deficiency problematic. Among patients with IgG1 or IgG3 deficiency, documentation of the ability to produce protective titers of anti-tetanus toxin and anti-diphtheria toxin antibodies following standard tetanus toxoid and diphtheria immunizations is a strong indication that replacement gammaglobulin therapy is likely unwarranted, especially if pulmonary function is normal and there is no history of recurrent infection. Documentation of a strong anti-pneumococcal polysaccharide response in patients with an apparent IgG2 deficiency would suggest gammaglobulin replacement is likely not required. Conversely, the lack of a response to vaccination calls for appropriate prophylactic antibiotic therapy before a trial of IVIG is attempted.

Clinical Manifestations

The clinical spectrum of isolated IgG subclass deficiency is quite variable and deficiencies of each of the four IgG subclasses have been described.[42] Some individuals present with only a mild reduction of total IgG, but most symptomatic patients have marked deficiencies of one or more IgG subclasses despite normal total IgG concentrations. Since IgG1 makes up the majority of serum IgG in most patients, a deficiency of IgG tends to correlate with depressed serum levels of total IgG.

IgG subclass levels are rarely measured in asymptomatic individuals. Most patients with isolated IgG2 deficiency present with recurrent upper or lower respiratory infections. Individuals may have few residual symptoms between infections, but some have severe chronic inflammation with refractory sinusitis, pulmonary fibrosis, or bronchiectasis. Because protective antibodies directed against carbohydrate antigens are usually of the IgG2 subclass, many affected patients exhibit an impairment of their ability to mount specific protective responses to encapsulated pathogens. However, normal responses have also been described.[43] There is general agreement that recurrent sinopulmonary infections (RESPI), IgG2 deficiency, and a response to less than half of the polysaccharide antigens in an unconjugated vaccine meets the standard for functional immune deficiency and thus warrants aggressive prophylactic therapy up to and including immunoglobulin replacement.

IgG3 deficiency may occur alone or in association with IgG1 deficiency. Recurrent infection of the respiratory tract with chronic lung disease has been reported. With a serum half-life of only 2 weeks, IgG3 levels may be consumed rapidly during the course of an active infection in an otherwise normal individual.[44] Before making the diagnosis of IgG3 deficiency, serum levels of IgG3 should be re-checked when the individual is asymptomatic.

When compared with the serum, IgG4 is over-represented in secretions, and IgG4-committed B cells are present at mucosal sites, suggesting a role in mucosal immunity. Since IgG4 is normally present in the serum in very low concentrations, the significance of a low serum level in a patient with recurrent infection remains unclear.

Origin and Pathogenesis

The origin of IgG subclass deficiency is unknown. Homozygous deletions of portions of the immunoglobulin heavy chain constant locus associated with total absence of IgG2, IgG3, and IgG4 or combinations of these isotypes have been described in healthy individuals. IgG2 deficiency is often found in association with selective IgAD with or without IgG4 deficiency. Patients with selective IgG subclass deficiencies often inherit the same MHC haplotypes seen in IgAD and CVID. Thus, IgG-subclass–deficient patients with recurrent infections likely have a more complex defect than just the absence one or more IgG isotype. In some instances, subclass deficiency is associated with a T-cell defect, as in chronic mucocutaneous candidiasis and ataxia-telangiectasia. IgG subclass deficiency may also be acquired. Acute infections, medications, chemotherapy, irradiation, surgery, and HIV infection have all been temporally linked to the development of a deficiency in one or more IgG subclass.[45]

Treatment and Prognosis

The natural history of IgG subclass deficiency (±IgAD), especially in children, varies.[46] Associated allergic rhinosinusitis and asthma must be aggressively treated with conventional therapy, as these conditions increase the risk of purulent sinusitis and pneumonia. Evaluation of possible anatomic obstruction should be performed when persistent infection of a sinus or pulmonary segment is the presenting complaint; the role of surgical therapy should not be overlooked.

Many patients with IgG subclass deficiency do well on prophylactic antibiotics and will never need immunoglobulin supplementation. However, in patients with severe, recurrent infections, IGRT can be beneficial. Patients who begin therapy should improve within the first 2 months, but to avoid the placebo effect, a full 6-month trial is recommended.

ANTIBODY DEFICIENCY WITH NORMAL SERUM IMMUNOGLOBULIN LEVELS

Occasional patients may present with normal serum immunoglobulin concentrations and a selective inability to respond to infections with pyogenic organisms. Diagnosis requires documentation of an inability to respond to antigenic challenge. These patients may respond to replacement immunoglobulin therapy if prophylactic antibiotics fail and aggressive management of other conditions such as asthma fail.[47,48] The antibody response to specific polysaccharide antigens can be very selective. In humans, most protective anti-*Haemophilus influenzae* type b (anti-Hib) antibodies utilize the rare Vκ A2 gene. The Navajo population in the Southwestern United States suffers a 5- to 10-fold increased incidence of Hib disease. This population also exhibits a high prevalence of an A2 allele with a defective recombination signal sequence, preventing use of germline-encoded antibodies that can generate protective antigen-binding sites.

Analysis of a group of well-characterized patients, mostly female, with a history of RESPI and normal serum immunoglobulin levels revealed a high prevalence of the same MHC haplotypes observed in IgAD, selective IgG subclass deficits, and CVID.[23] These patients tend to respond to aggressive antibiotic therapy, including prophylaxis.

SELECTIVE LIGHT CHAIN DEFICIENCY

Selective κ or λ light chain deficiencies have been reported.[49–51] In one such case, the patient was the offspring of a consanguineous (uncle-niece) union, and in the second, a molecular analysis demonstrated different loss-of-function mutations in the patient's Cκ alleles. The parents of these children had no health difficulties, but each of the patients required medical attention for RESPI and diarrhea. Two of the κ deficient patients exhibited IgAD and the remaining κ deficient and the λ deficient patients were panhypogammaglobulinemic.

TRANSIENT HYPOGAMMAGLOBULINEMIA OF INFANCY

Diagnosis

As infants make the transition from dependence on maternal immunoglobulin to reliance on endogenously produced antibodies, they suffer a physiologic nadir of serum immunoglobulin at 4 to 6 months of age, a period associated with susceptibility to mild upper respiratory infections and otitis media (see Fig. 33.2). Children who (a) exhibit serum concentrations of one or more of the three major immunoglobulin classes that fall below the 95% confidence interval for age on two or more occasions during infancy, (b) demonstrate a rise in these values to or toward normal over time, and (c) lack features consistent with other forms of primary immunodeficiency fall within the catch-all diagnosis of THI.[52,53] By definition, the diagnosis of THI can be made with certainty only in retrospect.

Clinical Manifestations

Immunoglobulin concentrations are rarely measured in infants unless there is some reason to suspect an immunodeficiency. Most patients with this diagnosis come to medical attention due either to recurrent infections or to routine screening studies of family members of immunodeficient patients. Bearing in mind that 2.5% of normal infants will fall below the 95% confidence range at any one time, the diagnosis of THI is remarkably rare. Among two major centers, one in the United States and one in Germany, the diagnosis was given to only 16 of 18,000 children in whom the index of suspicion warranted immunoglobulin determinations.[54,55]

Patients with THI are typically able to synthesize specific antibodies in response to immunization with T-dependent antigens (e.g., tetanus and diphtheria toxoids).[56] They may have difficulty responding to polysaccharide antigens (e.g., isohemagglutinins and vaccination with Pneumovax 23). Some will fail to sustain protective antibody responses to antigens. Most THI patients, especially those with mild upper respiratory infections alone or positive family histories, exhibit fewer infections over time. By 2 years of age, serum immunoglobulin levels frequently normalize in the great majority of THI infants. However, a minority fails to normalize IgG, continues to suffer with recurrent infections, and may develop evidence of autoimmune disease. These patients often become part of the hypogammaglobulinemia syndrome complex that includes CVID. They may end up requiring long-term IGRT, prophylactic antibiotics, or both.

Treatment and Prognosis

Children with suspected THI should be monitored with serial determination of serum immunoglobulins and isohemagglutinin titers in order to confirm acquisition of normal immune function. Development of normal IgG levels may be delayed for several years, and some children will remain IgG subclass or IgA deficient. Treatment of THI with IGRT is warranted should the child persistently suffer with recurrent, invasive infections, including pneumonia.

ON THE HORIZON

- Elucidation of the molecular basis of selective defects in humoral responses to pathogens, in part through the use of high-throughput sequencing to characterize the precise molecular composition of antibody responses in immune deficiencies.
- Further elucidation of the molecular basis of common variable immune deficiency, hypogammaglobulinemia, and IgA deficiency.
- New approaches to treatment such as hematopoietic stem cell transplantation and gene therapy.

FRONTIERS IN RESEARCH

Bruton reported the first case of agammaglobulinemia in 1952, as well as the first successful therapy for this classic primary antibody deficiency. Since that time, there has been remarkable progress in the identification of single-gene disorders. However, for the majority of patients the underlying pathogenesis of the most common manifestation of primary antibody deficiency, hypogammaglobulinemia in the adult, still remains unclear. It seems increasingly likely that this disorder is multifactorial in nature, dependent on the inheritance of one or more susceptibility loci in association with either environmental influences or random chance. In rare cases, patients with hypogammaglobulinemia have shown resolution of their symptoms, suggesting that a better understanding of pathogenesis might yield therapies of remission. The molecular basis of selective deficiencies

in the response to pathogens in the presence of normal serum immunoglobulin levels also remains unclear. The availability of whole exome and whole genome sequencing has helped identify the genetic cause of an increasing number of presumed CVID patients. More gene regulatory and signal transduction pathways will likely be defined in the next decade.

REFERENCES

1. Durandy A, Kracker S, Fischer A. Primary antibody deficiencies. *Nat Rev Immunol*. 2013;13(7):519–533.
2. Johnson ML, Keeton LG, Zhu ZB, Volanakis JE, Cooper MD, Schroeder HW Jr. Age-related changes in serum immunoglobulins in patients with familial IgA deficiency and common variable immunodeficiency (CVID). *Clin Exp Immunol*. 1997;108:477–483.
3. Litzman J, Freiberger T, Grimbacher B, et al. Mannose-binding lectin gene polymorphic variants predispose to the development of bronchopulmonary complications but have no influence on other clinical and laboratory symptoms or signs of common variable immunodeficiency. *Clin Exp Immunol*. 2008;153(3):324–330.
4. De Greef GE, van Tol MJ, Van Den Berg JW, et al. Serum immunoglobulin class and IgG subclass levels and the occurrence of homogeneous immunoglobulins during the course of ageing in humans. *Mech Ageing Dev*. 1992;66(1):29–44.
5. Stiehm ER, Fudenberg HH. Serum levels of immune globulins in health and disease: a survey. *Pediatrics*. 1966;37:715–727.
6. Soothill JF, Hayward AR, Wood CB. *Pediatric Immunology*. Oxford: Blackwell Scientific Publications; 1983. 463–464.
7. Gharib A, Caperton C, Gupta S. Anaphylaxis to IGIV in immunoglobulin-naive common variable immunodeficiency patient in the absence of IgG anti-IgA antibodies: successful administration of low IgA-containing immunoglobulin. *Allergy Asthma Clin Immunol*. 2016;12:23.
8. Wasserman RL. Personalized Therapy: Immunoglobulin replacement for antibody deficiency. *Immunol Allergy Clin North Am*. 2019;39(1):95–111.
9. Perez EE, Orange JS, Bonilla F, Chinen J, Chinn IK, Dorsey M, et al. Update on the use of immunoglobulin in human disease: a review of evidence. *J Allergy Clin Immunol*. 2017;139(3S):S1–S46.
10. Carr TF, Koterba AP, Chandra R, et al. Characterization of specific antibody deficiency in adults with medically refractory chronic rhinosinusitis. *Am J Rhinol Allergy*. 2011;25(4):241–244.
11. Lopez-Herrera G, Vargas-Hernandez A, Gonzalez-Serrano ME, et al. Bruton's tyrosine kinase—an integral protein of B cell development that also has an essential role in the innate immune system. *J Leukoc Biol*. 2014;95(2):243–250.
12. Nelson KS, Lewis DB. Adult-onset presentations of genetic immunodeficiencies: genes can throw slow curves. *Curr Opin Infect Dis*. 2010;23(4):359–364.
13. El-Sayed ZA, Abramova I, Aldave JC, et al. X-linked agammaglobulinemia (XLA): phenotype, diagnosis, and therapeutic challenges around the world. *World Allergy Organ J*. 2019;12(3):100018.
14. Duriez B, Duquesnoy P, Dastot F, Bougneres P, Amselem S, Goossens M. An exon-skipping mutation in the btk gene of a patient with X-linked agammaglobulinemia and isolated growth hormone deficiency. *FEBS Lett*. 1994;346(2-3):165–170.
15. Winkelstein JA, Conley ME, James C, Howard V, Boyle J. Adults with X-linked agammaglobulinemia: impact of disease on daily lives, quality of life, educational and socioeconomic status, knowledge of inheritance, and reproductive attitudes. *Medicine (Baltimore)*. 2008;87(5):253–258.
16. Bearden D, Collett M, Quan PL, Costa-Carvalho BT, Sullivan KE. Enteroviruses in X-linked agammaglobulinemia: update on epidemiology and therapy. *J Allergy Clin Immunol Pract*. 2016;4(6):1059–1065.
17. Conley ME, Dobbs AK, Farmer DM, et al. Primary B cell immunodeficiencies: comparisons and contrasts. *Annu Rev Immunol*. 2009;27:199–227.
18. Tarr PE, Sneller MC, Mechanic LJ, et al. Infections in patients with immunodeficiency with thymoma (Good syndrome). Report of 5 cases and review of the literature. *Medicine (Baltimore)*. 2001;80(2):123–133.
19. Durandy A, Kracker S. Immunoglobulin class-switch recombination deficiencies. *Arthritis Research & Therapy*. 2012;14(4):218.
20. de la Morena MT. Clinical phenotypes of hyper-IgM syndromes. *J Allergy Clin Immunol Pract*. 2016;4(6):1023–1036.
21. Yazdani R, Fekrvand S, Shahkarami S, et al. The hyper IgM syndromes: epidemiology, pathogenesis, clinical manifestations, diagnosis and management. *Clin Immunol*. 2019;198:19–30.
22. Schroeder HW Jr., Schroeder HWIII, Sheikh SM. The complex genetics of common variable immunodeficiency. *J Investig Med*. 2004;52(2):90–103.
23. Johnston DT, Mehaffey G, Thomas J, et al. Increased frequency of HLA -B44 in recurrent sino-pulmonary infections (RESPI). *Clin Immunol*. 2006;119:346–350.
24. Nielsen LK, Dziegiel MH. Recombinant human immunoglobulin (Ig) A1 and IgA2 anti-D used for detection of IgA deficiency and anti-IgA. *Transfusion*. 2008;48(9):1892–1897.
25. Borrelli M, Maglio M, Agnese M, et al. High density of intraepithelial gammadelta lymphocytes and deposits of immunoglobulin (Ig)M anti-tissue transglutaminase antibodies in the jejunum of coeliac patients with IgA deficiency. *Clin Exp Immunol*. 2010;160(2):199–206.
26. Bonilla FA, Barlan I, Chapel H, et al. International Consensus Document (ICON): common variable immunodeficiency disorders. *J Allergy Clin Immunol Pract*. 2016;4(1):38–59.
27. Cunningham-Rundles C. Common variable immune deficiency: dissection of the variable. *Immunol Rev*. 2019;287(1):145–161.
28. Kiaee F, Azizi G, Rafiemanesh H, et al. Malignancy in common variable immunodeficiency: a systematic review and meta-analysis. *Expert Rev Clin Immunol*. 2019;15(10):1105–1113.
29. Wang Y, Hwangpo TA, Martin MP, et al. Killer cell immunoglobulin-like receptors are associated with common variable immune deficiency pathogenesis. *J Allergy Clin Immunol*. 2016
30. Tangye SG, Al-Herz W, Bousfiha A, et al. Human inborn errors of immunity: 2019 update on the classification from the International Union of Immunological Societies Expert Committee. *J Clin Immunol*. 2020;40(1):24–64.
31. Yu JE, Knight AK, Radigan L, et al. Toll-like receptor 7 and 9 defects in common variable immunodeficiency. *J Allergy Clin Immunol*. 2009;124(2); 349-56, 56 e1-3.
32. Bogaert DJ, Dullaers M, Lambrecht BN, Vermaelen KY, De Baere E, Haerynck F. Genes associated with common variable immunodeficiency: one diagnosis to rule them all? *J Med Genet*. 2016;53(9):575–590.
33. Sathkumara HD, De Silva NR, Handunnetti S, De Silva AD. Genetics of common variable immunodeficiency: role of transmembrane activator and calcium modulator and cyclophilin ligand interactor. *Int J Immunogenet*. 2015;42(4):239–253.
34. Zhang L, Radigan L, Salzer U, et al. Transmembrane activator and calcium-modulating cyclophilin ligand interactor mutations in common variable immunodeficiency: clinical and immunologic outcomes in heterozygotes. *J Allergy Clin Immunol*. 2007;120(5):1178–1185.
35. Yong PF, Thaventhiran JE, Grimbacher B. "A rose is a rose is a rose," but CVID is Not CVID common variable immune deficiency (CVID), what do we know in 2011? *Adv Immunol*. 2011;111:47–107.
36. Salzer E, Kansu A, Sic H, et al. Early-onset inflammatory bowel disease and common variable immunodeficiency-like disease caused by IL-21 deficiency. *J Allergy Clin Immunol*. 2014;133(6):1651-9 e12.
37. Hwangpo TA, Wang Z, Ghably J, Bhatt S, Cui X, Schroeder HW, Jr. Retrospective study of spirometry in CVID reveals that FEF25-75% can be used to guide the dose of IgG used to protect pulmonary function. 2019.
38. Malamut G, Verkarre V, Suarez F, et al. The enteropathy associated with common variable immunodeficiency: the delineated frontiers with celiac disease. *Am J Gastroenterol*. 2010;105(10):2262–2275.
39. Sneller MC, Strober W, Eisenstein E, Jaffe JS, Cunningham-Rundles C. NIH conference. New insights into common variable immunodeficiency. *Ann Intern Med*. 1993;118(9):720–730.
40. Chase NM, Verbsky JW, Hintermeyer MK, et al. Use of combination chemotherapy for treatment of granulomatous and lymphocytic interstitial lung disease (GLILD) in patients with common variable immunodeficiency (CVID). *J Clin Immunol*. 2013;33(1):30–39.

41. Lefranc MP, Hammarstrom L, Smith CI, Lefranc G. Gene deletions in the human immunoglobulin heavy chain constant region locus: molecular and immunological analysis. *Immunodefic Rev*. 1991;2(4):265–281.
42. Barton J, Barton C, Bertoli L. Duration of frequent or severe respiratory tract infection in adults before diagnosis of IgG subclass deficiency. *PLoS One*. 2019;14(5):e0216940.
43. Shackelford PG, Granoff DM, Polmar SH, et al. Subnormal serum concentrations of IgG2 in children with frequent infections associated with varied patterns of immunologic dysfunction. *J Pediatr*. 1990;116(4):529–538.
44. Tabata N, Azuma E, Masuda S, Ido M, Sakurai M. Transient low level of IgG3 induced by sepsis. *Acta Paediatr Jpn*. 1995;37(2):201–202.
45. Morell A. IgG subclass deficiency: a personal viewpoint. *Pediatr Infect Dis J*. 1990;9(8 Suppl):S4–S8.
46. Kutukculer N, Karaca NE, Demircioglu O, Aksu G. Increases in serum immunoglobulins to age-related normal levels in children with IgA and/or IgG subclass deficiency. *Pediatr Allergy Immunol*. 2007;18(2):167–173.
47. Paris K, Sorensen RU. Assessment and clinical interpretation of polysaccharide antibody responses. *Ann Allergy Asthma & Immunol*. 2007;99(5):462–464.
48. Perez E, Bonilla FA, Orange JS, Ballow M. Specific antibody deficiency: controversies in diagnosis and management. *Front Immunol*. 2017;8:586.
49. Bernier GM, Gunderman JR, Ruymann FB. Kappa-chain deficiency. *Blood*. 1972;40(6):795–805.
50. Barandun S, Morell A, Skvaril F, Oberdorfer A. Deficiency of kappa- or lambda-type immunoglobulins. *Blood*. 1976;47(1):79–89.
51. Stavnezer-Nordgren J, Kekish O, Zegers BJ. Molecular defects in a human immunoglobulin kappa chain deficiency. *Science*. 1985;230(4724):458–461.
52. Stiehm ER. The four most common pediatric immunodeficiencies. *Journal of Immunotoxicology*. 2008;5(2):227–234.
53. Moschese V, Graziani S, Avanzini MA, et al. A prospective study on children with initial diagnosis of transient hypogammaglobulinemia of infancy: results from the Italian Primary Immunodeficiency Network. *Int J Immunopathol Pharmacol*. 2008;21(2):343–352.
54. Tiller TL Jr., Buckley RH. Transient hypogammaglobulinemia of infancy: review of the literature, clinical and immunologic features of 11 new cases, and long-term follow-up. *J Pediatr*. 1978;92(3):347–353.
55. Dressler F, Peter HH, Muller W, Rieger CH. Transient hypogammaglobulinemia of infancy: five new cases, review of the literature and redefinition. *Acta Paediatr Scand*. 1989;78(5):767–774.
56. Tiller TL Jr., Buckley RH. Transient hypogammaglobulinemia of infancy: review of the literature, clinical and immunologic features of 11 new cases, and long-term follow-up. *J Pediatr*. 1978;92:347–353.

34

Primary T-Cell Immunodeficiencies

Luigi D. Notarangelo

Adaptive immune responses consist of specific recognition of antigens, effector mechanisms if immunity, and development of immunological memory. T lymphocytes play an essential role in this process. By expressing either β or γ heterodimeric T-cell receptors (TCRs) on the cell surface, they are able to recognize antigenic epitopes. Moreover, while CD8 T cells are endowed with cytotoxic effector mechanisms that may contribute to killing of virus-infected cells, CD4 T lymphocytes participate in immune-mediated responses by providing soluble and membrane-bound signals that permit activation and differentiation of B lymphocytes, dendritic cells, and macrophages. CD4− CD8− T cells expressing TCR β or γ on their cell surface contribute to immune responses against mycobacteria and other pathogens. Finally, in the course of an immune response, a subset of antigen-specific T cells differentiates into memory cells that patrol the body, thereby triggering prompt and robust responses during subsequent encounters with the same antigen.

In this chapter, we review primary T-cell immunodeficiencies, a large group of genetically determined disorders characterized by abnormalities of T-cell development and/or function.

KEY CONCEPTS

T-Cell Immunodeficiencies

- A large group of disorders caused by genetic defects that affect development, maturation, and/or function of T lymphocytes
- Depending on the severity of the numerical and functional defect, T-cell immunodeficiencies can be distinguished in two categories: severe combined immune deficiency (SCID) and combined immune deficiency (CID)
- In most cases, the defect is intrinsic to the hematopoietic lineage; however, some forms of T-cell immunodeficiency may be due to abnormalities of thymus development and function
- The age of clinical onset may vary from infancy to adulthood, depending on the severity of the defect

SEVERE COMBINED IMMUNE DEFICIENCY—GENERAL CONSIDERATIONS

Severe combined immune deficiency (SCID) comprises a heterogeneous group of genetic disorders characterized by profound impairment of T-cell development and/or function (Fig. 34.1).[1,2] In some forms of SCID, the number and/or function of B and/or natural killer (NK) cells are also affected by the underlying gene defect, so that SCID may manifest with four distinct immunological phenotypes: (1) $T^-B^+NK^-$ SCID (the most common phenotype); (2) $T^-B^+NK^+$ SCID; (3) $T^-B^-NK^+$ SCID; or (4) $T^-B^-NK^-$ SCID (Table 34.1). However, because antibody responses to protein antigens require T/B-cell interaction, infants with SCID have impaired humoral immunity even if the gene defect does not directly affect the B-cell compartment.

The natural history of SCID is characterized by early-onset, life-threatening infections that may be sustained by a variety of pathogens (bacterial, viral, fungal), and even by opportunistic microorganisms. Interstitial pneumonia due to *Pneumocystis jiroveci*, cytomegalovirus (CMV), or other viruses (adenovirus, parainfluenzae virus, respiratory syncytial virus) in the first months of life, chronic diarrhea, and failure to thrive are common clinical features of infants with SCID. Because of the inability to control replication of live microorganisms, administration of live-attenuated vaccines often leads to severe, life-threatening complications in infants with SCID. Breastfeeding from a CMV positive mother may cause transmission of the virus and severe disease; therefore, screening of the mother's CMV seropositivity status is important, and if positive should prompt avoiding breastfeeding if possible. Transfusion of SCID babies with unirradiated blood products often leads to fatal graft-versus-host disease.

Engraftment of maternal T cells that cross the placenta occurs in more than 50% of infants with SCID.[2,3] Most often asymptomatic, it may cause skin rash or, less frequently, typical graft-versus-host disease with generalized rash, liver disease, profuse diarrhea, jaundice, and severe hematological abnormalities (thrombocytopenia, anemia, leukopenia) that are indicative of marrow damage.

SCID is inevitably fatal unless treated by allogeneic hematopoietic stem cell transplantation (HSCT) or, in selected cases, by gene therapy or enzyme replacement therapy (ERT).[2]

From a laboratory standpoint, severe lymphopenia and, in particular profound T-cell lymphopenia, are typical findings. Furthermore, in vitro proliferative responses to mitogens and antigens are severely reduced or absent. Immunoglobulin serum levels are often low, but serum IgG may be normal early in life, reflecting transplacental passage of maternal IgG. However, antibody response to vaccine antigens is severely impaired. Flow-cytometric analysis of the expression of surface markers specific for T, B, and NK cells permits one to define the immunological phenotype of SCID and may provide information regarding the underlying gene defect. Although most patients with SCID manifest profound T-cell lymphopenia, the number of circulating T cells may be less affected and may even be normal in patients with atypical forms of SCID (associated with residual capacity to produce T lymphocytes) or with maternal T-cell engraftment. In both cases, however, the proportion of circulating naïve ($CD45RA^+$ $CCR7^+$ or $CD45RA^+$ $CD62L^+$) T cells is severely reduced.

A major advance in the diagnostic approach to SCID has been the introduction of newborn screening, based on the quantification of T-cell receptor excision circles (TRECs) in dried blood spots collected at birth. DNA rearrangements that occur at the TCR loci during T-cell development result in the generation of

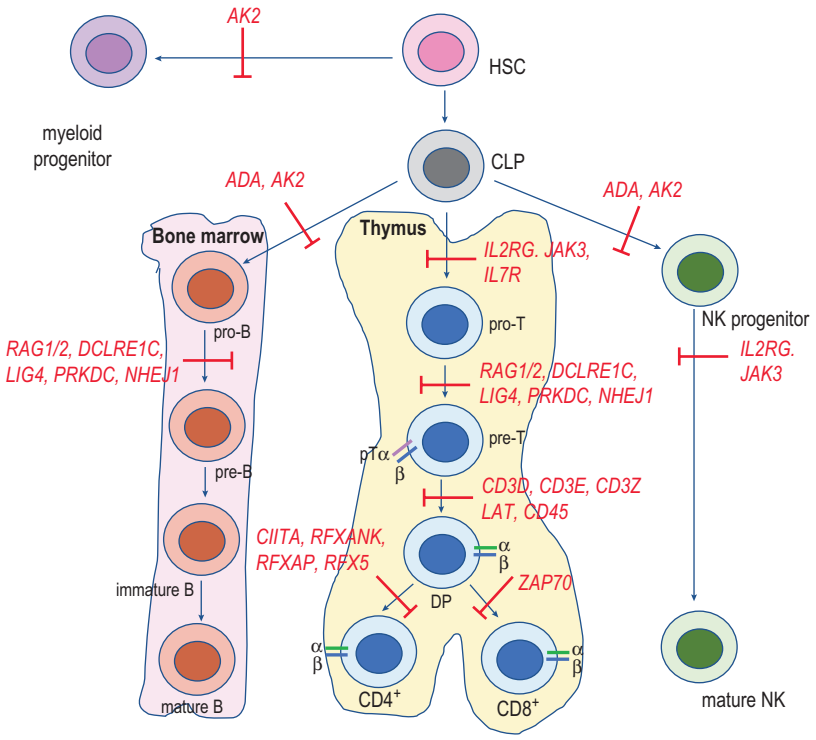


FIG. 34.1 Schematic Representation of Developmental Blocks That Characterize Various Forms of Severe T-Cell Deficiency. Common lymphoid progenitor *(CLP)* cells originate from hematopoietic stem cells *(HSCs)* and give rise to cells of T, B, and natural killer *(NK)* lineages. Differentiation of T cells in the thymus proceeds through discrete stages. Signaling through the interleukin-7 receptor *(IL-7R)* mediates expansion and survival of early T-cell progenitors. V(D)J recombination, initiated by the recombinase activating genes (*RAG1*, *RAG2*) and involving components of the nonhomologous end-joining pathway, allows DNA rearrangement at T-cell receptor (TCR) loci. Pre-T cells express on the surface a complex composed of TCRβ and pre-Tα (pTα) molecules along with CD3 subunits. Signaling through this complex promotes further rearrangements at the TCR α/δ locus, allowing generation of $CD4^+$ $CD8^+$ double-positive (DP) cells expressing the TCRαβ heterodimer. After positive and negative selection, $CD4^+$ and $CD8^+$ single-positive mature thymocytes are generated, that are exported to the periphery. Various defects along this differentiation pathway may cause severe T-cell deficiency. *Blocked red lines* identify specific stages at which known gene defects affect development of T, B, and NK cells. *ADA*, Adenosine deaminase; *AK2*, adenylate kinase 2; *IL2RG*, interleukin-2 receptor gamma chain; *JAK3*, Janus-associated kinase 3; *LAT*, linker of activated T cells; *LIG4*, DNA ligase 4; *ZAP-70*, Zeta-associated protein of 70 kDa.

TABLE 34.1 Immunological and Genetic Characteristics of Severe Combined Immune Deficiency

				CIRCULATING LYMPHOCYTES		
Phenotype	**Disease**	**Gene**	**Inheritance**	**T**	**B**	**NK**
T^- B^- NK^-						
	ADA deficiency	*ADA*	AR	↓↓	↓↓	↓↓
	Reticular dysgenesis	*AK2*	AR	↓↓	↓↓	↓
T^- B^- NK^+						
	RAG deficiency	*RAG1, RAG2*	AR	↓↓	↓↓	N
	ARTEMIS deficiency	*DCLRE1C*	AR	↓↓	↓↓	N
	DNA-PKcs deficiency	*PRKDC*	AR	↓↓	↓↓	N
	LIG4 deficiency	*LIG4*	AR	↓↓	↓↓	N
	CERNUNNOS deficiency	*NHEJ1*	AR	↓↓	↓↓	N
T^- B^+ NK^-						
	X-linked SCID	*IL2RG*	XR	↓↓	N	↓↓
	JAK3 deficiency	*JAK3*	AR	↓↓	N	↓↓
T^- B^+ NK^+						
	IL-7Rα deficiency	*IL7R*	AR	↓↓	N	N
	CD3 defects	*CD3D, CD3E, CD3Z*	AR	↓↓	N	N
	LAT deficiency	*LAT*	AR	↓↓	N	N
	CD45 deficiency	*CD45*	AR	↓↓	N	N

N, Normal.

excised DNA fragments that are ligated into circles (TRECs). In particular, DNA rearrangements at the TCRTCRα/δocus occur in approximately 70% of thymocytes, and generate a δrec-ψJα TREC.[4,5] Because TRECs are diluted with T-cell divisions, enumeration of TREC levels by quantitative polymerase chain reaction (qPCR) provides valid information on the capacity of the thymus to generate T cells.[6] Moreover, implementation of newborn screening for SCID has also permitted more precise definition of its incidence, now estimated to be 1:50,000 to 1:75,000 live births in the United States.[7] However, this figure is significantly higher in countries with a high rate of consanguinity and among certain genetically restricted ethnic groups due to a higher rate of autosomal recessive forms of SCID in these populations.

Ultimately, molecular analysis permits identification of the genetic basis of SCID. Appropriate gene panels or whole exome/whole genome sequencing (WES/WGS) can be used to correctly identify the gene defect. Defining the cellular and molecular bases of SCID is also important from a therapeutic standpoint. In particular, some forms of SCID are associated with defective DNA repair. These patients are at high risk of severe, potentially fatal complications if chemotherapy is used as part of the conditioning regimen during HSCT. Moreover, while the vast majority of SCID disorders are due to hematopoietic cell autonomous defects, severe T-cell lymphopenia may also reflect thymus-intrinsic defects. For the latter category of patients, thymus transplantation may be needed. Finally, the nature of the pathogenic variant often determines the severity of the immunological phenotype. Complete loss-of-expression and loss-of-function (null) mutations are associated with a more severe (typical) SCID phenotype. In contrast, hypomorphic mutations allowing residual development (and/or function) of T cells are often responsible for atypical ("leaky") forms of the disease. Diagnostic criteria distinguishing typical and atypical forms of SCID have been developed by the Primary Immune Deficiency Treatment Consortium (PIDTC) and are reported in Table 34.2. SCID genotypes have a different distribution among typical and atypical forms of SCID (Fig. 34.2).[8] Atypical forms of SCID are often associated with manifestations of immune dysregulation, with oligoclonal expansion of T cells that may infiltrate and damage peripheral organs. Infants with these features plus erythroderma, adenopathy, eosinophilia, and elevated immunoglobulin E (IgE)

KEY CONCEPTS

Severe Combined Immune Deficiency

- Severe combined immune deficiency (SCID) comprises a heterogeneous group of genetic disorders characterized by profound numerical and functional defects of T lymphocytes. The severity of T-cell lymphopenia is the main element that distinguishes typical and atypical forms of SCID. The latter is often due to hypomorphic variants in SCID-associated genes
- In some forms of SCID, also the number of B and/or NK lymphocytes is reduced. However, antibody responses are uniformly compromised, because of the lack of T-helper function
- Maternal T-cell engraftment is common in babies with SCID and may variably affect the clinical phenotype (from being clinically silent to causing severe graft-versus-host disease). In SCID infants with maternal T-cell engraftment, the T-cell count may vary, but the proportion of naïve T cells is very low
- The natural history of SCID is characterized by early-onset, life-threatening infections sustained by virus, bacteria, and fungi. Interstitial pneumonia, chronic diarrhea, and failure to thrive are common clinical features
- Omenn syndrome, characterized by generalized erythroderma, lymphadenopathy, and eosinophilia, is a peculiar phenotype associated with oligoclonal expansion of autologous T cells infiltrating the skin and other organs
- Newborn screening, with enumeration of T-cell receptor excision circles (TRECs), allows identification of babies with profound naïve T-cell lymphopenia. Additional phenotypic, functional, and genetic studies permit definitive diagnosis of SCID

TABLE 34.2 Definition of Typical and Atypical SCID According to Criteria of the Primary Immune Deficiency Treatment Consortium (PIDTC)

Typical SCID

- Absence of very low number of T cells (CD3 T cells <300/μL), **AND**
- No or very low T-cell function (<10% of lower limit of normal) as measured by response to phytohemagglutinin (PHA)
- **OR**
- T cells of maternal origin present

Atypical (Leaky) SCID

- Presence of maternal lymphocytes tested and not detected
- AND either one or both of the following:
  a. <50% of lower limit of normal T-cell function as measured by response to PHA or to anti-CD3/CD28 antibody
  b. Absent or less than 30% of lower limit of normal proliferative response to Candida and tetanus toxoid antigens
- **AND** at least two of the following:
  a. Reduced number of CD3 T cells [age ≤ 2 years: <1500 cells/μL; age >2 years ≤ 4 years: 800 cells/μL; age >4 years: <600 cells/μL]
  b. 80% of CD3 or CD4 T cells are $CD45R0^+$
  AND/OR >80% of CD3 or CD4 T cells are CD62L-negative
  AND/OR >50% of CD3 or CD4 T cells express Human Leukocyte Antigen DR (HLA-DR)
  (at <4 years of age)
  AND/OR are oligoclonal T cells
  c. Hypomorphic mutation in *IL2RG* in a male or homozygous hypomorphic mutation or compound heterozygosity with at least one hypomorphic mutation in an autosomal SCID-causing gene
  d. Low TRECs and/or the percentage of $CD4^+$ $CD45RA^+$ $CD31^+$ or $CD4^+$ $CD45RA^+$ $CD62L^+$ cells is below the lower limit of normal [for reference: see Schatorje et al.[89]]
  e. Functional testing in vitro that supports impaired, but not absent, activity of the mutant protein
- **AND** does not meet criteria for Omenn syndrome, which include infantile erythroderma rash, adenopathy, hepatomegaly, eosinophilia, elevated serum IgE, and oligoclonal T-cell expansion

SCID, Severe combined immune deficiency.

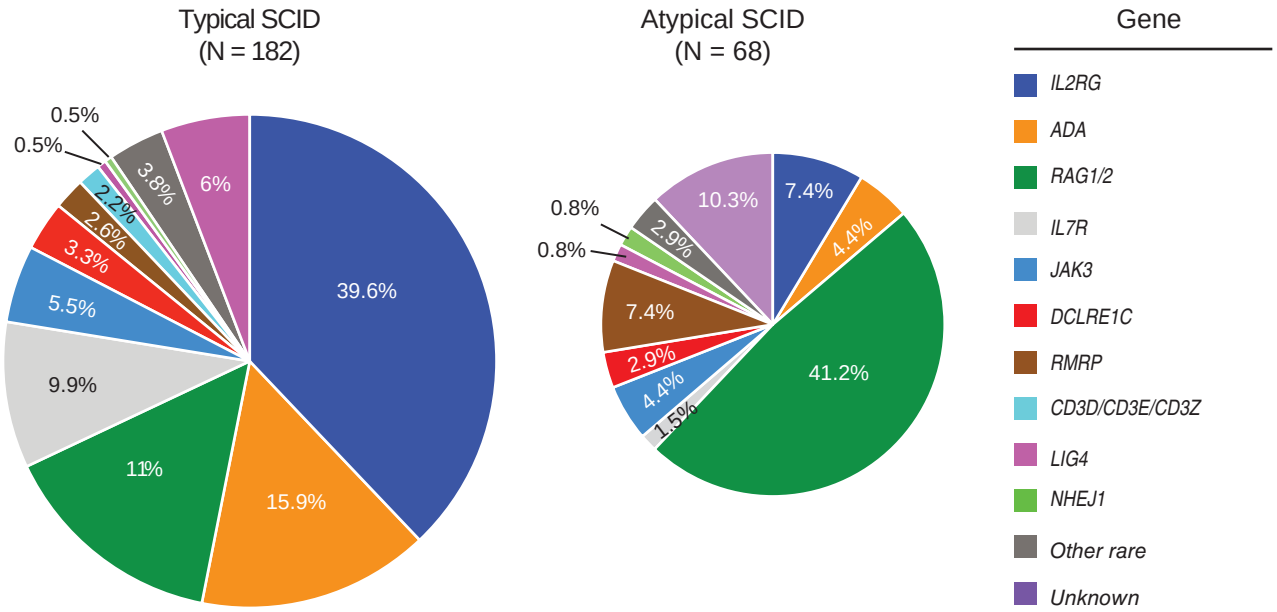


FIG. 34.2 Distribution of Scid Genotypes Among Patients with Typical and Atypical forms of SCID. Data are from reference 8 and pertain to patients with SCID identified in the period 2000–2008 and enrolled in protocols of the Primary Immune Deficiency Treatment Consortium. *SCID*, Severe combined immune deficiency.

are designated as Omenn syndrome. In patients with atypical SCID, conditioning regimes with chemotherapy and/or serotherapy may be necessary to eliminate the dysreactive T cells and permit robust and durable donor stem cell engraftment and immune reconstitution.

Finally, SCID can be divided into different groups based on the mechanisms underlying the T-cell lymphopenia:

SCID due to hematopoietic defects

- Metabolic defects affecting survival of T-cell progenitors
- Defects of cytokine-mediated signaling
- Defects of V(D)J recombination
- Defects of the CD3/TCR complex
- Other mechanisms

SCID due to non-hematopoietic defects

SCID DUE TO METABOLIC DEFECTS AFFECTING SURVIVAL OF T-CELL PROGENITORS

Adenosine Deaminase Deficiency

Adenosine deaminase (ADA), an enzyme of the purine salvage pathway, converts adenosine (Ado) and deoxyadenosine (dAdo) to inosine and deoxyinosine, respectively.[9] In the absence of ADA, intracellular concentrations of Ado, dAdo, and their phosphorylated derivatives (AXP, dAXP) increase. Although ADA is ubiquitously expressed, high levels of dAdo and dAXP are particularly toxic to developing lymphocytes; moreover, dATP inhibits ribonucleotide reductase, an enzyme required for DNA synthesis. Consequently, complete ADA deficiency (OMIM *102700) is characterized by a profound reduction of circulating T, B, and NK cells, and therefore is a cause of T^- B^- NK^- SCID.[9] Furthermore, ADA deficiency also interferes with the proliferative capacity of circulating T lymphocytes and with the function of regulatory T (Treg) cells and B-cell tolerance. ADA deficiency is inherited as an autosomal recessive trait and accounts for approximately 5% to 10% of all cases of SCID. Affected patients have typical manifestations of SCID, with early-onset, life-threatening infections. Besides lymphoid cells, ADA deficiency also affects other organs.[10] Cognitive and behavioral abnormalities and sensorineural deafness are common. Liver and renal dysfunction, costochondral abnormalities (scapular squaring, anterior rib cupping, flared costochondral junctions), and pulmonary alveolar proteinosis (due to dysfunction of alveolar macrophages) have also been frequently reported. Patients with ADA deficiency are at increased risk of Epstein-Barr Virus (EBV)-associated lymphoma and of multicentric dermatofibrosarcoma protuberans.[9]

Hypomorphic *ADA* mutations are often associated with a delayed and less severe clinical, immunological, and metabolic phenotype.[9] Somatic mosaicism with gene reversion or second-site mutations that restore (fully or in part) ADA activity and lymphocyte survival have been identified as another reason for milder clinical presentations. Immune dysregulation, with eczema, type 1 diabetes, other autoimmune manifestations, and hepatosplenomegaly, may occur in patients with residual ADA function.

Null *ADA* gene mutations are associated with low to undetectable levels of TRECs upon newborn screening. Hypomorphic variants allowing residual development of T cells may escape identification by newborn screening with TREC quantification but can potentially be correctly identified by tandem mass spectrometry, although this is not widely performed at present.[11] Measurement of ADA enzymatic activity and of dAdo and dAXP levels in erythrocytes represents the gold standard for diagnosis. However, if the patient has received red cell transfusions, measurement of ADA activity and of levels of toxic metabolites should be performed on other cell types, such as peripheral blood mononuclear cells. Mutation analysis provides definitive confirmation.

HSCT and gene therapy represent efficacious definitive therapies for ADA deficiency[12] and are discussed in Chapters 90 and 91. ERT with intramuscular injections of bovine recombinant

ADA conjugated with polyethylene glycol (PEG-ADA) allows rapid detoxification and is often used as a bridge therapy between diagnosis and definitive therapy. Improvement of T- and B-cell count and function has been observed in approximately 80% of the patients treated with ERT, although persistent lymphopenia and progressive decline of immune function over time have been reported in most of them.

Reticular Dysgenesis

Reticular dysgenesis (OMIM *267500) is an autosomal recessive form of SCID, characterized by extreme lymphopenia, absence of neutrophils, and sensorineural deafness.[13] The disease is caused by mutations of the adenylate kinase 2 (*AK2*) gene that result in apoptosis of myeloid precursors of neutrophils and of lymphoid progenitor cells. Hypomorphic mutations that allow residual lymphopoiesis and granulopoiesis are associated with atypical presentations, including generalized erythroderma suggestive of Omenn syndrome, autoimmunity, or isolated hypogammaglobulinemia. HSCT represents the only available treatment.[13,14]

SCID DUE TO DEFECTS OF CYTOKINE-MEDIATED SIGNALING

X-Linked Severe Combined Immune Deficiency

X-linked SCID (OMIM *300400) is the most common form of typical SCID. It affects only males as it is caused by hemizygous mutations of the X-linked *IL2RG* gene, encoding the membrane-spanning common gamma chain (γ_c), shared by the cytokine receptors for interleukin (IL)-2, IL-4, IL-7, IL-9, IL-15, and IL-21 (Fig. 34.3).[15] Proliferation of thymic T-cell progenitors depends on IL-7, and NK-cell development requires IL-15.[16] Consequently, patients with X-linked SCID manifest a T^- B^+ NK^- phenotype. Moreover, while human B-cell development is spared in patients with X-linked SCID, B-cell function is defective because IL-21 is required for plasmablast differentiation.[17] Surface γ_c expression is abolished in patients with null *IL2RG* mutations but may be preserved in those with missense variants or truncations of the terminal exons encoding the intracellular domains of the molecule.

Most patients with X-linked SCID present with typical manifestations of SCID starting in the first months of life.[2] Their thymi are severely atrophic and not visible by imaging studies. However, milder clinical features and delayed clinical onset, associated with preserved presence of T and/or NK cells (consistent with atypical SCID) and partially preserved thymic architecture, have been reported in unusual patients with hypomorphic mutations or with somatic gene revertants. In the latter, revertant T cells may persist for up to several years.[2] Carrier females are asymptomatic and display nonrandom X-chromosome inactivation in T and NK cells.

Treatment of X-linked SCID requires allogeneic HSCT (see Chapter 90) or transplantation of autologous gene-corrected cells (see Chapter 91). Very promising results have been recorded also after gene therapy in both infants and older patients who had failed previous attempts with HSCT for X-linked SCID.[18]

Interleukin-7 Receptor Deficiency

Autosomal recessive interleukin-7 receptor (IL-7R) deficiency[2] (OMIM *146661) is characterized by a T^- B^+ NK^+ phenotype, reflecting the critical role played by IL-7 in thymocyte survival and proliferation. Clinical manifestations are typical of SCID. Treatment is allogeneic HSCT.

JAK3 Deficiency

All of the γ_c-containing cytokine receptors depend on Janus-associated kinases (JAKs) for signaling (see Fig. 34.3). In particular, the γ_c is constitutively associated with JAK3. Accordingly, autosomal recessive JAK3 deficiency (OMIM *600802) shares similar clinical and laboratory features with X-linked SCID, but due to the gene location on chromosome 19 instead of the X chromosome, homozygous or compound heterozygous mutations affect females as well as males. Patients present early in life with typical features of SCID and manifest a T^- B^+ NK^- phenotype.[19] Atypical forms, including delayed-onset disease and presentation with severe warts and increased risk of lymphoma, have been reported in patients with hypomorphic mutations.[2]

SCID DUE TO DEFECTS OF V(D)J RECOMBINATION

Expression of antigen-specific receptors on the surface of T and B lymphocytes depends on rearrangement and ligation of the variable (V), diversity (D), and joining (J) gene segments of the TCR and immunoglobulin loci by V(D)J recombination (Chapters 4, 7 and 9). Any of several gene defects may impair this process, causing T^- B^- SCID.

Defects of Recombination Activating Genes (RAG) 1 and RAG2

RAG1 and RAG2 are lymphoid-specific proteins that form a heterotetramer (with two subunits of each protein). They recognize recombination signal sequences that flank the *V*, *D*, and *J* gene elements and introduce DNA breaks at the TCR and immunoglobulin loci, thereby initiating V(D)J recombination.[20] RAG1 and RAG2 deficiencies are inherited as autosomal recessive traits and account for approximately 20% of all cases of SCID.[8] Patients with null mutations in these genes manifest T^- B^- NK^+ SCID and suffer from typical early-onset clinical features. However, hypomorphic mutations allowing residual generation of T (and in some cases, B) cells are associated with a spectrum of clinical phenotypes that includes Omenn syndrome, atypical SCID, and combined immunodeficiency with granulomas and/or autoimmunity (CID-G/AI).[21] In particular, Omenn syndrome is characterized by generalized erythroderma, hepatosplenomegaly, and lymphadenopathy. Immunological features of the disease include a variable number of autologous, oligoclonal T cells with an activated phenotype that infiltrate the skin and other target organs. Eosinophilia, hypogammaglobulinemia, and elevated serum IgE (despite profound hypogammaglobulinemia) are prominent features of immune dysregulation. While hypomorphic *RAG* mutations are the most common cause of Omenn syndrome, genetic defects in other SCID-associated genes that drastically reduce (but do not completely abrogate) the capacity to generate T cells may also cause the syndrome. When Omenn syndrome is caused by *RAG* defects, circulating B cells are absent or markedly reduced in number. Autoimmune manifestations (in particular, autoimmune hemolytic anemia) may occur in patients with RAG deficiency manifesting as atypical SCID. A subset of these patients develops an expansion of $TCR\gamma\delta^+$ T cells after CMV infection, and increased risk of EBV-induced lymphoproliferation has been reported. In contrast to patients

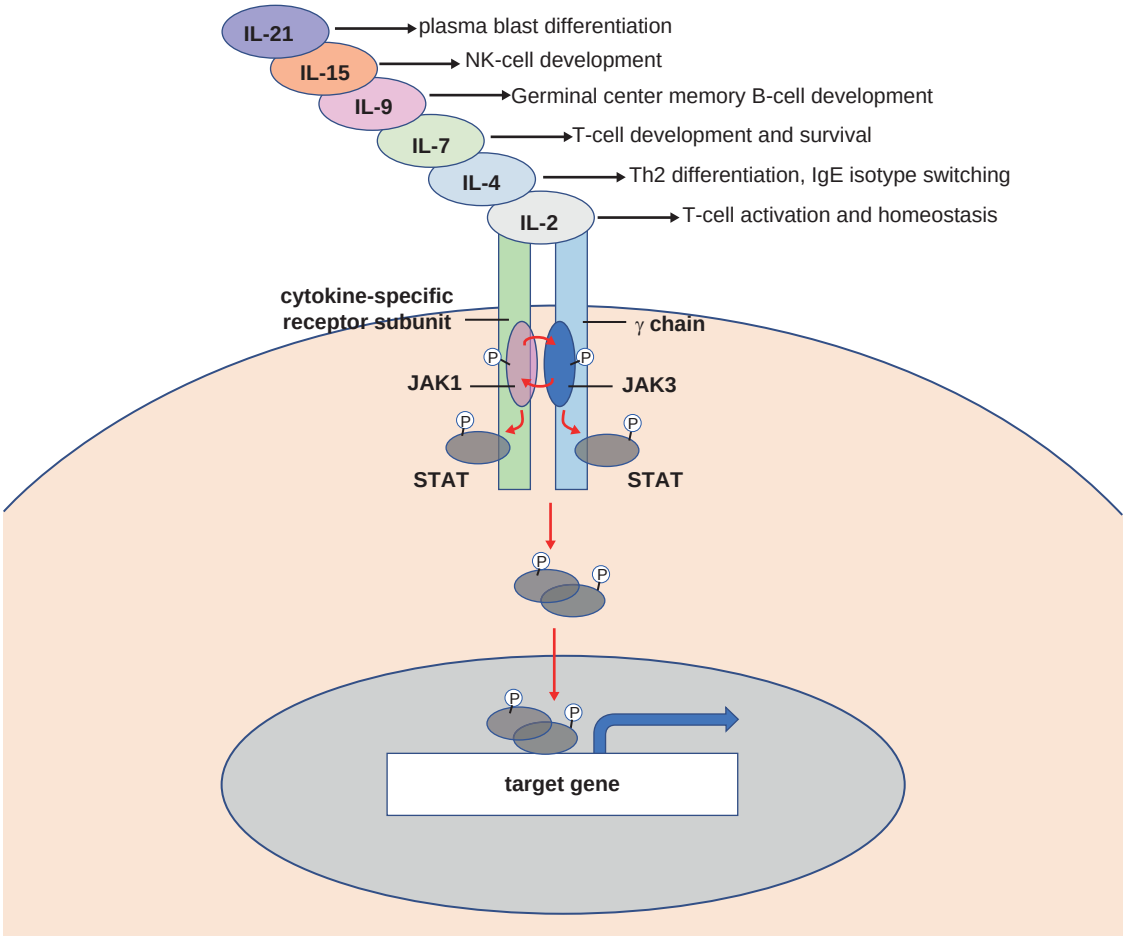


FIG. 34.3 Functional Role of Cytokines Signaling Through the γc//JAK/STAT Pathway. Cytokine receptor engagement triggers cross-phosphorylation and activation of the JAK proteins, allowing phosphorylation of the cytokine receptor chain and docking of the STAT proteins. The JAK proteins also mediate phosphorylation of the STAT proteins, which dimerize and translocate to the nucleus, inducing transcription of cytokine-responsive genes. *JAK*, Janus-associated kinase; *STAT*, signal transducer and activator of transcription.

with T^- B^- NK^+ SCID, Omenn syndrome, or atypical SCID, those with CID-G/AI may present later in life (even in adulthood). Autoimmune cytopenias, organ-specific autoimmunity, and granulomas involving the skin or other organs are typical manifestations.[22] The autoimmunity is secondary to defects of negative selection of self-reactive T cells and impaired Treg generation in the thymus, as well as to defective receptor editing, a process mediated by RAG gene re-expression in the bone marrow and that allows to purge the developing B-cell repertoire from self-reactive specificities. Presence of the rubella virus vaccine strain has been documented in the granulomas of several patients with CID-G/AI; this phenomenon is not unique to patients with hypomorphic RAG defects, but can also be observed in patients with various forms of combined immune deficiency.[23] Because patients with CID-G/AI have a variable number of T and B cells and variable hypogammaglobulinemia, the diagnosis is often delayed. Furthermore, while the TREC assay at birth identifies patients with RAG mutations manifesting as SCID or Omenn syndrome, it is unclear to what extent it also detects patients with CID-G/AI, whose RAG mutations often support levels of recombination activity that allows generation of a relatively diversified repertoire of T cells. Treatment of RAG deficiency is based on HSCT, but chemotherapy and serotherapy may be needed to attain durable stem cell engraftment and immune reconstitution.[24] The autoimmune and inflammatory manifestations of CID-G/AI are often refractory to medical management but tend to resolve after HSCT.

ARTEMIS Deficiency

ARTEMIS, encoded by the *DCLRE1C* gene, is required to open the hairpins that seal the DNA broken ends generated by RAG proteins during V(D)J recombination. This triggers the interventions of several other proteins involved in DNA repair. ARTEMIS deficiency (OMIM *605988) is an autosomal recessive form of T^- B^- NK^+ SCID.[2,25] Besides typical manifestations of SCID, oral and genital ulcers are common. In addition, patients with ARTEMIS deficiency manifest a generalized cellular radiosensitivity, and when exposed to ionizing radiation or alkylating agents they are at high risk of serious complications, including growth failure, malabsorption requiring parenteral nutrition, endocrine abnormalities, renal disease, and dental problems. Patients with hypomorphic *DCLRE1C* are at increased risk of EBV-driven lymphoma and may develop granulomas. HSCT represents the mainstay of treatment, but caution must be used in the choice and tissue exposure of agents for conditioning. A clinical trial of gene therapy is currently under way.

CLINICAL RELEVANCE

T-cell Immunodeficiencies With Radiation Sensitivity

Clinical features and management strategies:

- Gene defects that affect DNA repair are responsible for severe combined immune deficiency (SCID) and other combined immunodeficiencies associated with increased cellular radiosensitivity
- Besides an increased risk of infections, these disorders are often characterized by other clinical features, including microcephaly, growth retardation, neurodevelopmental delay and other neurological signs, accelerated aging, bone marrow failure, and increased rate of malignancies
- Genetic tests and in vitro radiosensitivity assays are the mainstays of diagnosis
- Exposure to ionizing radiation and to alkylating agents carries serious risks in patients with radiosensitive T-cell disorders and should be avoided unless strictly necessary
- Radiosensitive SCID can be cured with hematopoietic stem cell transplantation. However, patients remain at risk of some complications (short stature, tooth problems, malabsorption) as the result of cell damage in non-hematopoietic tissues and of the use of chemotherapeutic agents

DNA-PKcs Deficiency

The DNA-dependent protein kinase catalytic subunit (DNA-PKcs), encoded by the *PRKDC* gene, is involved in the non-homologous end-joining (NHEJ) pathway of DNA repair. Deficiency of DNA-PKcs (OMIM #615966) is inherited as an autosomal recessive trait and is characterized by T^- B^- NK^+ SCID associated with cellular radiosensitivity.[25] Neurodevelopmental delay has been reported in some patients. Hypomorphic mutations may cause atypical SCID and autoimmunity. HSCT can cure the immunodeficiency, but use of alkylating agents should be avoided.

DNA Ligase IV Deficiency

DNA ligase IV is another component of the NHEJ DNA repair pathway. Deficiency of this enzyme (OMIM #606593) causes autosomal recessive T^- B^- NK^+ SCID, associated with microcephaly, developmental delay, and an increased risk of bone marrow failure and hematological malignancies.[25] Hypomorphic variants are associated with milder cellular defects, contributing to significant variability of the clinical and immunological phenotype.

Cernunnos/XLF Deficiency

Cernunnos deficiency (OMIM #611291), caused by mutations of the *NHEJ1* gene, is another form of autosomal recessive radiosensitive SCID and is characterized by progressive T- and B-cell lymphopenia, microcephaly, and growth retardation.[25]

SCID DUE TO DEFECTS OF THE CD3/TCR COMPLEX

The T-cell receptor heterodimeric (TCRαβ or TCRγδ) molecules are expressed on the surface of T cells in association with CD3γ, CD3δ, CD3ε, and CD3ζ invariant chains. Antigen recognition by the TCR induces activation of the p56Lck kinase, which phosphorylates the CD3 molecules, allowing recruitment and activation of the zeta-associated protein of 70 kDa (ZAP-70). This in turn triggers activation of downstream molecules (LAT, RHOH, STK4, ITK) and initiation of Ca^{2+} flux (Fig. 34.4).

CD3 deficiency includes various genetic defects affecting CD3δ (OMIM *186790), CD3ε (OMIM *186830), and CD3ζ (OMIM *186780), causing a block in thymocyte development leading to T^- B^+ NK^+ SCID.[26] Hypomorphic defects of these genes may cause atypical presentations. In contrast, CD3γ deficiency (OMIM *186740) results in milder T-cell deficiency and variability of the clinical phenotype that often includes autoimmunity.[26]

Deficiency of the linker of activated T cells (LAT) (OMIM *602354) causes T^- B^+ NK^+ SCID;[27] atypical variants with splenomegaly, lymphadenopathy, and autoimmunity have also been reported.

CD45 deficiency (OMIM *151460) is an autosomal recessive form of SCID due to mutations of the CD45 phosphatase involved in signaling. Patients manifest T^- B^+ NK^+ SCID and profound hypogammaglobulinemia.

SCID DUE TO OTHER HEMATOPOIETIC DEFECTS

Coronin-1A Deficiency

Coronin-1A is involved in intracellular signaling, actin cytoskeleton regulation, and cell motility. Deficiency of this protein (OMIM *615401) causes autosomal recessive T^- B^+ NK^+ SCID; hypomorphic mutations have been associated with CD4 T-cell lymphopenia, warts, granulomatous lesions, and an increased risk of EBV-induced lymphoproliferative disease and severe varicella.[28]

RAC2 Gain-of-Function Mutations

Gain-of-function (GOF) mutations of the small GTPase RAC2 cause an autosomal dominant form of SCID (OMIM *602049) with low numbers of T and B cells, as well as neutropenia associated with increased content of F-actin in leukocytes and defective cell migration.[29]

SCID With Multiple Intestinal Atresia

This autosomal recessive condition (OMIM *609332) is due to mutations of the tetratricopeptide repeat domain 7 A (*TTC7A*) gene.[30] The degree of T-cell lymphopenia is variable but is often very severe. The number of circulating B cells may be normal or low, but immunoglobulin serum levels are markedly reduced. Patients are at increased risk of life-threatening bacterial, viral, and fungal infections. Intestinal atresias may affect multiple areas of the gastrointestinal tract. There is a high mortality rate early in life. HSCT may correct the immune defect but not the gastrointestinal manifestations.

Veno-Occlusive Disease With Immunodeficiency

Veno-occlusive disease with immunodeficiency (VODI) (OMIM #235550) is an autosomal recessive disease characterized by liver abnormalities (veno-occlusive disease, fibrosis, often progressing to liver failure) and immunodeficiency, with onset in the first months of life.[31] Patients are prone to recurrent infections with viruses, bacteria, and opportunistic pathogens (*P. jirovecii, Candida*, CMV). Thrombocytopenia is frequent. Immunological defects include low number of memory T and B lymphocytes, defective B-cell differentiation in vitro into antibody-secreting cells, and hypogammaglobulinemia.[31] The disease is caused by mutations of the *SP110* gene, which encodes a nuclear body protein that acts as a transcription factor. The prognosis is dismal. HSCT is the only curative approach, but there is a high risk of severe liver toxicity due to conditioning.[32]

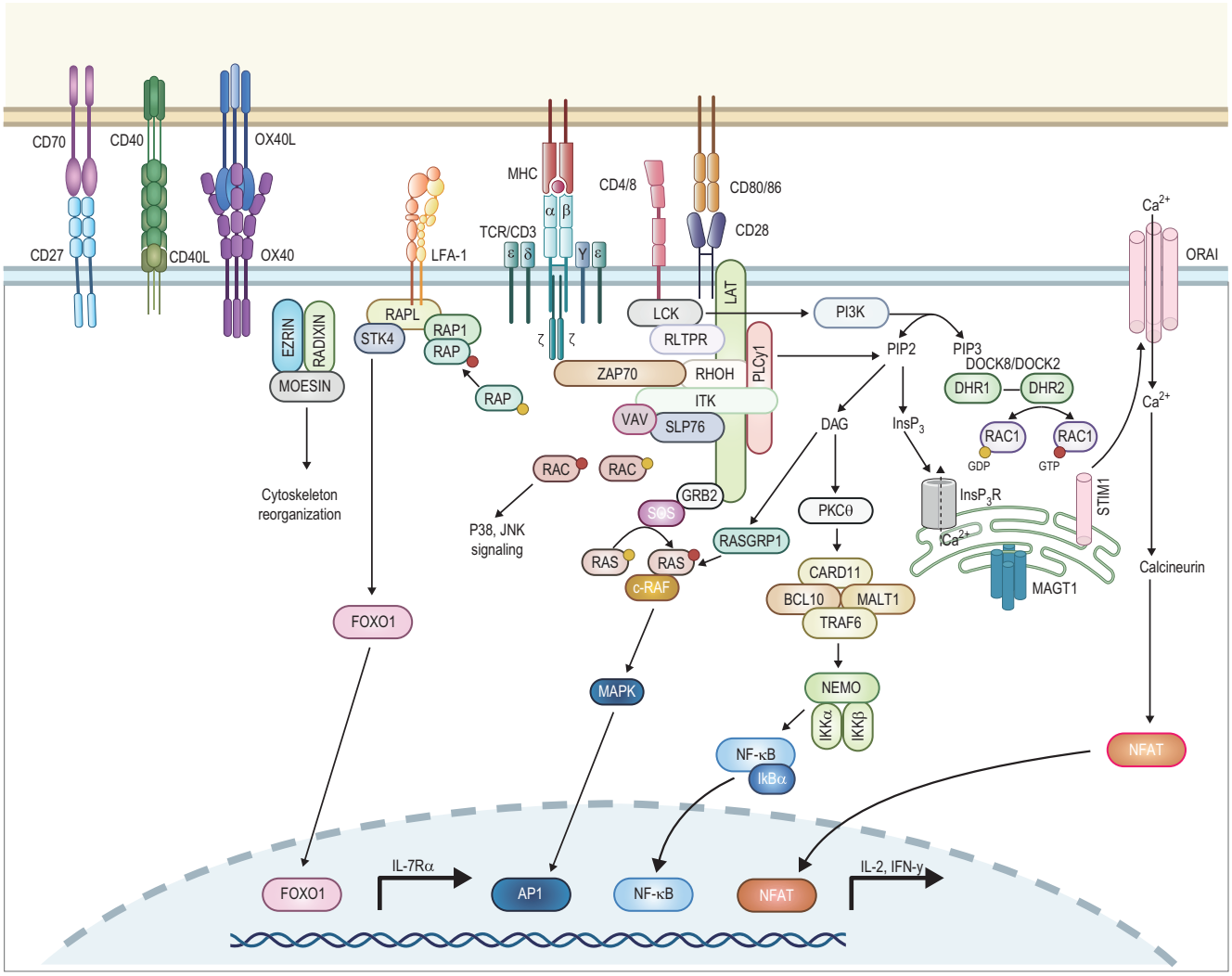


FIG. 34.4 Schematic Representation of Tcr-Mediated Intracellular Signaling and of Receptor/Ligand Pairs That Participate at T-Cell Activation. Activation of T cells through the T-cell receptor *(TCR)* induces a signaling cascade that ultimately results in cytoskeleton reorganization, calcium flux, and activation of transcription factors that drive expression of target genes. CD27, CD40LG, and OX40 are T-cell surface molecules that interact with counter-receptors expressed by other cell types (B cells, dendritic cells, monocyte/macrophages) and contribute to T-cell activation. Several forms of T-cell immunodeficiency are due to mutations of the genes encoding molecules involved in T-cell activation. *AP-1*, Activator protein 1; *BCL10*, B-cell lymphoma/leukemia 10; *CARD11*, caspase recruitment domain family member 11; *DAG*, diacylglycerol; *DHR1/2*, Dock homology region 1/2; *DOCK2*, dedicator of cytokinesis 2; *DOCK8*, dedicator of cytokinesis 8; *GRB2*, growth factor receptor bound protein 2; *IKBα*, inhibitory subunit of NF-κB α; *IKKα*, IκB kinase α; *IKKβ*, IκB kinase β; *$InsP_3$*, inositol 1,4,5-trisphosphate; *$InsP_3R$*, $InsP_3$ receptor; *ITK*, interleukin-2 induced tyrosine kinase; *LAT*, linker of activated T cells; *LCK*, lymphocyte-specific protein tyrosine kinase; *LFA-1* lymphocyte function-associated antigen 1; *MAGT1*, magnesium transporter 1; *MALT1*, mucosa-associated lymphoid tissue lymphoma translocation protein 1; *MAPK*, mitogen-activated protein kinase; *MHC*, major histocompatibility complex; *NEMO*, NF-κB essential modulator; *NFAT*, nuclear factor of activated T cells; *NF-κB*, nuclear factor-κB; *PI3K*, phosphoinositide 3-kinase; *PIP2*, phosphatidylinositol 4,5-bisphosphate; *PIP3*, phosphatidylinositol (3,4,5)-trisphosphate; *PLCγ1*, phospholipase C-γ1; *RAP*, ras-related protein; *RAPL*, regulator of adhesion and polarization enriched in lymphocytes; *RASGRP1*, Ras guanyl releasing protein 1; *RHOH*, Ras homolog family member H; *RLTPR*, RGD-leucin-rich repeat, tropomodulin domain, and proline-rich domain-containing protein; *SLP76*, SH2 domain-containing leukocyte protein of 76 kDa; *SOS*, son of sevenless; *STIM1*, stromal interaction molecule 1; *STK4*, serine/threonine kinase 4; *TCR*, T-cell receptor; *TRAF6*, tumor necrosis factor receptor-associated protein 6; *ZAP-70*, zeta-associated protein of 70 kDa.

SCID DUE NON-HEMATOPOIETIC DEFECTS

Although the vast majority of patients with SCID harbor gene defects that affect hematopoietic and lymphoid development, extra-hematopoietic defects, and in particular defects of thymus development, may also cause severe T-cell lymphopenia at birth. Establishing a correct diagnosis is important because the latter group of patients do not benefit from HSCT but may respond to implantation of thymus tissue from a suitable donor source. In vitro assays analyzing the T-cell differentiation potential of hematopoietic stem cells help to distinguish between hematopoietic cell-autonomous and extra-hematopoietic causes of SCID.[33]

Complete DiGeorge Syndrome

DiGeorge syndrome (DGS) (OMIM #188400) is a multisystem developmental disorder characterized by the triad of congenital heart disease, immunodeficiency and hypoparathyroidism, stemming from defects in tissues derived from the fetal branchial arch and pouch structures. Congenital defects of the cardiac outflow tract include interrupted aortic arch type B and truncus arteriosus; hypocalcemia is due to parathyroid insufficiency, and immunodeficiency is secondary to thymus aplasia or hypoplasia.[34] Other features include developmental disabilities and renal and craniofacial anomalies, including micrognathia, hypertelorism, downward slanting eyes, and ear malformations. However, there is significant phenotypic variability. A third of DGS patients have velopharyngeal incompetence, leading to feeding difficulties and speech delay; 10% have a cleft palate. As young adults, many DGS patients develop social, behavioral, and psychiatric problems. In most cases, DGS is due to a heterozygous interstitial deletion of chromosome 22q11.2. However, approximately 2% of the patients have small deletions in chromosome 10p, and others have intragenic mutations resulting in haploinsufficiency of the *TBX1* gene, located in the DiGeorge minimal deletion region of 22q11.2.

There is a high incidence of autoimmune diseases, such as cytopenia and thyroiditis in DGS, reflecting perturbed thymic architecture and a reduced number of regulatory T cells. Around 20% of DGS patients have naïve T-cell lymphopenia in infancy, with those at the lowest end of the lymphopenic spectrum detected with low TREC levels upon newborn screening. According to the severity of the immunodeficiency, "partial" and "complete" forms of the disease are recognized. Patients with "partial DGS" have a low T-cell count but with residual (>50 cells/μL) naïve T cells. In contrast, approximately 1% of all cases have "complete DGS" with virtual absence of naïve T cells, thereby mirroring what is observed in SCID. The term "complete atypical DGS" refers to oligoclonal expansion of activated T cells in patients with DGS infiltrating various tissues, often associated with clinical manifestations of Omenn syndrome. Cardiovascular anomalies require prompt attention and appropriate medical treatment of hypocalcemia. Experimental thymus implants have become the treatment of choice for patients with complete typical or complete atypical DGS.[35]

CHARGE Syndrome

Coloboma, Heart defects, Atresia of the choanae, Retarded growth, Genital hypoplasia, and Ear anomalies (CHARGE) syndrome is caused by heterozygous mutations of the *CHD7* or the *SEMA3E* genes (OMIM *608892, 608166) that may also affect thymus development. The number and function of T cells may be reduced, such that some patients meet criteria for SCID.[36]

FOXN1 Deficiency

The FOXN1 transcription factor plays a critical role in development of thymic epithelial cells and hair follicles and in regulation of keratin expression. Biallelic *FOXN1* mutations cause thymic aplasia, alopecia totalis, and nail dystrophy (OMIM *601705). The T-cell count at birth is often very low, consistent with a SCID phenotype. In rare cases, the disease may manifest with only immunodeficiency. Heterozygous loss-of-function *FOXN1* mutations may also cause T-cell lymphopenia which, however, is less severe and tends to improve with time.[37]

PAX1 Deficiency

PAX1 is a transcription factor involved in differentiation of the third and fourth pharyngeal pouches. PAX1 deficiency causes a syndromic form of SCID with impaired thymus development associated with ear, facial, and vertebral defects (OMIM *615560). Some patients manifest an Omenn syndrome phenotype.[38]

COMBINED IMMUNE DEFICIENCY—DEFINITION

According to criteria defined by the European Society for Immune Deficiencies, genetically determined combined immune deficiency (CID) comprises a heterogeneous group of conditions characterized by reduced (but not absent) number and/or function of T and B lymphocytes, associated with infections and/or clinical features of immune dysregulation, such as autoimmunity, lymphoproliferation and granuloma formation and/or occurrence of malignancies (Table 34.3).[39] However, it should be noted that these criteria also require exclusion of syndromic disorders associated with impaired T-cell function, such as ataxia telangiectasia (A-T) and cartilage hair hypoplasia (CHH), and yet these conditions are often associated with laboratory and clinical features of CID. Therefore, they are also discussed in this chapter. As compared to infants with SCID, patients with CID often present with later onset (beyond 1-year of age). Poor control of viral infections is a common feature. Along with hypomorphic defects in SCID-associated genes allowing

KEY CONCEPTS

Combined Immunodeficiencies

- Combined immunodeficiencies include a heterogeneous group of conditions with numerical and/or functional defects of T cells. The T-cell lymphopenia (>300 cells/μL) is less severe than in patients with SCID
- Chronic viral infections (EBV, CMV, HPV, molluscum) are particularly common and may be a cause of lymphoproliferative disease and malignancies (lymphoma, squamous cell carcinoma)
- Autoimmune manifestations (cytopenias in particular) and other manifestations of immune dysregulation (granulomas, elevated serum IgE) are also frequent and indicate abnormalities of T-cell homeostasis
- Flow-cytometric analysis of total T cells and of naïve and memory T-cell subsets is an essential test in the diagnosis of combined immunodeficiencies.
- Enumeration of T-cell receptor excision circles (TRECs) in dried blood spots collected at birth detects some, but not all, cases of combined immunodeficiency
- Functional assays, such as in vitro proliferation to mitogens and antigens, are valuable tools in the characterization of the severity of the disease. Other tests, such as cellular radiosensitivity, cytokine expression, regulatory T-cell function, and analysis of T-cell repertoire diversity, may contribute to the diagnosis in selected cases.
- Flow cytometry and Western blotting may be used to investigate expression of specific proteins associated with individual forms of T-cell immunodeficiency. However, while lack of protein expression may contribute to the diagnosis, demonstration of preserved protein expression cannot be used to rule out the diagnosis, because many cases of T-cell immunodeficiencies are due to hypomorphic mutations
- Ultimately, genetic diagnosis permits definitive diagnosis. T-cell immunodeficiency gene panels, whole exome sequencing, and whole genome sequencing are valuable strategies to achieve genetic diagnosis. However, the potential impact of the genetic variants detected must be critically assessed
- Many forms of combined immunodeficiency can be cured with hematopoietic stem cell transplantation. Use of a conditioning regimen is required to attain robust and durable donor stem cell engraftment and immune reconstitution

CMV, Cytomegalovirus; SCID, severe combined immune deficiency.

TABLE 34.3 Diagnostic Criteria of Combined Immune Deficiency (CID) Elaborated by the European Society for Immune Deficiencies (ESID)[39]

At least one of the following:
- At least one severe infection (requiring hospitalization)
- One manifestation of immune dysregulation (autoimmunity, IBD, severe eczema, lymphoproliferation, granuloma)
- Malignancy
- Affected family member

AND 2 of 4 T-cell criteria fulfilled:
- Reduced CD3 or CD4 or CD8 T-cell count (using age-related reference values)
- Reduced naïve CD4 and/or CD^+ T cells
- Elevated TCRγδ T cells
- Reduced proliferation to mitogen or TCR stimulation

AND HIV infection excluded

AND exclusion of a clinical diagnosis associated with CID (e.g., defined syndromic disease, such as dyskeratosis congenita, ataxia-telangiectasia, cartilage hair hypoplasia)

From: Seidel MG, Kindle G, Gathmann B, et al. The European Society for Immunodeficiencies (ESID) Registry Working Definitions for the Clinical Diagnosis of Inborn Errors of Immunity. *J Allergy Clin Immunol Pract.* 2019;7(6):1763–1770.

development of a low number of T cells, a growing number of gene defects that partially compromise T-cell differentiation and/or function, causing CID, have been described.[1]

CID DUE TO METABOLIC DEFECTS

Purine Nucleoside Phosphorylase Deficiency

Purine nucleoside phosphorylase (PNP), an enzyme of the purine salvage pathway, catalyzes the phosphorylation of inosine, guanosine, and deoxyguanosine.[40] In the absence of PNP, high intracellular levels of deoxyguanosine triphosphatase cause lymphoid and neuronal toxicity. Immature thymocytes are particularly susceptible to PNP deficiency. PNP deficiency (OMIM *613179) is inherited as an autosomal recessive trait. Its immunological phenotype is characterized by decreased T-cell counts, causing increased risk of infections. Compared to infants with ADA deficiency, patients with PNP deficiency tend to present after 1 year of age. Bacterial, fungal, and in particular viral infections are common. There is an increased risk of non-Hodgkin lymphoma. Although development of B and NK lymphocytes is often unaffected, there is an increased risk of severe autoimmune hemolytic anemia. Neurodevelopmental delay, hypotonia, and spasticity are frequent and may develop even before immune problems. HSCT represents the only curative option for the immune phenotype, but it is not expected to rescue the neurological phenotype.

Cytidine 5-Triphosphate Synthase 1 Deficiency

Cytidine 5-triphosphate synthase 1 (CTPS1) is involved in de novo synthesis of cytidine 5-triphosphate (CTP), a nucleotide required for DNA and RNA metabolism. CTPS1 expression is upregulated following TCR stimulation. *CTPS1* mutations (OMIM #615897) have been identified in several infants from Northwestern England, who manifested severe bacterial and viral infections since early in life, and an increased risk of EBV-driven non-Hodgkin lymphoma. CD4 T-cell lymphopenia, increased proportion of effector memory T cells, and reduced in vitro proliferation to mitogens and antigens have been reported.[41]

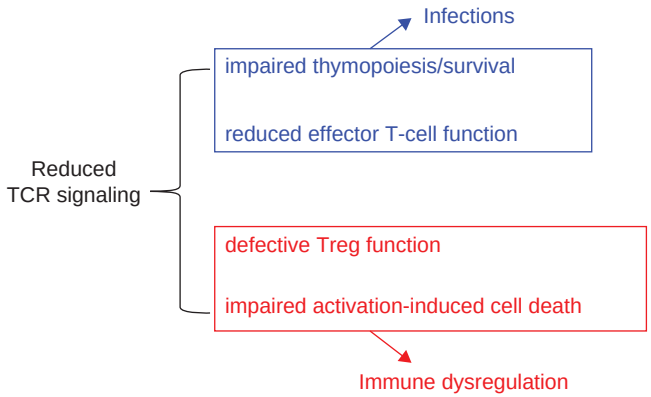


FIG. 34.5 Mechanisms of Increased Susceptibility to Infections and Autoimmunity in Patients with Defects of T-Cell Receptor Signaling.

CID Due to Defects of TCR-Mediated Signaling

Several forms of CID are due to genetic defects that impair intracellular signaling, and in particular signaling through the CD3/TCR complex (see Fig. 34.4). In these disorders, reduced signaling through the pre-TCR and the TCR may compromise generation of a diversified repertoire of T cells in the thymus and impair effector T-cell responses in the periphery, contributing to increased susceptibility to infections. Chronic viral infections (CMV, EBV) are particularly common and may cause lymphoproliferative disease and lymphoma. At the same time, reduced TCR signaling is also responsible for defective Treg function and impaired activation-induced cell death, which result in an increased rate of autoimmune manifestations (Fig. 34.5).

Deficiency of the Zeta-Associated Protein of 70 kDa (ZAP-70)

The ZAP-70 protein plays a critical role in TCR-mediated signaling.[42] Upon TCR ligation, ZAP-70 is recruited to the phosphorylated ITAM motifs of the CD3ζ chain and is itself phosphorylated by LCK. This process induces ZAP-70 kinase activity, permitting phosphorylation of various downstream molecules, including SLP-76, LAT, and PLC-γ1. ZAP-70 deficiency (OMIM *269840) is inherited as an autosomal recessive trait.[43] Clinical manifestations include life-threatening infections of bacterial, viral, and fungal origin beginning early in life, reminiscent of SCID. Some patients manifest with skin rash that may be severe, mirroring Omenn syndrome. A peculiar phenotype predominantly characterized by autoimmunity has been reported in patients in whom one of the alleles carried a gain-of-function mutation.

The typical immunological phenotype of ZAP-70 deficiency includes lack or markedly reduced number of CD8 T cells. CD4 T cells are present in normal number, but in vitro T-cell proliferation to mitogens and antigens is profoundly impaired. The differential diagnosis is with major histocompatibility complex (MHC) class I deficiency and CD8α deficiency, two conditions that are also characterized by a severe reduction of CD8 T cells. Treatment is based on HSCT.

LCK Deficiency

TCR ligation promotes activation of the p56Lck kinase, which mediates phosphorylation of components of the CD3 complex, initiating TCR-mediated signal transduction. LCK deficiency

(OMIM *615758) is a rare cause of CID manifesting with infections, autoimmune cytopenias, and vasculitis, associated with a low CD4 T-cell count and impaired T-cell function.

Deficiency of the IL-2 Inducible Tyrosine Kinase (ITK)

ITK participates in TCR signaling by inducing phosphorylation of PLC-γ1, thereby generating second messengers that activate protein kinase C (PKC), allowing Ca^{2+} release. ITK deficiency (OMIM *186973) is inherited as an autosomal recessive trait. The immunological phenotype is characterized by progressive CD4 T-cell lymphopenia and decreased T-cell function, contributing to a high rate of EBV infection and lymphoproliferative disease, including non-Hodgkin lymphoma in addition to respiratory tract infections. Hypogammaglobulinemia is common.[44]

STK4 Deficiency

STK4 is a kinase that activates various transcription factors, regulating cell survival and proliferation. STK4 deficiency (OMIM *614868) is an autosomal recessive form of CID characterized by recurrent bacterial and viral (Human Papilloma Virus (HPV), EBV, varicella-zoster virus [VZV], molluscum) infections, mucocutaneous candidiasis, lymphoproliferation, autoimmune cytopenias, and increased risk of lymphoma. Congenital heart disease and intermittent neutropenia are other common findings. Immunological abnormalities include CD4 T-cell lymphopenia with a low proportion of naïve T cells and impaired T-cell function.

RHOH Deficiency

RHOH is a small GTPase involved in signal transduction. RHOH deficiency (OMIM *602037) is characterized by increased susceptibility to HPV infection (causing epidermodysplasia verruciformis), recurrent pneumonia, and a higher risk of Burkitt lymphoma. Naïve T-cell lymphopenia, oligoclonal representation of the T-cell repertoire, and decreased T-cell function are typical immunological findings.[44]

T-Cell Receptor α Constant (TRAC) Chain Deficiency

Biallelic mutations of the T-cell receptor α constant chain gene (*TRAC*) cause TRAC deficiency (OMIM #615387), a rare form of SCID characterized by markedly reduced levels of the TCRαβ complex on the surface of T cells.[45] In addition to infections, immune dysregulation (with eczema, eosinophilia) and autoimmunity have been reported.

CID Due to Defects of Nuclear Factor-κB (NF-κB) Signaling

Following TCR signaling, the complex composed of caspase recruitment domain-containing protein (CARD)-11, BCL-10 and MALT1 proteins (CBM complex) is activated, permitting recruitment of TRAF6 and activation of IKK, leading to nuclear translocation of the p50 and p65 subunits of NF-κB and expression of NF-κB–dependent genes. Defects of the CBM complex due to biallelic loss-of function mutations of any one of the *MALT1*, *CARD11*, and *BCL10* genes (OMIM #615469, 615206, 616098) are associated with increased susceptibility to bacterial, viral and fungal infections.[46] Although the number of circulating T lymphocytes is normal, generation of memory T cells is impaired and proliferative responses to CD3 stimulation are decreased. Patients with *CARD11* mutations have a block in B-cell development at the transitional stage. Autosomal dominant CARD11 dominant negative mutations (OMIM #617638) are a cause of CID with bacterial and viral infections, early-onset atopic disease, and autoimmune cytopenias. Immunological abnormalities include skewing to Th2 cells and defective T-cell activation.[46] In contrast, autosomal dominant gain of function CARD11 mutations (OMIM #616452) cause hyperactivation of NF-κB, leading to a phenotype that is distinct from CID and is characterized by B-cell lymphocytosis, splenomegaly, and lymphadenopathy. In vitro T-cell activation through the TCR is reduced, and susceptibility to infections is variable.

The NF-κB family of proteins is composed of five members: NF-κB1 (p105), NF-κB2 (p100), RelA, RelB, and c-Rel.[47] Upon cleavage, the p105 and p100 proteins generate p50 and p52, which are involved in the canonical and noncanonical pathway of NF-kB activation, respectively, by forming a heterodimer with one of the Rel proteins. In the canonical pathway, the NF-κB heterodimer is kept in the cytoplasm by the IκBα protein. Following phosphorylation by the IKK kinase, IκBα is degraded upon phosphorylation by the IKK kinase, and NF-κB can translocate to the nucleus, inducing expression of target genes. The IKK complex is composed of two catalytic subunits (IKK-α and IKK-β, encoded by the *IKBKA* and *IKBKB* genes) and by the regulatory subunit IKK-γ (also known as NEMO), encoded by the *IKBKG* gene. Mutations of *IKBKG* cause an X-linked immunodeficiency characterized by recurrent infections, variable antibody responses, anhidrotic ectodermal dysplasia, and—in some cases—lymphedema or osteopetrosis (OMIM #300291).[48,49] The infections associated with this condition may be sustained by various pathogens, including pyogenic bacteria, viruses, and fungi. *Mycobacterium avium* and *M. kansasii* infections have been reported in up to 40% of the patients. Inflammatory manifestations are common, especially in the gastrointestinal tract.

Autosomal recessive IKBKB deficiency (OMIM #615592) manifests with early-onset severe infections sustained by various bacterial mycobacterial, viral, and fungal pathogens.[50] T and B cells exhibit uniquely a naïve phenotype. Low in vitro T-cell proliferation to mitogens and hypogammaglobulinemia are other typical features.

Autosomal dominant gain-of-function IKBKB mutations (OMIM #618204) cause increased phosphorylation of IKB-α, and therefore enhanced NF-κB activity. The clinical phenotype includes recurrent and severe infections, immune dysregulation, and signs of ectodermal dysplasia. T-cell lymphopenia, impaired generation of memory T and B cells, hypogammaglobulinemia, and defective antibody responses are the main immunological abnormalities.[51]

Heterozygous, gain-of-function mutations of the *IKBA* gene, that prevent phosphorylation and degradation of the protein, cause defective NF-κB activation and are responsible for a CID with ectodermal dystrophy and immune dysregulation (OMIM #612132).[52] Profound B-cell deficiency, hypogammaglobulinemia, and impaired in vitro T-cell proliferation to mitogens are common immunological abnormalities.

NFKB1 haploinsufficiency (OMIM #616576) is caused by heterozygous mutations that reduce levels and phosphorylation of p105, and therefore also of p50, affecting activation of the NF-κB canonical pathway. The phenotype is more often consistent with common variable immunodeficiency.[53] Autosomal

dominant NFKB2 deficiency (OMIM #615577) is caused by dominant negative mutations that impair processing of the p100 precursor, leading to reduced levels of p52 and defective activation of the NF-κB noncanonical pathway. In addition to a CVID-like phenotype, these patients also manifest adrenal insufficiency and alopecia.[53] Gain-of-function *NFKB2* mutations lead to constitutive activation of NF-κB and manifest with CID (recurrent pneumonia, severe EBV and CMV infections, warts), sclerosing cholangitis, and splenomegaly.[54] Low B-cell number, T-cell lymphocytosis, and hypogammaglobulinemia are the main immunological abnormalities.

Autosomal dominant RELA haploinsufficiency (OMIM #618287) is a rare disease characterized by CD4 T-cell lymphoproliferation (with low proportion of naïve T cells), autoimmune cytopenias, and mucocutaneous ulcers.[53] Autosomal recessive RELB (OMIM #617585) and c-REL deficiencies cause CID with recurrent infections and impaired T-cell diversity and function.[53] Decreased number of naïve T cells and low levels of TRECs have been demonstrated in RELB deficiency, and a low proportion of memory T and B cells and B-cell lymphopenia have been reported in c-REL deficiency.

Finally, NF-κB-inducing kinase (NIK) deficiency is caused by biallelic mutations of the *MAP3K14* gene. Patients suffer from recurrent bacterial, viral, and *Cryptosporidium* infections since childhood, associated with B-cell lymphopenia and decreased memory T cells.[53]

Clinical management of the defects of the NF-κB pathway varies depending on the underlying disease. Immunoglobulin replacement therapy and antimicrobial prophylaxis are required in most patients. However, HSCT is typically required to correct the profound immunodeficiency associated with IKBKB, RELB, and NIK deficiency and with gain-of function *IKBKA* mutations. A special case is represented by X-linked NEMO deficiency, because of the variable severity of the clinical phenotype. Moreover, persistence of inflammatory manifestations (especially of the gastrointestinal tract) has been observed in some of these patients, in spite of donor chimerism and immune reconstitution.

CID DUE TO DEFECTS OF MAJOR HISTOCOMPATIBILITY COMPLEX MOLECULE EXPRESSION

Expression of MHC class I and class II molecules is essential for positive selection of CD8 and CD4 cells, respectively, and for the development of adaptive immune responses.

MHC class I deficiency (OMIM *604571) is inherited as an autosomal recessive trait and may be caused by defects in the genes encoding the Transporter associated with Antigen Processing 1 (TAP1) and TAP2, the TAP-binding protein (TABP) TAPASIN, or in the β_2-microglobulin (*B2M*) gene.[55] These defects interfere with intracellular transport of peptide antigens, their loading onto MHC class I molecules, and with cell-surface expression of the complex. The clinical phenotype includes recurrent respiratory infections, chronic inflammatory lung disease, and ulcerative skin lesions with necrotizing granulomas. Glomerulonephritis and herpes zoster infections have been reported in TAPASIN deficiency. The number of circulating CD8 T cells is reduced, but in vitro T-cell proliferation is normal, which facilitates differential diagnosis with ZAP-70 deficiency. Treatment is largely based on supportive interventions. Treatment is supportive. Immunosuppressive drugs may worsen the disease manifestations and should be avoided.

MHC class II deficiency includes autosomal recessive conditions due to defects in genes (*CIITA, RFXANK, RFX5,* and *RFXAP*) that encode transcription factors and regulators that bind to the proximal promoters of the MHC class II genes, allowing their expression. The disease manifests early in life with increased susceptibility to bacterial, viral, and opportunistic respiratory tract infections. Chronic diarrhea and sclerosing cholangitis, often secondary to *Cryptosporidium* or CMV infection, are frequently observed.[55] The number of circulating CD4 T cells is markedly reduced. Hypogammaglobulinemia and poor antibody response to immunization antigens are common. Differential diagnoses include HIV infection and idiopathic CD4 lymphopenia; however, in these conditions, expression of MHC class II molecules is preserved. If untreated, most patients die in infancy or childhood. Antibiotic prophylaxis, immunoglobulin replacement therapy, and nutritional support are mainstays of treatment, but the only definitive cure is represented by HSCT. However, results are worse than in other forms of CID, and graft-versus-host disease is common, especially in patients with preexisting viral infections.[55]

CID Due to Defects of Co-Stimulatory Molecules

Several forms of CID are due to defects in genes that encode cell-surface molecules involved in the interaction of T cells with other cells types, including B lymphocytes, dendritic cells, and monocytes/macrophages.

Defects of CD40 Ligand and CD40

CD40 ligand (CD40LG) is predominantly expressed on the surface of activated CD4 T cells and interacts with CD40 expressed by B cells, monocytes, and dendritic cells, and various other cell types. CD40LG/CD40-mediated cross-talk between activated CD4 T cells and B cells promotes class-switch recombination and memory B-cell generation. In addition, interaction between $CD40LG^+$ T cells and monocytes and dendritic cells induces activation of these myeloid cells, promoting IL-12 production, which in turn acts upon T cells to induce expression of Interferon (IFN). CD40LG deficiency (OMIM *308230) is inherited as an X-linked trait. Affected males suffer from recurrent infections, including recurrent bacterial infections, *P. jirovecii, CMV, Cryptosporidium*, and *Cryptococcus*.[56] Chronic *Cryptosporidium* infections are often complicated by ascending cholangitis and sclerosing cholangitis. There is an increased risk of lymphomas, biliary tract tumors, and peripheral neuroectodermal tumors of the gastrointestinal tract. Chronic or intermittent neutropenia is common. Pure red cell aplasia secondary to parvovirus B19 infection has been reported. Immunological abnormalities include hypogammaglobulinemia with normal or elevated IgM, reduced proportion of switched memory B cells, and impaired T-cell proliferation to antigens. Flow cytometry may be used to document impaired expression of CD40LG, but mutation analysis is ultimately needed to confirm the diagnosis, especially in males expressing mutant forms of the molecule. Use of trimethoprim-sulfamethoxazole to prevent *P. jirovecii* pneumonia and regular immunoglobulin replacement therapy are the mainstays of treatment. Recombinant G-CSF may help in patients with severe neutropenia. Hygiene measures to prevent *Cryptosporidium* infection and laboratory and imaging studies to monitor liver and biliary tract status should be part of the monitoring plan. HSCT is indicated in patients with a more

severe clinical course and may be combined with liver transplantation for patients with liver failure.[56]

CD40 deficiency (OMIM *606843) is an autosomal recessive disease, whose clinical and immunological phenotype recapitulates that of CD40LG deficiency.

ICOS and ICOS Ligand (ICOSLG) Deficiency

The inducible T-cell co-stimulator (ICOS) is expressed on the surface of activated T cells; it promotes cytokine production, T-cell proliferation, and generation of follicular helper T cells.[57] These activities are mediated by ICOS interaction with ICOSL, which is expressed by B cells. Autosomal recessive ICOS and ICOSLG deficiencies are characterized by recurrent infections, autoimmunity, splenomegaly, granulomas, and increased risk of malignancies. In addition to respiratory tract infections, warts due to HPV infection and herpes simplex virus (HSV) and *Cryptococcus* infection have been also reported. Hypogammaglobulinemia with markedly reduced number of switched memory B cells are the main immunological defects in both conditions; however, ICOSLG deficiency is also characterized by T-cell lymphopenia.[58]

OX40 Deficiency

OX40 is a T-cell activation molecule that promotes cell survival. Autosomal recessive OX40 deficiency (OMIM *615593) has been described in a young adult with Kaposi sarcoma. Naïve and memory CD8 T cells were present in higher and lower than normal numbers, respectively. In vitro responses to antigens and IFN-γ production were reduced.

CD27 and CD70 Deficiency

CD27 and CD70 are counter-receptors expressed on the surface of T and B cells, respectively. Autosomal recessive defects of these molecules (OMIM #615122, 618261) are associated with EBV-driven lymphoproliferative disease, including Hodgkin lymphoma and diffuse large B-cell lymphoma.[59] Other manifestations include respiratory tract infections, severe varicella, autoimmunity, lymphadenopathy, hepatosplenomegaly, uveitis, and aphthous ulcers. Hypogammaglobulinemia has been frequently reported. HSCT can correct the disease phenotype.

CID DUE TO DEFECTS OF DNA REPAIR AND DNA REPLICATION

Ataxia-telangiectasia (A-T, OMIM *208900) is an autosomal recessive condition characterized by progressive neurodegeneration, development of telangiectasias, a variable degree of immunodeficiency, and increased risk of malignancies.[60] The disease is caused by mutations of the Ataxia-Telangiectasia Mutated (*ATM*) gene, which encodes a protein kinase that is recruited to sites of DNA double-strand breaks. ATM phosphorylates multiple substrates involved in DNA repair, cell cycle checkpoints, intracellular signaling, and gene transcription, and therefore controls multiple cellular processes. Increased cellular radiosensitivity is a prominent feature of the disease. Patients with A-T manifest oculocutaneous telangiectasias at around 2 to 3 years of age. Neurological problems (ataxia, dysarthria, oculomotor apraxia, hypotonia, tremors) appear a few years later, and progress, making these patients wheel-chair dependent by age 10 to15. The neurological features of the disease are associated with cerebellar involution and degeneration of Purkinje cells. The immunodeficiency of A-T manifests mainly with increased susceptibility to respiratory tract infections. A significant proportion of patients develop cutaneous granulomas. In some cases, presence of the rubella virus vaccine strain has been documented. Autoimmunity (in particular, diabetes mellitus) is also frequently observed. Other endocrinopathies include hypogonadism and growth hormone deficiency. There is accelerated aging, marked by development of progeric somatic changes. Up to 20% to 30% of A-T patients develop cancer in the course of their life. In most cases, malignancies are of lymphoid origin, and translocations involving chromosomes 7 and 14 (with breakpoints at the TCR and immunoglobulin loci) are frequently present in the malignant clones. Tumors of non-hematological origin are more frequently observed in older patients. The disease has a dismal prognosis, and death occurs in most cases by the fourth or fifth decade of life. Immunological abnormalities include a variable degree of T- and B-cell lymphopenia and hypogammaglobulinemia. Severe T-cell lymphopenia may already be present at birth, resulting in low TREC levels at newborn screening[61] but may worsen over time in other patients. The thymus is hypoplastic and lacks Hassall corpuscles. Low levels of IgA and IgG2, and high levels of serum IgM are the main abnormalities of humoral immunity. The high IgM also reflects the presence of monomeric IgM molecules. Elevated levels of serum alpha-fetoprotein (AFP) are a hallmark of the disease; however, this test cannot be used at birth because normal newborns also have elevated AFP levels. Therefore, for infants manifesting with T-cell lymphopenia at birth, mutation analysis at the *ATM* locus, and specific tests to assess ATM protein expression and phosphorylation are necessary to rule out A-T. Treatment is largely supportive. Administration of immunoglobulins and antimicrobial prophylaxis are beneficial in patients with recurrent infections. Exposure to ionizing radiation should be minimized. No definitive cure is currently available.

Nijmegen breakage syndrome (NBS, OMIM *251260) is an autosomal recessive disease due to mutations in the gene encoding nibrin, involved in DNA double-strand break, meiotic recombination, and telomere length maintenance.[60] Clinical manifestations include microcephaly, facial dysmorphisms, recurrent respiratory tract infections, cutaneous manifestations (granulomas, vitiligo, café-au-lait spots), and autoimmunity (cytopenias, thyroiditis, celiac disease, interstitial lung disease). More than 50% of the patients develop malignancies, in particular leukemias and lymphomas, by the third decade of life. Other solid tumors are more common in older patients. Immunological defects include T- and B-cell lymphopenia, reduced T-cell proliferative responses, and poor antibody responses. Treatment is largely supportive, but HSCT with reduced intensity conditioning has been successfully performed in patients with lymphoma.[62] Exposure to ionizing radiations should be avoided.

Bloom syndrome (OMIM *210900) is due to a defect in a helicase that regulates DNA replication and DNA double-strand break repair. Clinical manifestations include short stature, microcephaly, erythematous skin lesions, recurrent respiratory tract infections, autoimmunity, and increased risk of cancer (especially carcinomas, lymphomas, and leukemias) associated with cellular radiosensitivity. The immunodeficiency is rarely severe.

ERCC6L2 deficiency (OMIM *615715) is due to a defect of another helicase. Patients manifest short stature, microcephaly, variable degrees of developmental delay, and bone marrow failure. The number of B and T cells (especially of naïve $CD4^+$ cells) is reduced.

DNA ligase I deficiency is another autosomal recessive DNA repair defect, causing immunodeficiency of variable severity, ranging from hypogammaglobulinemia to profound CID (with severe lymphopenia) requiring HSCT. Learning disability may be present. Deficiency of NSMCE3 causes lung disease, Immunodeficiency, and Chromosomal breakage Syndrome (LICS, OMIM *617241), a condition with failure to thrive and severe viral infections leading to lung damage and early death. The number and function of T cells are reduced. RIDDLE syndrome (OMIM *611943) is a rare disorder characterized by Radiosensitivity, Immune Deficiency, facial Dysmorphisms, LEarning disability, and short stature. The disease is due to mutations in the *RNF168* gene encoding an E3 ubiquitin ligase involved in DNA double-strand break repair. Hypogammaglobulinemia with elevated IgM is a common immunological abnormality.

Several forms of immunodeficiency are due to defects of proteins involved in the initiation and progression of DNA replication. DNA polymerase δ (Polδ) deficiency is a CID due to biallelic loss-of-function mutations of the *POLD1* and *POLD2* genes, encoding two of the four subunits of Polδ, which plays an important role in DNA replication and maintenance of genomic stability. Clinical features include recurrent bacterial and viral infections and intellectual disability; short stature has also been observed in Polδ2 deficiency.[63,64] T-cell lymphopenia and hypogammaglobulinemia are the main immunological abnormalities.

DNA polymerase ε (POLE) deficiency includes defects of the POLE1 and the POLE2 subunits. Both disorders are inherited as autosomal recessive traits. POLE1 deficiency may manifest with facial dysmorphisms, immunodeficiency, livedo, and short stature (FILS syndrome, OMIM #615139) or with intrauterine growth retardation, metaphyseal dysplasia, congenital adrenal hypoplasia, genital anomalies, and immune deficiency (IMAGE-I, OMIM #618336) syndrome. A low number of naïve T cells and of memory B cells, decreased T-cell proliferation, reduced levels of serum IgM, and impaired antibody response to polysaccharide antigens have been frequently reported. POLE2 deficiency is characterized by dysmorphisms, microcephaly, growth retardation, and increased susceptibility to severe infections and autoimmunity, associated with severe T-cell lymphopenia, impaired proliferation to mitogens, and a reduced number of NK cells. GINS1 deficiency (OMIM *617827) is a rare disease characterized by growth retardation, mild skeletal and facial anomalies, and recurrent bacterial and viral infections associated with low T-cell count and profound NK-cell deficiency. Neutropenia with reduced number of mature granulocytes in the bone marrow is frequently observed. MCM4 deficiency (OMIM *602638) is characterized by growth retardation, adrenal insufficiency, severe viral infections, and a profound defect of NK lymphocytes. There is increased risk of malignancies associated with cellular radiosensitivity. Finally, Immune deficiency, Centromeric instability, and Facial anomalies (ICF syndrome) includes a group of disorders with mutations in the *DNMT3B (OMIM #242860)*, *ZBTB24 (OMIM #614069)*, or less frequently in the *CDCA7* (OMIM #616910) and *HELLS* (OMIM #616911) genes.[65] Infections are a prominent feature and a leading cause of death early in life. Besides facial dysmorphisms, neurodevelopmental and gross motor skill deficits, congenital malformations (macroglossia, hypospadias, cleft palate, syndactyly), and chronic diarrhea are common findings. Profound hypo-gammaglobulinemia with normal B-cell numbers is the main immunological abnormality, but progressive decline of T and B cells has been reported.

CID ASSOCIATED WITH CYTOSKELETAL DEFECTS

Several inborn errors of immunity are characterized by defects of the actin cytoskeleton that compromise the response to cell-surface receptor engagement, intracellular signaling, and cell motility.

Wiskott-Aldrich Syndrome and Related Disorders

Wiskott-Aldrich syndrome (WAS, OMIM *301000) is an X-linked disorder due to a defect in the gene encoding the WAS protein (*WASP*), which regulates actin polymerization in hematopoietic cells by interacting with the Arp2/3 complex.[66] Clinical manifestations include chronic thrombocytopenia, eczema, recurrent infections, immunodeficiency, and a high incidence of autoimmune diseases and malignancies The thrombocytopenia of WAS is associated with reduced platelet size, and is responsible for hemorrhagic manifestations of variable severity, ranging from petechiae and bruises to severe gastrointestinal and brain hemorrhages. Because of defective innate and adaptive immunity, patients with classic WAS are highly susceptible to bacterial, fungal, and viral infections. Autoimmune complications occur in 70% of WAS patients, most frequently presenting as cytopenias, arthritis, vasculitis, and inflammatory bowel disease. Lymphoma, often EBV-associated, and leukemias are common hematological malignancies in patients with WAS. A milder phenotype, X-linked thrombocytopenia (XLT), is associated with mutations that result in residual expression of mutated protein. Although XLT patients manifest only thrombocytopenia early in life (sometimes associated with mild eczema), severe infections, autoimmunity, and malignancy may occur at an older age. The immunological abnormalities of WAS include reduced T- and B-cell counts, impaired T-cell proliferation to anti-CD3, low serum IgM with elevated IgA and IgE, defective antibody responses to polysaccharide antigens, poor NK-cell and Treg function, and defective directional migration of lymphoid and myeloid cells. The T-cell lymphopenia is often progressive but may occasionally already be present at birth and be detected by low TREC levels at newborn screening. Flow-cytometric analysis of WASP expression may help in the diagnosis, which is ultimately confirmed by mutation analysis. Treatment includes antibiotic prophylaxis and IgG replacement therapy. Immunosuppressive therapy may be needed if autoimmune symptoms occur. Splenectomy ameliorates the bleeding tendency by increasing the number of circulating platelets. However, it substantially increases the risk of systemic infections due to encapsulated bacteria. HSCT is the treatment of choice, and excellent outcomes have been reported in recent years. After initial attempts with gene therapy based on use of gammaretroviral vectors was associated with a high rate of leukemias due to insertional mutagenesis, more promising results have been recently reported with use of self-inactivating lentiviral vectors.

The WASP-interacting protein (WIP) binds to and stabilizes WASP. WIP deficiency (OMIM *614493) is an autosomal recessive condition whose clinical and immunological phenotype mimics WAS.[67] Intracellular levels of WASP are reduced. HSCT can cure the disease.

ARPC1B deficiency (OMIM *617718) is an autosomal recessive disease due to mutations of the Arp2/3 complex component ARPC1B.[68] The clinical phenotype consists of eczema, food allergies, autoimmune complications, recurrent bacterial and viral infections, and mild bleeding tendency. The thrombocytopenia is less severe than in WAS. Abnormal chemotaxis, impaired NK- and Treg-cell function, and markedly increased IgA and IgE

levels are the main immunological abnormalities. Treatment is symptomatic, but HSCT has been successfully used in patients with severe manifestations. Autosomal dominant CDC42 deficiency (OMIM *616737) is characterized by growth dysregulation, facial dysmorphisms, recurrent infections, developmental delay, and macrothrombocytopenia. CDC42 encodes a small GTPase of the RHO family that modulates multiple signaling pathways controlling cell migration, endocytosis, and cell cycle.

Dedicator of Cytokinesis 8 (DOCK8) Deficiency

DOCK8 is a guanine nucleotide exchange factor (GEF) that regulates cytoskeleton reorganization and intracellular signaling in hematopoietic cells.[69] DOCK8 deficiency (OMIM #243700) is inherited as an autosomal recessive trait and is characterized by recurrent and severe bacterial, fungal, and viral infections, eczema, and immune dysregulation. Cutaneous viral infections (warts, molluscum contagiosum, herpes simplex) are especially common, and warts often progress to squamous cell carcinoma. Patients are also at increased risk of systemic infections due to VZV, CMV, and EBV. Vascular thrombosis in the central nervous system has been described in several patients. Lymphopenia affecting especially naïve T cells, increased proportion of CD8 T_{EMRA} cells, decreased in vitro proliferation to CD3 stimulation, and defects of NK cytolytic function are characteristic defects of cellular immunity. High IgE and low IgM serum levels are observed in most patients, and specific antibody responses are blunted.[69] If untreated, the disease has a poor prognosis but can be cured by allogeneic HSCT.[70,71]

DOCK2 Deficiency

DOCK2 is another GEF that regulates cytoskeleton reorganization in response to engagement of TCR, BCR, and various chemokine receptors. Autosomal recessive DOCK2 deficiency (OMIM #616433) is characterized by early-onset invasive bacterial and viral infections. Lymphopenia and defective function of T-, B-, and NK-cell responses are common findings.[72] Lymphocyte migration in response to chemokines is reduced,[72] and neutrophil production of reactive oxygen species is partially impaired.[73] Production of type 1 interferon by immune and nonimmune cells upon viral infections is markedly defective. HSCT is the only curative approach.

Moesin Deficiency

Moesin is a member of the ezrin-radixin-moesin family of proteins that link the cortical actin filaments to the cell membrane. Moesin deficiency (OMIM #300988) is inherited as an X-linked trait. Patients suffer from bacterial and viral infections. Severe VZV infection has been reported in several cases. Lymphopenia, hypogammaglobulinemia, and intermittent neutropenia are immunological findings.[74] Low levels of TRECs may be detected at birth. However, in most cases HSCT is not necessary, and the disease can be managed with immunoglobulin replacement therapy and prophylactic antibiotics.[75]

CID DUE TO DEFECTS OF CLATHRIN-MEDIATED ENDOCYTOSIS AND IRON INTERNALIZATION

Iron internalization plays an important role in promoting intracellular metabolism. Transferrin receptor deficiency (OMIM #616740) is an autosomal recessive CID due to a homozygous missense mutation in the transferrin receptor (*TFRC)* gene, impeding transferrin receptor and iron internalization. There are normal numbers, but reduced function of T and B cells, intermittent neutropenia and thrombocytopenia, and mild anemia. HSCT corrects the disease peneotype.[76]

The FCHO1 protein is involved in the early stages of clathrin-mediated endocytosis. Autosomal recessive FCHO1 deficiency (OMIM #619164) is a CID with low number and impaired function of T cells, and recurrent infections sustained by various pathogens. The disease can be cured by HSCT.[77]

CID With Immune-Osseous Dysplasia

Cartilage Hair Hypoplasia

CHH (OMIM *250250) is an autosomal recessive condition characterized by short-limbed dwarfism and light-colored, hypoplastic hair.[78] Bone marrow dysplasia, increased susceptibility to malignancies, Hirschsprung disease, defects of spermatogenesis, and a variable degree of immune deficiency are other typical features. The disease is more common in certain populations, such as the Amish and Finns, and is caused by mutations in the gene encoding an untranslated RNA component of the ribonuclease mitochondrial RNA processing (*RMRP*) complex, which is involved in cleavage of ribosomal RNA, processing of mitochondrial RNA, and cell-cycle control. T-cell lymphopenia in some patients may be so severe to manifest as SCID with low TREC levels at birth. Impaired in vitro proliferative responses to mitogens is another manifestation of impaired cellular immunity, which may cause increased susceptibility to severe varicella and other viral infections. Defects of humoral immunity may also be observed and contribute to recurrent respiratory tract infections and bronchiectasis. Autoimmune manifestations (hemolytic anemia, neutropenia, and thrombocytopenia) occur in about 10% of the patients. Bone marrow dysplasia may manifest with anemia, leukopenia, and thrombocytopenia. HSCT has been successfully used to correct CHH with severe immunodeficiency.

Schimke Syndrome

Schimke syndrome (OMIM *606622) is an autosomal recessive disorder characterized by short stature due to spondyloepiphyseal dysplasia, progressive renal impairment evolving to renal failure, facial dysmorphisms, micro/hypodontia, immunodeficiency (ranging from moderate T-cell lymphopenia to SCID), increased occurrence of marrow failure, and early onset of CNS vasculopathy and strokes.[79] The disease is caused by mutations of the *SMARCAL1* gene that encodes a chromatin remodeling protein. Recurrent infections of bacterial, viral, and fungal origin are seen in half of the patients. Severe presentations lead to death in the first decade of life, and development of renal failure is common among those who survive. Combined HSCT and renal transplantation may correct immune deficiency and renal problems.

Phosphoglucomutase 3 (PGM3) Deficiency

PGM3 deficiency (OMIM *615816) is a congenital disorder of glycosylation with a broad clinical phenotype that includes severe atopy, scoliosis, short stature, developmental delay, bone marrow failure, and recurrent bacterial, viral, and fungal infections.[80] Chronic lung disease and persistent EBV viremia have been reported in several patients. Markedly elevated IgE serum levels and eosinophilia are common findings. Some patients have severe T-cell lymphopenia causing a SCID phenotype.

Other Immuno-Osseous Dysplasias

Roifman syndrome (OMIM *616651) is an autosomal recessive spondyloepiphyseal dysplasia with extreme intrauterine growth retardation, retinal dystrophy, dysmorphisms, recurrent respiratory and cutaneous infections, and lymphadenopathy. Atopic and autoimmune manifestations are also frequently seen. The disease is due to mutations of the *RNU4ATAC* gene, encoding a small nuclear RNA (snRNA). Immunological abnormalities include hypogammaglobulinemia, low number of T and memory B cells, defective T-cell proliferation to antigens, and decreased NK cell function. Immunoglobulin replacement therapy is used to control infections.

The exosostin-like glycosyltransferase 3 (EXTL3) enzyme is involved in heparan sulfate biosynthesis. Autosomal recessive *EXTL3* deficiency (OMIM *605744) is characterized by syndromic manifestations, with immune deficiency ranging from SCID to moderate T-cell lymphopenia, associated with severe skeletal abnormalities, microcephaly, neurodevelopmental delay, and narrowing of the laryngotracheal tract.[81] *MYSM1* deficiency (OMIM *618116) manifests with postnatal short stature, neurodevelopmental delay, microcephaly, increased risk of bone marrow failure and myelodysplasia, and recurrent infections. Impaired B-cell development in the bone marrow, T-cell lymphopenia, and hypogammaglobulinemia are the main immunological abnormalities. Spondyloenchondrodysplasia with immune dysregulation (SPENCDI, OMIM *607944) is an autosomal recessive condition due to mutations of the *ACP5* gene, encoding tartrate-resistant acid phosphatase. Patients manifest bone abnormalities, mild developmental delay, spasticity, cerebral calcifications, and autoimmunity (lupus, vasculitis, pancytopenia). Cellular and humoral immune defects may be variably present.

CID ASSOCIATED WITH PROMINENT IMMUNE DYSREGULATION

Although CID is often associated with autoimmunity, in some forms of CID immune dysregulation represents a prominent feature and may even dominate the clinical picture.

IL-2 plays a critical role in immune homeostasis by promoting Treg fitness and function. CD25 deficiency (OMIM *606367), caused by mutations of the *IL2RA* gene, encoding the IL-2R α chain, is a CID manifesting with chronic diarrhea, eczema, endocrinopathies, autoimmune cytopenias, lymphocytic infiltrates in various organs, lymphadenopathy, hepatosplenomegaly, and increased risk of severe infections (CMV pneumonitis and colitis, other respiratory tract infections, chronic EBV viremia).[82] IL2RB deficiency (OMIM *618495), due to mutations of the gene encoding the IL-2R β chain, is characterized by a similar phenotype, with autoimmune colitis, hepatosplenomegaly, cytopenias, food allergy, and severe infections, in particular due to CMV and EBV. Downstream from the IL-2R and from other γ_c-containing receptors, STAT5B deficiency (OMIM *245590) manifests with autoimmunity (thyroiditis, thrombocytopenia, arthritis), eczema, viral infections, and pulmonary disease. Short stature with growth hormone insensitivity is a cardinal feature of this condition, reflecting the role of STAT5B in signaling also through the growth hormone receptor.

Signal transducer and activator of transcription 1 (STAT1) gain-of-function mutations cause an autosomal dominant disorder (OMIM #614162) that was initially reported to manifest with chronic mucocutaneous candidiasis (CMC). However, it was rapidly found that autoimmunity (cytopenias, enteropathy, type 1 diabetes, thyroiditis, alopecia, psoriasis), recurrent infections of bacterial, viral. and fungal origin, vascular aneurysms, eczema, and asthma are also common features.[83] Immunological abnormalities include a high number of circulating follicular helper-T (Tfh) cells and impaired generation of Th17 cells. Antifungals are used to treat CMC, and other infections require a targeted drug approach. JAK inhibitors have shown promising results in the treatment of autoimmunity but may aggravate viral and fungal infections. Results of HSCT remain unsatisfactory.

Upon cellular activation, the STIM1 molecule senses the Ca^{2+} concentration in the endoplasmic reticulum and activates Ca^{2+} release-activated channels (CRAC). ORAI1 constitutes the pore-forming subunits of the CRAC in the cell membrane. Mutations of *STIM1* and *ORAI1* genes (OMIM *612783 and 612782) cause an autosomal recessive immunodeficiency with increased susceptibility to severe infections, autoimmunity (cytopenia, hepatosplenomegaly), nonprogressive myopathy, and ectodermal dysplasia.[84] T cells are present in a normal number, but their proliferative capacity is decreased. Allogeneic HSCT can correct the immune defect.

X-linked immunodeficiency with magnesium defect, EBV infection, and neoplasia (XMEN disease, OMIM #300853) is an X-linked disorder due to mutations of the *MAGT1* gene that encodes for a magnesium transporter that also plays a role in protein glycosylation.[85] Patients suffer from chronic EBV infection, with high rate of progression to lymphoma. Recurrent infections due to other viruses have been also reported. Impaired T-cell activation and reduced expression of NKG2D on NK and CD8 T cells are the main immunological findings.

IL-21/IL-21R interaction plays a critical role in promoting B-cell maturation and plasmablast generation. Autosomal recessive IL-21 and IL-21R deficiencies (OMIM *615767 and 615207) are characterized by recurrent sinopulmonary infections.[86] In addition, *Cryptosporidium* infection and liver disease have been reported in IL-21R deficiency and severe early-onset colitis in IL-21 deficiency. Immunological abnormalities include B-cell lymphopenia, hypogammaglobulinemia with elevated IgE, and impaired T-cell proliferation to antigens.

OTHER FORMS OF CID DUE TO DEFECTS OF TRANSCRIPTION FACTORS

The IKAROS transcription factor, encoded by the *IKZF1* gene, regulates gene expression in hematopoietic cells via chromatin remodeling. Dominant negative *IKZF1* mutations cause a CID with increased susceptibility to bacterial and viral infections and *P. jirovecii* pneumonia, and to T-cell acute lymphoblastic leukemia.[87] Laboratory studies demonstrate a low number of T and B cells (with absence of memory cells), neutrophils, and eosinophils. HSCT can rescue the phenotype. A de novo, heterozygous, dominant-negative mutation in the transcription factor *BCL11B* has been identified in an infant with a profound numerical and functional T-cell deficiency and low TRECs at birth (OMIM #617237). Craniofacial and dental abnormalities, and absence of corpus callosum, were also present. The immune defects were reversed by HSCT.[88]

MANAGEMENT OF PATIENTS WITH CID

Irrespective of the nature of the underlying gene defect, all forms of T-cell immunodeficiency are characterized by significant morbidity, with increased susceptibility to infections, often associated with a high frequency of autoimmunity and malignancies. Careful review of the patient's medical and family history and evaluation of the immune status (with enumeration of lymphocyte subsets, distribution of naïve and memory subsets, and measurement of serum immunoglobulins and specific antibody responses) may facilitate recognition of immune deficiency. However because of the heterogeneity of CID and the overlapping clinical and laboratory features, genetic tests have become an essential step in the diagnostic approach to these disorders. Comprehensive gene panels, WES, and WGS represent valuable tools to identify the gene defect. Yet, because many mutations are private (i.e., not previously reported in other patients), functional validation with protein expression and functional studies are often necessary to demonstrate the disease-causing role of gene variants. Antimicrobial prophylaxis and immunoglobulin replacement therapy are beneficial in most forms of CID. Major caution should be used in the administration of live vaccines. Unless strictly necessary, exposure to ionizing radiation and to alkylating agents should be avoided in patients with possible defects in DNA repair. Regular follow-up visits with proper clinical, laboratory, and imaging studies should be planned for patients whose CID is associated with increased risk of malignancy. Finally, HLA typing and search for possible donors should be performed in patients whose CID may require HSCT.

THERAPEUTIC PRINCIPLES

Clinical Management of Patients With Severe Combined Immune Deficiency

- Strict hygiene and isolation measures, use of prophylactic antimicrobials, and immunoglobulin replacement therapy are the mainstay of treatment while preparing for definitive cure
- Immunization with live attenuated vaccines should be strictly avoided
- Breastfeeding is a potential source of cytomegalovirus (CMV) transmission to severe combined immune deficiency (SCID) babies and should be discontinued if the mother is CMV seropositive
- Blood products should be from CMV-negative donors and should be irradiated
- Some forms of T-cell immunodeficiency are associated with radiation sensitivity. In these patients, exposure to ionizing radiation and alkylating agents may cause severe, even fatal, consequences and should be avoided
- If untreated, SCID is inevitably fatal within the first years of life. However, hematopoietic stem cell transplantation (HSCT) can cure the disease. By identifying babies with SCID at birth, before the disease becomes clinically manifest, newborn screening permits prompt referral to HSCT, with improved outcome (>90% survival)

ON THE HORIZON

- Implementation of newborn screening will facilitate diagnosis and further improve clinical outcome of severe combined immune deficiency worldwide. Whole exome sequencing of DNA extracted by dried blood spots may permit newborn screening also of other congenital T-cell immunodeficiencies
- Databases of gene variants for which functional validation of the disease-causing effect has been obtained will help in the diagnostic approach
- Multicenter studies are needed to better define the natural history and the response to treatment of rare forms of combined immunodeficiency
- Development of novel drugs is needed to prevent and treat viral infections in patients with severe T-cell disorders
- Hematopoietic stem cell transplantation protocols tailored to treat individual gene defects may help further improve clinical outcome
- Gene therapy will likely be extended to treat a broader group of T-cell immunodeficiencies. Gene editing is a novel and attractive therapeutic approach, especially for disorders due to gain-of-function and dominant negative mutations

REFERENCES

1. Tangye SG, Al-Herz W, Bousfiha A, et al. Human inborn errors of immunity: 2019 update on the classification from the International Union of Immunological Societies Expert Committee. *J Clin Immunol*. 2020;40(1):24–64.
2. Fischer A, Notarangelo LD, Neven B, et al. Severe combined immunodeficiencies and related disorders. *Nat Rev Dis Primers*. 2015;1:15061.
3. Wahlstrom J, Patel K, Eckhert E, et al. Transplacental maternal engraftment and posttransplantation graft-versus-host disease in children with severe combined immunodeficiency. *J Allergy Clin Immunol*. 2017;139(2):628–633. e610.
4. Douek DC, McFarland RD, Keiser PH, et al. Changes in thymic function with age and during the treatment of HIV infection. *Nature*. 1998;396(6712):690–695.
5. Hazenberg MD, Otto SA, Cohen Stuart JW, et al. Increased cell division but not thymic dysfunction rapidly affects the T-cell receptor excision circle content of the naive T cell population in HIV-1 infection. *Nat Med*. 2000;6(9):1036–1042.
6. Chan K, Puck JM. Development of population-based newborn screening for severe combined immunodeficiency. *J Allergy Clin Immunol*. 2005;115(2):391–398.
7. Kwan A, Abraham RS, Currier R, et al. Newborn screening for severe combined immunodeficiency in 11 screening programs in the United States. *JAMA*. 2014;312(7):729–738.
8. Dvorak CC, Haddad E, Buckley RH, et al. The genetic landscape of severe combined immunodeficiency in the United States and Canada in the current era (2010–2018). *J Allergy Clin Immunol*. 2019;143(1):405–407.
9. Flinn AM, Gennery AR. Adenosine deaminase deficiency: a review. *Orphanet J Rare Dis*. 2018;13(1):65.
10. Whitmore KV, Gaspar HB. Adenosine deaminase deficiency—more than just an immunodeficiency. *Front Immunol*. 2016;7:314.
11. la Marca G, Canessa C, Giocaliere E, et al. Tandem mass spectrometry, but not T-cell receptor excision circle analysis, identifies newborns with late-onset adenosine deaminase deficiency. *J Allergy Clin Immunol*. 2013;131(6):1604–1610.
12. Kohn DB, Hershfield MS, Puck JM, et al. Consensus approach for the management of severe combined immune deficiency caused by adenosine deaminase deficiency. *J Allergy Clin Immunol*. 2019;143(3):852–863.
13. Hoenig M, Pannicke U, Gaspar HB, et al. Recent advances in understanding the pathogenesis and management of reticular dysgenesis. *Br J Haematol*. 2018;180(5):644–653.
14. Chou J, Alazami AM, Jaber F, et al. Hypomorphic variants in AK2 reveal the contribution of mitochondrial function to B cell activation. *J Allergy Clin Immunol*. 2020;146(1):192–202.
15. Noguchi M, Yi H, Rosenblatt HM, et al. Interleukin-2 receptor gamma chain mutation results in X-linked severe combined immunodeficiency in humans. *Cell*. 1993;73(1):147–157.
16. Leonard WJ, Lin JX, O'Shea JJ. The γc family of cytokines: basic biology to therapeuticutic ramifications. *Immunity*. 2019;50(4):832–850.
17. Recher M, Berglund LJ, Avery DT, et al. IL-21 is the primary common gamma chain-binding cytokine required for human B-cell differentiation in vivo. *Blood*. 2011;118(26):6824–6835.
18. Fischer A, Hacein-Bey-Abina S. Gene therapy for severe combined immunodeficiencies and beyond. *J Exp Med*. 2020;217(2):e20190607.
19. Macchi P, Villa A, Giliani S, et al. Mutations of Jak-3 gene in patients with autosomal severe combined immune deficiency (SCID). *Nature*. 1995;377(6544):65–68.

20. Notarangelo LD, Kim MS, Walter JE, Lee YN. Human RAG mutations: biochemistry and clinical implications. *Nat Rev Immunol*. 2016;16(4):234–246.
21. Villa A, Notarangelo LD. RAG gene defects at the verge of immunodeficiency and immune dysregulation. *Immunol Rev*. 2019;287(1):73–90.
22. Farmer JR, Foldvari Z, Ujhazi B, et al. Outcomes and treatment strategies for autoimmunity and hyperinflammation in patients with RAG deficiency. *J Allergy Clin Immunol Pract*. 2019;7(6):1970–1985. e1974.
23. Perelygina L, Chen MH, Suppiah S, et al. Infectious vaccine-derived rubella viruses emerge, persist, and evolve in cutaneous granulomas of children with primary immunodeficiencies. *PLoS Pathog*. 2019;15(10):e1008080.
24. Haddad E, Logan BR, Griffith LM, et al. SCID genotype and 6-month post-transplant CD4 count predict survival and immune recovery: a PIDTC retrospective study. *Blood*. 2018;132(17):1737–1749.
25. Woodbine L, Gennery AR, Jeggo PA. The clinical impact of deficiency in DNA non-homologous end-joining. *DNA Repair (Amst)*. 2014;16:84–96.
26. Fischer A, de Saint Basile G, Le Deist F. CD3 deficiencies. *Curr Opin Allergy Clin Immunol*. 2005;5(6):491–495.
27. Bacchelli C, Moretti FA, Carmo M, et al. Mutations in linker for activation of T cells (LAT) lead to a novel form of severe combined immunodeficiency. *J Allergy Clin Immunol*. 2017;139(2):634–642. e635.
28. Moshous D, de Villartay JP. The expanding spectrum of human coronin 1A deficiency. *Curr Allergy Asthma Rep*. 2014;14(12):481.
29. Hsu AP, Donko A, Arrington ME, et al. Dominant activating RAC2 mutation with lymphopenia, immunodeficiency, and cytoskeletal defects. *Blood*. 2019;133(18):1977–1988.
30. Notarangelo LD. Multiple intestinal atresia with combined immune deficiency. *Curr Opin Pediatr*. 2014;26(6):690–696.
31. Cliffe ST, Bloch DB, Suryani S, et al. Clinical, molecular, and cellular immunological findings in patients with SP110-associated veno-occlusive disease with immunodeficiency syndrome. *J Allergy Clin Immunol*. 2012;130(3):735–742. e736.
32. Ganaiem H, Eisenstein EM, Tenenbaum A, et al. The role of hematopoietic stem cell transplantation in SP110 associated veno-occlusive disease with immunodeficiency syndrome. *Pediatr Allergy Immunol*. 2013;24(3):250–256.
33. Bosticardo M, Pala F, Calzoni E, et al. Artificial thymic organoids represent a reliable tool to study T-cell differentiation in patients with severe T-cell lymphopenia. *Blood Adv*. 2020;4(12):2611–2616.
34. McDonald-McGinn DM, Sullivan KE, Marino B, et al. 22q11.2 deletion syndrome. *Nat Rev Dis Primers*. 2015;1:15071.
35. Sullivan KE. Chromosome 22q11.2 deletion syndrome and DiGeorge syndrome. *Immunol Rev*. 2019;287(1):186–201.
36. Mehr S, Hsu P, Campbell D. Immunodeficiency in CHARGE syndrome. *Am J Med Genet C Semin Med Genet*. 2017;175(4):516–523.
37. Du Q, Huynh LK, Coskun F, et al. FOXN1 compound heterozygous mutations cause selective thymic hypoplasia in humans. *J Clin Invest*. 2019;129(11):4724–4738.
38. Yamazaki Y, Urrutia R, Franco LM, et al. PAX1 is essential for development and function of the human thymus. *Sci Immunol*. 2020;5(44).
39. Seidel MG, Kindle G, Gathmann B, et al. The European Society for Immunodeficiencies (ESID) Registry working definitions for the clinical diagnosis of inborn errors of immunity. *J Allergy Clin Immunol Pract*. 2019;7(6):1763–1770.
40. Grunebaum E, Cohen A, Roifman CM. Recent advances in understanding and managing adenosine deaminase and purine nucleoside phosphorylase deficiencies. *Curr Opin Allergy Clin Immunol*. 2013;13(6):630–638.
41. Martin E, Palmic N, Sanquer S, et al. CTP synthase 1 deficiency in humans reveals its central role in lymphocyte proliferation. *Nature*. 2014;510(7504):288–292.
42. Au-Yeung BB, Shah NH, Shen L, Weiss A. ZAP-70 in signaling, biology, and disease. *Annu Rev Immunol*. 2018;36:127–156.
43. Notarangelo LD. Functional T cell immunodeficiencies (with T cells present). *Annu Rev Immunol*. 2013;31:195–225.
44. Notarangelo LD. Combined immunodeficiencies with nonfunctional T lymphocytes. *Adv Immunol*. 2014;121:121–190.
45. Morgan NV, Goddard S, Cardno TS, et al. Mutation in the TCRalpha subunit constant gene (TRAC) leads to a human immunodeficiency disorder characterized by a lack of TCRalphabeta+ T cells. *J Clin Invest*. 2011;121(2):695–702.
46. Lu HY, Biggs CM, Blanchard-Rohner G, et al. Germline CBM-opathies: from immunodeficiency to atopy. *J Allergy Clin Immunol*. 2019;143(5):1661–1673.
47. Perkins ND. Integrating cell-signalling pathways with NF-kappaB and IKK function. *Nat Rev Mol Cell Biol*. 2007;8(1):49–62.
48. Doffinger R, Smahi A, Bessia C, et al. X-linked anhidrotic ectodermal dysplasia with immunodeficiency is caused by impaired NF-kappaB signaling. *Nat Genet*. 2001;27(3):277–285.
49. Zonana J, Elder ME, Schneider LC, et al. A novel X-linked disorder of immune deficiency and hypohidrotic ectodermal dysplasia is allelic to incontinentia pigmenti and due to mutations in IKK-gamma (NEMO). *Am J Hum Genet*. 2000;67(6):1555–1562.
50. Pannicke U, Baumann B, Fuchs S, et al. Deficiency of innate and acquired immunity caused by an IKBKB mutation. *N Engl J Med*. 2013;369(26):2504–2514.
51. Cardinez C, Miraghazadeh B, Tanita K, et al. Gain-of-function IKBKB mutation causes human combined immune deficiency. *J Exp Med*. 2018;215(11):2715–2724.
52. Courtois G, Smahi A, Reichenbach J, et al. A hypermorphic IkappaBalpha mutation is associated with autosomal dominant anhidrotic ectodermal dysplasia and T cell immunodeficiency. *J Clin Invest*. 2003;112(7):1108–1115.
53. Scott O, Roifman CM. NF-kappaB pathway and the Goldilocks principle: lessons from human disorders of immunity and inflammation. *J Allergy Clin Immunol*. 2019;143(5):1688–1701.
54. Kuehn HS, Niemela JE, Sreedhara K, et al. Novel nonsense gain-of-function NFKB2 mutations associated with a combined immunodeficiency phenotype. *Blood*. 2017;130(13):1553–1564.
55. Hanna S, Etzioni A. MHC class I and II deficiencies. *J Allergy Clin Immunol*. 2014;134(2):269–275.
56. de la Morena MT, Leonard D, Torgerson TR, et al. Long-term outcomes of 176 patients with X-linked hyper-IgM syndrome treated with or without hematopoietic cell transplantation. *J Allergy Clin Immunol*. 2017;139(4):1282–1292.
57. Warnatz K, Bossaller L, Salzer U, et al. Human ICOS deficiency abrogates the germinal center reaction and provides a monogenic model for common variable immunodeficiency. *Blood*. 2006;107(8):3045–3052.
58. Roussel L, Landekic M, Golizeh M, et al. Loss of human ICOSL results in combined immunodeficiency. *J Exp Med*. 2018;215(12):3151–3164.
59. Ghosh S, Kostel Bal S, Edwards ESJ, et al. Extended clinical and immunological phenotype and transplant outcome in CD27 and CD70 deficiency. *Blood*. 2020;136(23):2638–2655.
60. Slatter MA, Gennery AR. Update on dna-double strand break repair defects in combined primary immunodeficiency. *Curr Allergy Asthma Rep*. 2020;20(10):57.
61. Puck JM. Newborn screening for severe combined immunodeficiency and T-cell lymphopenia. *Immunol Rev*. 2019;287(1):241–252.
62. Wolska-Kusnierz B, Gennery AR. Hematopoietic stem cell transplantation for dna double strand breakage repair disorders. *Front Pediatr*. 2019;7:557.
63. Conde CD, Petronczki OY, Baris S, et al. Polymerase delta deficiency causes syndromic immunodeficiency with replicative stress. *J Clin Invest*. 2019;129(10):4194–4206.
64. Cui Y, Keles S, Charbonnier LM, et al. Combined immunodeficiency caused by a loss-of-function mutation in DNA polymerase delta 1. *J Allergy Clin Immunol*. 2020;145(1):391–401.
65. Sterlin D, Velasco G, Moshous D, et al. Genetic, cellular and clinical features of ICF syndrome: a French national survey. *J Clin Immunol*. 2016;36(2):149–159.
66. Candotti F. Clinical manifestations and pathophysiological mechanisms of the Wiskott-Aldrich syndrome. *J Clin Immunol*. 2018;38(1):13–27.
67. Lanzi G, Moratto D, Vairo D, et al. A novel primary human immunodeficiency due to deficiency in the WASP-interacting protein WIP. *J Exp Med*. 2012;209(1):29–34.
68. Volpi S, Cicalese MP, Tuijnenburg P, et al. A combined immunodeficiency with severe infections, inflammation, and allergy caused by ARPC1B deficiency. *J Allergy Clin Immunol*. 2019;143(6):2296–2299.
69. Su HC, Jing H, Angelus P, et al. Insights into immunity from clinical and basic science studies of DOCK8 immunodeficiency syndrome. *Immunol Rev*. 2019;287(1):9–19.

70. Aydin SE, Freeman AF, Al-Herz W, et al. Hematopoietic stem cell transplantation as treatment for patients with DOCK8 deficiency. *J Allergy Clin Immunol Pract*. 2019;7(3):848–855.
71. Pillay BA, Avery DT, Smart JM, et al. Hematopoietic stem cell transplant effectively rescues lymphocyte differentiation and function in DOCK8-deficient patients. *JCI Insight*. 2019:5.
72. Dobbs K, Dominguez Conde C, Zhang SY, et al. Inherited DOCK2 deficiency in patients with early-onset invasive infections. *N Engl J Med*. 2015;372(25):2409–2422.
73. Moens L, Gouwy M, Bosch B, et al. Human DOCK2 deficiency: report of a novel mutation and evidence for neutrophil dysfunction. *J Clin Immunol*. 2019;39(3):298–308.
74. Lagresle-Peyrou C, Luce S, Ouchani F, et al. X-linked primary immunodeficiency associated with hemizygous mutations in the moesin (MSN) gene. *J Allergy Clin Immunol*. 2016;138(6):1681–1689. e1688.
75. Delmonte OM, Biggs CM, Hayward A, et al. First case of X-linked moesin deficiency identified after newborn screening for SCID. *J Clin Immunol*. 2017;37(4):336–338.
76. Jabara HH, Boyden SE, Chou J, et al. A missense mutation in TFRC, encoding transferrin receptor 1, causes combined immunodeficiency. *Nat Genet*. 2016;48(1):74–78.
77. Calzoni E, Platt CD, Keles S, et al. F-BAR domain only protein 1 (FCHO1) deficiency is a novel cause of combined immune deficiency in human subjects. *J Allergy Clin Immunol*. 2019;143(6):2317–2321. e2312.
78. Vakkilainen S, Taskinen M, Makitie O. Immunodeficiency in cartilage-hair hypoplasia: pathogenesis, clinical course and management. *Scand J Immunol*. 2020;92(4):e12913.
79. Morimoto M, Lewis DB, Lucke T, et al. Schimke Immunoosseous Dysplasia Seattle (WA). In: Adam MP, Ardinger HH, Pagon RA, eds. *GeneReviews((R))*; 1993
80. Bergerson JRE, Freeman AF. An update on syndromes with a hyper-IgE phenotype. *Immunol Allergy Clin North Am*. 2019;39(1):49–61.
81. Volpi S, Yamazaki Y, Brauer PM, et al. EXTL3 mutations cause skeletal dysplasia, immune deficiency, and developmental delay. *J Exp Med*. 2017;214(3):623–637.
82. Cepika AM, Sato Y, Liu JM, et al. Tregopathies: monogenic diseases resulting in regulatory T-cell deficiency. *J Allergy Clin Immunol*. 2018;142(6):1679–1695.
83. Bezrodnik L, Gaillard MI, Caldirola MS. Dysregulatory syndromes: the role of signal transducers and activators of transcription. *Curr Opin Pediatr*. 2018;30(6):821–828.
84. Vaeth M, Feske S. Ion channelopathies of the immune system. *Curr Opin Immunol*. 2018;52:39–50.
85. Ravell JC, Chauvin SD, He T, et al. An update on XMEN disease. *J Clin Immunol*. 2020
86. Kotlarz D, Zietara N, Milner JD, et al. Human IL-21 and IL-21R deficiencies: two novel entities of primary immunodeficiency. *Curr Opin Pediatr*. 2014;26(6):704–712.
87. Boutboul D, Kuehn HS, Van de Wyngaert Z, et al. Dominant-negative IKZF1 mutations cause a T, B, and myeloid cell combined immunodeficiency. *J Clin Invest*. 2018;128(7):3071–3087.
88. Punwani D, Zhang Y, Yu J, et al. Multisystem anomalies in severe combined immunodeficiency with mutant BCL11B. *N Engl J Med*. 2016;375(22):2165–2176.
89. Schatorje EJ, Gemen EF, Driessen GJ, et al. Paediatric reference values for the peripheral T cell compartment. *Scand J Immunol*. 2012;75(4):436–444.

Genetic Disorders of Interferons-, Interleukin-17, Interleukin-18, and Nuclear Factor-κB-Mediated Immunity

Jacinta Bustamante, Shen-Ying Zhang, Bertrand Boisson, Vivien Béziat, Qian Zhang, Emmanuelle Jouanguy, Stéphanie Boisson-Dupuis, Anne Puel, and Jean-Laurent Casanova

In the last 25 years, primary immunodeficiencies (PIDs) or inborn errors of immunity (IEIs) affecting the immunity mediated by interferon (IFN)-γ, IFN-α/β-λ, Toll-like receptor (TLR) and interleukin (IL)-1 receptor (TIR) signaling, nuclear factor (NF)-κB pathway, TLR-3 pathway, IL-17, and IL-18 have been identified. Some of these genetic defects are "conventional" PIDs, associated with a broad range of infections, but others provide a molecular explanation for severe infectious diseases previously thought to be idiopathic (Table 35.1). These "non-conventional" PIDs may be associated with severe and/or recurrent infections caused by a single family of microorganisms, a situation strongly contrasting with that for "conventional" PIDs. Standard immunological explorations are generally normal in these patients, whether they are susceptible to one or several infectious agents.[1] Despite the lack of a clear immunological abnormality, infections in these patients are typically severe, and often fatal. Most IEIs affect leukocytes and other cell types.[2] This chapter is devoted to the description of these IEIs (see Table 35.1). They include disorders of the IFN-γ mediated immunity associated with the syndrome of Mendelian susceptibility to mycobacterial disease (MSMD) and tuberculosis (TB).[3–6] We also describe genetic defects affecting primarily the TLR-3 pathway. The predominant infectious phenotype of patients with these defects is herpes simplex virus (HSV)-1 encephalitis (HSE).[7] We include also the first genetic defects associated with severe isolated influenza, rhinovirus, human papillomaviruses (HPVs), fulminant hepatitis, and primary cytomegalovirus (CMV).[8–13] We describe IEIs associated with impaired signaling downstream from or via the TLR-IL-1R (TIR) pathway.[14–16] The main infectious phenotype of patients with any of these defects is the occurrence of pyogenic bacterial infections. Finally, we describe the group of defects affecting the IL-17A/F mediated immunity and conferring predisposition to chronic mucocutaneous candidiasis (CMC).[17,18]

KEY CONCEPTS

- New inborn errors of immunity (IEIs) should be considered in patients with unexplained serious infectious diseases.
- Children with severe infectious diseases should be actively investigated for a potential IEI.
- The exploration of idiopathic infections can lead to the discovery of new IEIs, which results in a better understanding of immunity to pathogens and new approaches to therapy.

GENETIC DISORDERS OF INTERFERON-γ-DEPENDENT IMMUNITY AND MENDELIAN SUSCEPTIBILITY TO MYCOBACTERIAL DISEASE

MSMD (OMIM#209950) is a rare group of genetic disorders associated with a selective susceptibility to weakly pathogenic mycobacteria, such as environmental mycobacteria (EM) and/or bacille Calmette–Guérin (BCG) vaccines, in otherwise healthy patients, normally resistant to other microbes (Fig. 35.1).[4] It occurs in about 1/50,000 individuals worldwide. The first clinical report of a probable idiopathic disseminated BCG (BCG-osis) infection following vaccination was in 1951. MSMD was the first IEI characterized by a selective predisposition to one or a few infectious agents. Various EM species have been isolated from MSMD patients.[3] The more virulent *Mycobacterium tuberculosis* has also been reported as disease-causing with these genetic disorders.[6] Patients with MSMD may have a wide range of clinical symptoms, from localized to persistent, disseminated infections and with impaired granuloma formation. Macrophage activation syndrome may occur in rare cases, probably because of uncontrolled infection.[4] Some patients spontaneously improve with age, or even remain clinically silent due to incomplete penetrance. About half of the patients are also particularly susceptible to non-typhoidal *Salmonella* and their spectrum of clinical diseases is broad, ranging from gastroenteritis to septicemia and disseminated infection.[3] A significant proportion of MSMD patients also suffer from CMC.[3] Occasionally other viral (caused by CMV, human herpes virus 8, parainfluenza virus type 3, respiratory syncytial virus [RSV] or varicella-zoster virus [VZV]), parasitic (leishmaniasis, toxoplasmosis), fungal (histoplasmosis, paracoccidioidomycosis, coccidioidomycosis) or bacterial (listeriosis, nocardiosis, klebsiellosis) infections have more rarely been reported.[3]

The first genetic etiology of MSMD was reported in 1996 with the discovery of autosomal recessive (AR) complete IFN-γR1 deficiency.[19,20] Since this initial description, mutations in 16 genes have been shown to be responsible for MSMD (Fig. 35.2).[4,21] The causal mutations in *IFNGR1* and *IFNGR2*, encoding the two chains of the receptor for IFN-γ, and some mutations in *STAT1*, encoding a transcription factor essential to the IFN-γR signaling pathway, result in defective cellular response to IFN-γ. MSMD-causing mutations in the genes encoding the p40 subunit of IL-12 *(IL12B)* and the β_1 chain of the receptor for IL-12 receptor *(IL12RB1)* affect the IL-12 and IL-23-dependent production of IFN-γ. MSMD-causing mutations in the genes encoding the β_2 chain of the IL-12 receptor *(IL12RB2)* affect the IL-12-dependent production by IFN-γ

TABLE 35.1 Genetic Disorders of Interferons-, Interleukin-17, Interleukin-18, and Nuclear Factor-κB-Mediated Immunity

Gene	Form	Inheritance	Mycobacteria	Salmonella	Viruses	HSE	Pyogenic Bacteria	Fungi	EDA	Inflammatory Signs
IFNRG1	Amorphic	AR	++	+	+	–	–	–	–	N
	Hypomorphic	AR	++	+	–	–	+	–	–	N
	Hypomorphic	AD	++	+	–	–	–	±–[a]	–	N
IFNRG2	Amorphic	AR	++	–	+	–	–	–	–	N
	Hypomorphic	AR	++	–	–	–	–	–	–	N
	Hypomorphic	AD	++	–	–	–	–	–	–	N
IFNG	Amorphic	AR	++	–	–	–	–	–	–	N
JAK1	Hypomorphic	AR	++	–	+	–	–	–	–	N
IRF8	Complete	AR	++	–	–	–	+	+	–	N
	Hypomorphic	AD	++	–	–	–	–	–	–	N
IL12B	Amorphic	AR	++	++	–	–	–	+/–	–	N
IL12RB1	Amorphic	AR	++	++	–	–	–	+/–[b]	–	N
IL12RB2	Amorphic	AR	++	–	–	–	–	–	–	N
IL23R	Amorphic	AR	++	–	–	–	–	–	–	N
STAT1	Amorphic	AR	++	–	++	++	–	–	–	N
	Hypomorphic	AR	++	+	+	–	–	–	–	N
	Hypomorphic	AD	++	–	–	–	–	–	–	N
	Hypermorphic	AD	–/+	–	–	–	–	++	–	N
ISG15	Amorphic	AR	++	–	–	–	–	–	–	N
TYK2	Amorphic	AR	++	–	VZV, HSV-1	–	–	–	–	N
	Hypomorphic	AR	++	–	–	–	–	–	–	N
CYBB	Hypomorphic	XR	++	–	–	–	–	–	–	N
TLR3	Amorphic	AR	–	–	HSV-1	++	–	–	–	N
	Hypomorphic	AR	–	–	HSV-1	++	–	–	–	N
	Hypomorphic	AD	–	–	HSV-1, IAV	++	–	–	–	N
TRIF	Amorphic	AR	–	–	HSV-1	++	–	–	–	N
	Hypomorphic	AD	–	–	HSV-1	++	–	–	–	N
UNC93B1	Amorphic	AR	–	–	HSV-1	++	–	–	–	N
TRAF3	Hypomorphic	AD	–	–	HSV-1	++	–	–	–	N
TBK1	Hypomorphic	AD	–	–	HSV-1	++	–	–	–	N
SNORA31	Hypomorphic	AD	–	–	HSV-1	++	–	–	–	N
DBR1	Hypomorphic	AR	–	–	HSV-1, IBV, norovirus	++	–	–	–	N
IRF3	Hypomorphic	AD			HSV-1	+	–	–	–	N
IRF7	Amorphic	AR	–	–	IAV	–	–	–	–	N
IRF9	Amorphic	AR	–	–	IAV	–	–	–	–	N
NOS2	Amorphic	AR	–	–	CMV	–	–	–	–	N
POLR3A[c]	Hypomorphic	AD	–	–	VZV	–	–	–	–	N
POLR3C[c]	Hypomorphic	AD	–	–	VZV	–	–	–	–	N
POLR3F	Hypomorphic	AD	–	–	VZV	–	–	–	–	N
IFIH1	Amorphic	AR	–	–	rhinovirus, RSV, EBV	–	–	–	–	N
IL18BP	Amorphic	AR	–	–	HAV	–	–	–	–	N
TMC6	Amorphic	AR	–	–	HPV	–	–	–	–	N
TMC8	Amorphic	AR	–	–	HPV	–	–	–	–	N
CIB1	Amorphic	AR	–	–	HPV	–	–	–	–	N
NEMO	Hypomorphic	XR	+	+	+	+	++	+	+/–	Weak
IKKB	Amorphic	AR	+	+	+	–	+	+	+/–	N
	Hypermorphic	AD	–	–	–	–	+	–	+/–	N
NFKBIA	Hypermorphic	AD	–	+	+	+	++	+	+	Weak
IRAK4	Amorphic	AR	–	–	–	–	++	–	–	Weak
MYD88	Amorphic	AR	–	–	–	–	++	–	–	Weak
HOIL-1	Amorphic	AR	–	–	+	–	++	–	–	Strong
HOIP	Hypomorphic	AR	–	–	+	–	++	–	–	Strong
IL17F	Hypomorphic	AD	–	–	–	–	–	++	–	N
IL17RA	Amorphic	AR	–	–	–	–	+/–	++	–	N
IL17RC	Amorphic	AR	–	–	–	–	–	++	–	N
ACT1	Amorphic	AR	–	–	–	–	+/–	++	–	N
MAPK8	Hypomorphic	AD	–	–	–	–	+/–	++	–	N
RORC	Amorphic	AR	++	–	–	–	–	+	–	N
STAT3	Hypomorphic	AD	–	–	–	–	++	++	–	Weak
ZNF341	Amorphic	AR	–	+	–	–	+	++	–	N
CARD9	Amorphic	AR	–	–	–	–	–	++	–	N

[a]One AD IFN-γR1-deficient patient presented one episode of *Histoplasma capsulatum* infection, another patient presented coccidioidomycosis.
[b]One IL-12Rβ1-deficient patient presented one episode of *Paracoccidioides brasiliensis* infection and some patients display forms of chronic mucocutaneous candidiasis.
[c]POLR3A and POLR3C can be monogenic or digenic deficiencies.
AD, autosomal dominant; *AR*, autosomal recessive; *CMV*, cytomegalovirus; *EDA*, anhidrotic ectodermal dysplasia; *HPV*, human papillomavirus; *HSE*, herpes simplex encephalitis; *IAV*, influenza A virus; *IBV*, influenza B virus; *N*, normal; *RSV*, respiratory syncytial virus; *VZV*, varicella-zoster virus; *XR*, X-linked recessive.

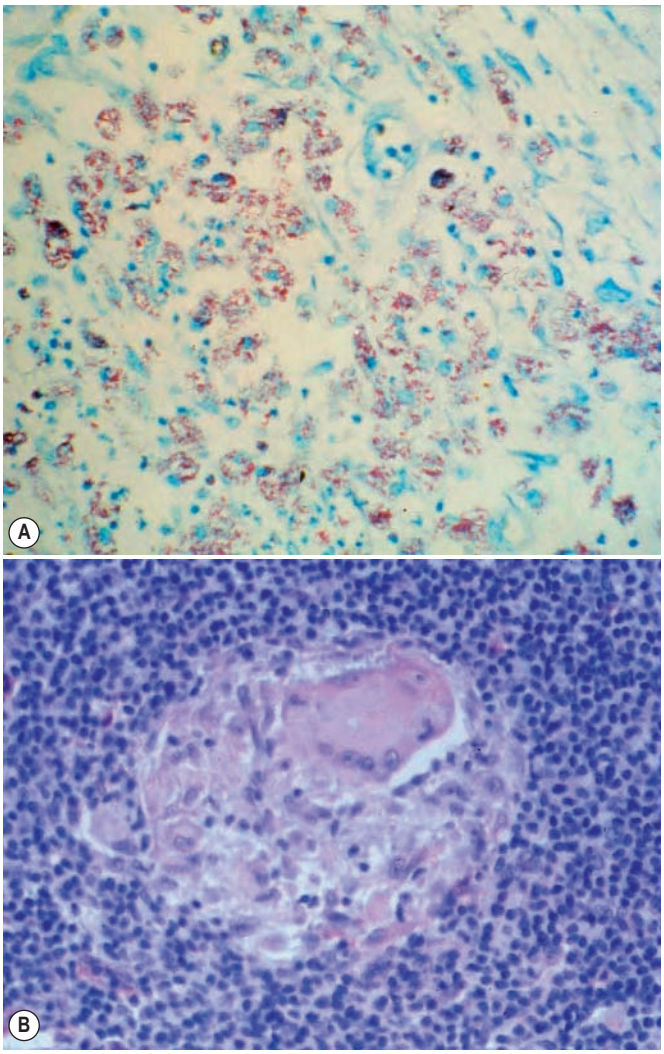


FIG. 35.1 Two Types of Granuloma in Patients With Mendelian Susceptibility to Mycobacterial Disease or Tuberculosis. (A) The lepromatous-like type consisted of poorly defined, poorly differentiated granulomas, with few, if any giant cells and lymphocytes, but widespread macrophages loaded with acid-fast bacilli. (B) The tuberculoid type consisted of well-circumscribed and differentiated granulomas, with epithelioid and multinucleated giant cells containing very few acid-fast rods, surrounded by lymphocytes and fibrosis, occasionally with central caseous necrosis.

and MSMD-causing mutations in the genes encoding the IL-23 receptor *(IL23R)* affect the IL-23-dependent production of IFN-γ. Bi-allelic mutations of *ISG15*, an IFN-γ-inducing molecule that acts in synergy with IL-12, have also been reported to cause MSMD, by impairing, but not abolishing IFN-γ production. Monoallelic mutations of *IRF8*, encoding an interferon regulator factor (IRF) inducible by IFN-γ, and homozygous mutations of *SPPL2A* impair IL-12 secretion by dendritic cells (DCs) interfering with the production of IFN-γ. Recently, a homozygous mutation in *IFNG* encoding IFN-γ has been described. MSMD can also be seen with hypomorphic defects in the X-linked gene encoding the NF-κB essential modulator (NEMO), which is involved in the CD40-dependent induction of IL-12, and in mutations affecting TYK2, which is involved in IL-12-dependent IFN-γ immunity. The MSMD-causing mutations in *CYBB*/gp91^{phox} affect the respiratory burst selectively in macrophages.[3,4] Some mutations of *STAT1* and most mutations of *NEMO* are associated with a broader range of infectious diseases (see below), and complete IRF8 deficiency is associated with a lack of circulating monocytes and DCs, as well as severe clinical disease, mimicking combined immunodeficiency (CID). The pathogenesis of MSMD in patients with these disorders results from impaired IFN-γ-mediated immunity. Among patients with MSMD, the level of IFN-γ mediated immunity determines the severity of mycobacterial disease.[4,21]

The high allelic heterogeneity at these 16 loci has led to the definition of 31 different genetic etiologies of MSMD, based on the mode of inheritance (recessive or dominant, autosomal or X-linked), mutant allele expression (e.g., presence or absence of encoded protein), the functional impact of the mutation (null or hypomorphic), and the mechanism underlying the dysfunction of expressed protein (e.g., phosphorylation, DNA-binding for transcription factors or both) (see Fig. 35.2).[3] In addition, mycobacteriosis has also been reported in patients with "syndromic MSMD" due to inherited AR TYK2, STAT1, or JAK1 deficiency, with susceptibility to mycobacterial and viral diseases due to impaired responses to IFN-γ (type II IFN) and IFN-α/β (type I IFN).[4] In addition, AR complete IRF8 deficiency, may present with mycobacterial, viral, and fungal infectious diseases due to impaired lymphoid and myeloid immunity; AR complete ROR-γ/ROR-γT deficiency, with susceptibility to mycobacterial and fungal diseases due to impaired IFN-γ and IL-17 immunities[22] and AR ISG15 associated with mycobacterial diseases plus type I interferonopathy. The identification of inborn errors of IL-12-, IL-23-, or ISG15-dependent induction of IFN-γ led to the successful, life-saving treatment of affected patients with recombinant IFN-γ.[3] Patients with MSMD are currently treated with antibiotics (ATB), but the prognosis for MSMD remains poor in some patients, while hematopoietic stem cell transplantation (HSCT) can be very difficult and is currently restricted to patients with complete deficiency of the IFN-γ receptor, STAT1, or IRF8.[3,4] Importantly, molecular diagnoses for siblings and offspring potentially carrying the MSMD-associated genotype may also lead to the selective prevention of mycobacterial disease with antimycobacterial antibiotics and/or recombinant IFN-γ[4] plus identification of individuals in whom vaccination with live BCG is absolutely contraindicated.[3]

THERAPEUTIC PRINCPLES

Treatment of Mendelian Susceptibility to Mycobacterial Disease Patients

- Vaccination with live bacille Calmette-Guerin (BCG) is contraindicated.
- Multiple antibiotics against the specific mycobacteria should be administered without interruption in patients with complete IFN-γR1 or IFN-γR2 deficiency.
- IFN-g therapy may be considered in addition to appropriate antimycobacterial antibiotics in patients with partial IFN-γR1, IFN-γR2, IRF8, gp91phox (or CYBB), or signal transducer and activator of transcription 1 (STAT1) deficiency as well as complete IL-12p40, IL-12Rβ1, or ISG15 deficiency.
- Hematopoietic stem cell transplantation (HSCT) should be considered in select patients with complete deficiency of IFN-γR1, IFN-γR2, STAT1, or IRF8.

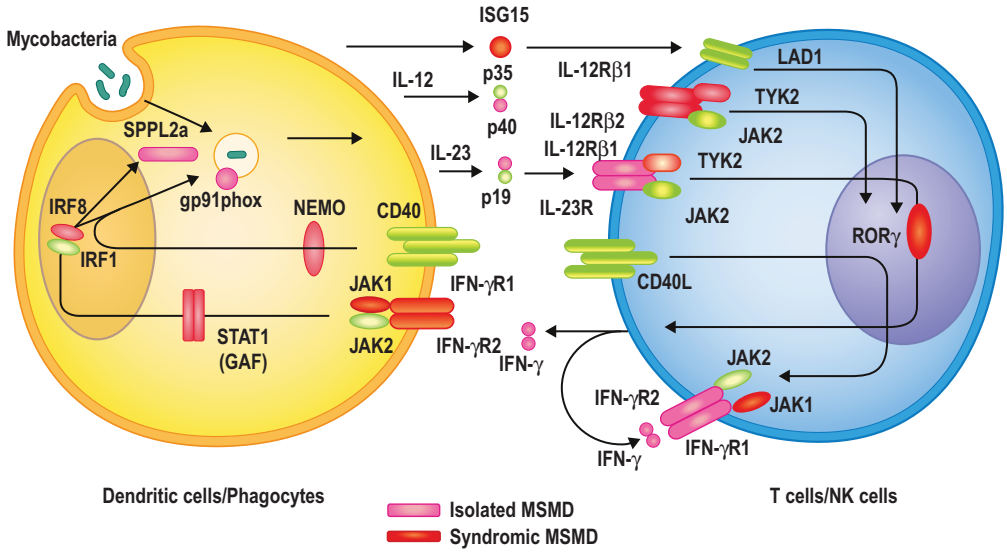


FIG. 35.2 Genetic Etiologies of Mendelian Susceptibility to Mycobacterial Infection. Cells involved in the production (T cells/NK cells) and response (dendritic cells, phagocytes) to interferon-γ (IFN-g). Proteins for which a mutation of the corresponding gene has been recognized to cause isolated Mendelian susceptibility to mycobacterial disease (MSMD) are shown in *pink* and those responsible for syndromic MSMD are shown in *red*. MSMD-causing mutations of *IL12B*, *IL12RB1*, *IL12RB2*, *IL23R*, *IFNG*, *ISG15*, *IRF8*, *SPPL2A*, *RORC*, *TYK2*, and *NEMO* impair the production of IFN-γ. MSMD-causing mutations of *IFNGR1*, *IFNGR2*, *STAT1*, *JAK1*, *CYBB*, and *IRF8* impair the action of IFN-γ.

GENETIC DISORDERS PREDISPOSING TO TUBERCULOSIS

TB caused by *M. tuberculosis* is still endemic in many countries (10 million new cases and 1.3 million deaths in 2017). However, only a minority of infected individuals (no more than 5% to 10%) develop clinical disease. Multiple experimental evidence has revealed a strong genetic basis for TB in humans.[6] However, classic association studies focusing on common variants of many candidate genes have not yielded consistent or reproducible results. Patient-based studies performed in parallel to these population-based studies yielded greater progress. Rare IEIs, including some of those linked to MSMD, were found to confer a predisposition to TB; these include IL-12Rβ1 and TYK2 deficiencies (found in no more than 1/600,000 individuals, hence representing a very small number of TB patients), and common IEI, such as homozygosity for P1104A TYK2 found in many more patients (this genotype is found in 1/600 Europeans and 1/5,000 other populations with the exception of sub-Saharan Africans and Eastern Asians).[23] Direct evidence includes the identification of siblings of MSMD probands carrying the same rare genetic defect but suffering from TB, with or without MSMD. Additional proof-of-concept was provided when several patients with TB but no family history of clinical MSMD were found to carry mutations in MSMD genes, including AR complete TYK2 deficiency found in two patients with isolated TB. These rare genetic etiologies underlying this common infectious disease provided proof-of-principle that TB susceptibility can be monogenic, but these situations are rare and do not account for TB as a global public health problem.[6]

Recently, a strong enrichment of individuals homozygous for a common TYK2 variant P1104A was found in a cohort of TB patients originating from countries outside Europe in which TB is currently endemic.[23] The minor allele frequency (MAF) of P1104A is 4.2% in the European population based on the Genome Aggregation Database (gnomAD), giving a frequency of homozygous individuals of approximately 1/600 in Europe. The P1104A mutant TYK2 protein catalytic activity is impaired such that it cannot phosphorylate itself or any other JAK or STAT substrate. Cells from patients homozygous for P1104A display impaired responses to IL-23, similar to cells from patients with complete TYK2 deficiency, whereas the other three TYK2-dependent cytokine pathways are unaffected. The role of P1104A homozygosity in TB was confirmed in a large cohort of European ancestry (United Kingdom [UK] Biobank).[6] Homozygosity for P1104A is the first common monogenic cause of TB susceptibility and the most common AR disorder in European population identified to date. It is estimated that TB has killed up to 1 billion people in Europe over the last 2000 years, suggesting that approximately 10 million people may have died due to TYK2 P1104A homozygosity. Interestingly, this homozygous variant was also found to have a strong protective effect against various inflammatory or autoimmune disorders, justifying the use of TYK2 inhibitors to treat some of these conditions.[6] Further understanding of genetic predisposition to TB is of clinical importance for new approaches to prevention and alternative approaches to therapy of the disease (e.g., recombinant IFN-γ).

GENETIC DISORDERS OF THE TLR3 PATHWAY IN HERPES SIMPLEX VIRUS-1 ENCEPHALITIS

This group of genetic disorders leads to impaired TLR3 signaling and susceptibility to HSE in childhood (Fig. 35.3).[7] The affected

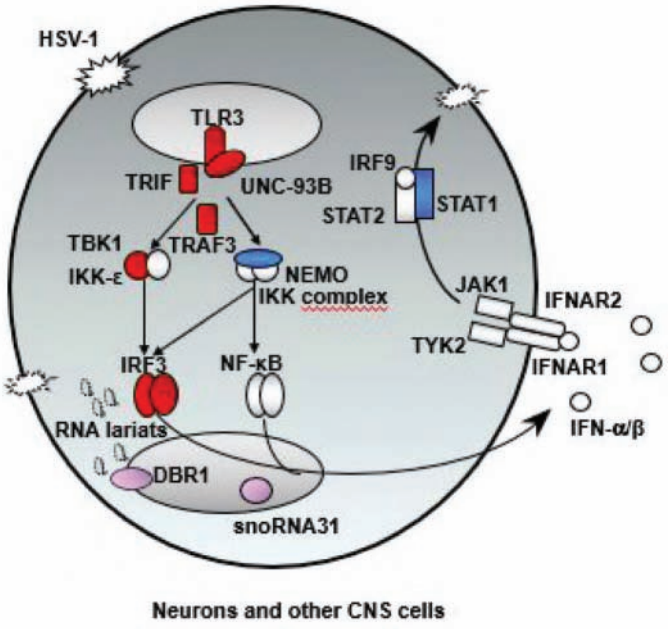


FIG. 35.3 Genetic Etiologies Underlying Herpes Simplex Encephalitis and Other Viral Encephalitis. Viruses enter the cells *(neurons, and other CNS cells)* and viral dsRNa is detected by TLR3. This recognition induces activation of IRF3 and nuclear factor-κB (NF-κB) pathway, leading the production of type I and type III interferons (IFNs). The binding of IFNs to their receptor induces the phosphorylation of JAK1 and TYK2 activating STAT1, STAT2, and IRF9. This complex is translocated as a heterotrimer to IFN-inducible genes. Proteins for which genetic mutations have been identified with susceptibility to herpes simplex virus (HSV)-1 encephalitis (HSE) andTLR3 signaling pathways are shown in *red*. Mutations in other genes of TLR3-IFNs pathway are shown in *blue*. Impaired TLR3 dependent, IFN-mediated cortical neuron and oligodendrocyte-autonomous anti-HSV-1 immunity underlie the pathogenesis of HSE in SnoRNA31 and DBR1 deficiencies (these deficiencies are shown in *pink*).

patients carry mutations of *TLR3*, *UNC93B1*, *TRIF*, *TRAF3*, *TBK1*, or *IRF3*. UNC-93B is a 12-transmembrane domain protein present in the endoplasmic reticulum (ER) that delivers the nucleotide-sensing receptors TLR3, 7, 8, and 9 from the ER to the endolysosomes. The TLR3 signaling pathway is exclusively mediated by the TRIF adapter and leads to activation of the transcription factors IRF3 and NF-κB. TRIF recruits TRAF6 and activates TAK1 in order for NF-κB to be activated. TRIF also recruits a signaling complex involving TBK1 and IKKε via TRAF3 for IRF3 activation.[7] This signaling pathway induces the production of type I and III IFNs (-α, -β, and -λ) as well as inflammatory cytokines that are important for antiviral immunity. TRAF3, TBK1, and IRF3 have functions downstream from multiple TNF family receptors as well as the receptors inducing IFN-α, -β, and -λ production, including TLR3. Genetic studies of isolated HSE of the forebrain led to the discovery of single-gene inborn errors of the TLR3-dependent pathway inducing IFN-α/β and -λ. These included mono- or biallelic mutations of the six TLR3 pathway genes (see Fig. 35.3).[7]

These findings, together with the previous observation of syndromic HSE in patients with X-linked recessive (XR) NEMO deficiency or AR complete STAT1 deficiency, suggested that TLR3-dependent IFN-α/β and/or -λ immunity is crucial for host defense against HSV-1 in the central nervous system (CNS). It has also been suggested that other mutations of these and other TLR3 or IFN pathway genes may underlie HSE in children or adults.[7] Interestingly, a few other patients with mutations of *TLR3* developed influenza A virus (IAV) pneumonitis or VZV ophthalmicus.[7] Importantly, TLR3-mediated responses to dsRNA and antiviral immunity seem to be redundant in most TLR3-expressing (non-CNS) cell types, including leukocytes in particular, which likely accounts for the lack of viral dissemination during the course of HSE.

The hypothesis that CNS-specific cell-intrinsic immunity rather than leukocyte-mediated immunity is crucial for host defense against HSV-1 was tested experimentally, initially with dermal fibroblasts as surrogate cells, and then with induced pluripotent stem cell (iPSC)-derived CNS- and peripheral nervous system (PNS)-resident cells from patients with forebrain HSE and mutations affecting TLR3 pathway genes. This work was consistent with studies showing that murine Tlr3 is required for responses to HSV in neurons and astrocytes. TLR3 pathway-deficient human fibroblasts[7,24,25] and iPSC-derived cortical neurons and oligodendrocytes[26] were found to be much more susceptible to HSV-1 infection than control cells. This phenotype was rescued by the addition of exogenous IFN-α/β. By contrast, in vitro differentiated human UNC-93B-deficient astrocytes or neural stem cells as well as TLR3-deficient peripheral trigeminal ganglia (TG) neurons did not show increased HSV-1 susceptibility compared to that of control cells.[26] Microglial cells, the CNS-resident macrophages that have been shown to rely on the cGAS/STING pathway to orchestrate anti-HSV-1 defense in mice, have not yet been tested in the human response to HSV-1 infection. Taken together, these data suggest that TLR3-dependent, IFN-mediated cortical neuron- and oligodendrocyte-autonomous anti-HSV-1 immunity appears to be crucial for host defense against HSV-1 infection in the human forebrain. These data provided a plausible cellular basis for the pathogenesis of genetically driven forebrain HSE, suggesting that cell-intrinsic immunity was crucial for host defense against HSV-1 in the human forebrain, as opposed to the innate and adaptive immunity mediated by leukocytes and related cells.[24,25] Treatment with recombinant IFN–α, in parallel with acyclovir, may

THERAPEUTIC PRINCPLES

Treatment of Patients With Genetic Susceptibility to Herpes Simplex Virus-1 Encephalitis

- In patients with TLR3 pathway deficiencies consideration can be given for adding interferon (IFN)-α to acyclovir therapy with the possibility of improving outcome.
- Serological monitoring for herpes simplex virus-1 (HSV-1) infection should be considered in individuals carrying mutations in TLR3 pathway genes with no detectable serum anti-HSV-1 antibody.
- In the absence of an effective vaccine against HSV-1, acyclovir may be considered an appropriate prophylactic treatment in individuals carrying mutations of TLR3 pathway genes, even if serologically negative for HSV-1.

help to improve disease outcomes in patients with TLR3 pathway deficiencies and HSE. With the exception of AR complete TLR3 deficiency and AR complete TRIF deficiency that have only been reported in a single patient, the other TLR3 pathway deficiencies displayed complete penetrance at the cellular level, but incomplete penetrance at the clinical level.

GENETIC DISORDERS OF OTHER INTERFERON-INDUCING PATHWAYS IN OTHER SEVERE VIRAL INFECTIONS

IEI affecting other virus-sensing, IFN-α-β-λ-inducing pathways have been reported recently, namely POL III deficiency underlying severe VZV infection, and MDA5 deficiency underlying severe pulmonary infections caused by various viruses including rhinovirus and RSV.[7,27] POL III is a protein complex comprising 17 subunits organized into different subcomplexes. It is a cytosolic DNA sensor recognizing and transcribing AT-rich DNA to RNA, and then triggering IFN induction through the RIG-I pathway. Four children with VZV pneumonitis or CNS infection have been found to have rare heterozygous missense Mutations of *POLR3A*, *POLR3C*, or both. Leukocytes from all patients displayed poor induction of IFN-α/β and -λ in response to synthetic or VZV-derived AT-rich DNA, or upon VZV infection resulting in poor control of VZV replication.[28] This work suggested an important contribution of type I and III IFNs to host defenses against VZV. Heterozygous *POLR3F* mutations were subsequently reported in monozygotic adult twins experiencing repeated CNS vasculitis presenting in a stroke-like manner with hemiparesis, sensory deficits, and headache, clinically diagnosed as being caused by recurrent VZV reactivation.[29] MDA5 is an IFN-α/β- and -λ-inducing cytosolic sensor of double-stranded (ds)RNA associated with viral byproducts and intermediates. Biallelic or monoallelic mutations of *IFIH1* encoding MDA5 have been reported in children with pulmonary rhinovirus, RSV, or EBV infection.[27] The *IFIH1* mutations are associated with impaired (ds)RNA sensing, reduced type I IFN induction, and increased cellular susceptibility to rhinovirus and RSV. Thus, human MDA5 deficiency is a novel inborn error of innate and/or intrinsic immunity that predisposes to respiratory viral infections. These new findings further highlighted the non-redundant role of type I and III IFNs in antiviral immunity in humans.

GENETIC ETIOLOGIES OF LIFE-THREATENING INFLUENZA PNEUMONITIS

Influenza viruses, including IAV and influenza B virus (IBV), have caused upper respiratory tract diseases throughout human history. Typically influenza virus infection causes a relatively mild disease of the upper respiratory tract that is readily cleared with little need for medical intervention.[27] However, infection with seasonal or more virulent pandemic influenza strains can cause life-threatening or fatal disease (e.g. acute respiratory distress syndrome [ARDS]). Four IEIs—GATA2, IRF7, IRF9, and TLR3 deficiencies—have been identified to be associated with life-threatening influenza pneumonitis (see Table 35.1).[8,27] Autosomal dominant (AD) GATA2 deficiency is the only one of these IEIs leading to a pleiotropic syndromic disorder that manifests as a lack of lymphoid and myeloid progenitors in the bone marrow, small numbers of DCs, monocytes, T, B, and NK cells, together with high susceptibility to viral, mycobacterial, and fungal infections. The deaths of four adults from severe IAV infection despite the presence of neutralizing antibodies in the serum have been reported. The lack of plasmacytoid DCs (pDCs) may contribute to susceptibility to IAV, because pDCs are known to produce a massive amount of type I IFN in response to viral infections. Type II IFN-mediated hypercytokinemia during infection was observed in two of the four patients, suggesting that immune dysregulation may also contribute to the severity of pneumonitis.[27]

Unlike patients with AD GATA2-deficiency, those with AR IRF7 or AR IRF9 deficiency have an essentially isolated susceptibility to influenza (see Table 35.1).[8,9,27,30] No overt immunological abnormality has been detected in these patients except for two recently reported IRF9-deficient siblings born to a consanguineous family.[27] Defects of IRF7 or IRF9 interrupt both type I and type III IFN signaling. IRF7 deficiency hinders the early production of type I IFNs by pDC. In contrast, IRF9 deficiency prevents IFN-stimulated gene factor 3 (ISGF3, a complex of STAT1, STAT2, and IRF9) formation that blocks downstream type I and III IFN responses. The cells of IRF7- and IRF9-deficient patients are susceptible to IAV infections in vitro, and to laboratory strains of other viruses.[27] Most recently, three unrelated, otherwise healthy, patients who carry heterozygous *TLR3* mutations have been reported to developed influenza pneumonitis but not HSE (see Table 35.1).[30] Two of these patients carried the same mutation as previously detected in four HSE patients.[30] Defects of TLR3 impair both type I and type III responses to IAV that can be rescued in vitro by exogenous type I and/or III IFNs, depending on the distribution of IFN receptors on the infected cell types. For example, either type I or type III IFNs can rescue defective TLR3 signaling in IAV-infected pulmonary epithelial cells (PECs) because both receptors are strongly expressed.[30] Observations in cases of TLR3 deficiency suggest that the tissue distribution of type I or type III IFN signaling molecules can determine the susceptibility to IAV. To date, no patient suffering from both HSE and IAV ARDS has been reported, indicating incomplete penetrance in both diseases' susceptibility. Recently, one adult Iranian patient has been reported with a CMV pneumonitis carrying a homozygous mutation in nitric oxidase synthase 2 (NOS2).[10] The mechanism connecting CMV infection and NOS2 deficiency has not yet been identified and requires further investigation.

GENETIC ETIOLOGIES IN HUMAN PAPILLOMAVIRUSES

HPVs are ubiquitous small DNA viruses with strict tropism for cutaneous or mucosal stratified epithelia. HPVs cause common, plantar, and flat warts on the skin, multifocal epithelial hyperplasia (MEH), mucosal condyloma, as well as cervical and oropharyngeal lesions. Persistent infections can lead to benign tumors and, in some cases, malignant transformation and progression to invasive cancer. Severe HPV infections are frequent in patients with HIV infection or on immunosuppressive treatment, indicating an important role for adaptive T-cell immunity in the control of HPV infection. In contrast to these syndromic HPV susceptibilities, some patients have isolated severe HPV infection

with milder T-cell deficiencies resulting mostly in decreased numbers or partially impaired function of CD4 T cells, as seen in patients with complete CD4 or IL-7 deficiency.[31–33] Among isolated susceptibilities to HPV infection, epidermodysplasia verruciformis is the best studied. The patients are specifically susceptible to β-HPVs that cause flat warts. AR amorphic mutations of *TMC6*, *TMC8*, and *CIB1*, encoding EVER1, EVER2, and calcium and integrin binding protein1 (CIB1), respectively, have been reported in these patients.[11,12,34] The EVER and CIB1 proteins have been shown to form a complex, and are thought to be viral restriction factors in skin keratinocytes.[12] Recently, juvenile recurrent respiratory papillomatosis, driven by certain α-HPVs, was shown to be due to biallelic gain of function (GOF) in *NLRP1*, suggesting an important role of cell intrinsic immunity against HPVs other than β-HPVs (α, γ, μ and ν).[11,34] Altogether, the studies of inherited susceptibility to HPVs reveal a central role of CD4 T cells and of keratinocyte intrinsic immunity for the control of the infection.

HUMAN GENETIC BASIS OF FULMINANT VIRAL HEPATITIS

Fulminant viral hepatitis (FVH) is a severe condition of liver function characterized by massive hepatocyte necrosis and an inflammatory infiltrate.[13] FVH is life-threatening, but the patients are typically otherwise healthy, and normally resistant to other microbes. Hepatitis A virus (HAV) and hepatitis B virus (HBV) are frequently involved in FVH. The actual prevalence and incidence of FVH worldwide are not precisely known, but previous studies have suggested that FVH develops in no more than 0.5% and 0.1% of individuals with symptomatic HAV and HBV infections, respectively. The outcome is poor, with fewer than 20% of patients surviving in the absence of liver transplantation. FVH is typically sporadic, but rare familial forms also exist, suggesting that there may be causal monogenic IEI. AR complete IL-18 binding protein (IL-18BP) deficiency was identified in 2019 as the first genetic etiology of FVH following HAV infection.[35] A girl born to a consanguineous family from Algeria but living in France died of FVH. The patient carried a homozygous private mutation in *IL18BP* gene, which encodes the IL-18B.[35] IL-18BP has a high affinity for IL-18 and blocks the binding of this cytokine to its membrane-bound receptor, IL-18R. It is known that IL-18 is hepatotoxic and was initially identified as the cytokine responsible for liver failure in a murine model. The mechanism is thought to involve enhanced IL-18- and IFN-γ-dependent killing of hepatocytes mediated by NK and CD8 T cytotoxic cells. Moreover, the FVH-causing *IL18BP* genotype suggests that excessive IL-18 immunity may be a general mechanism underlying FVH, perhaps through the enhancement of IFN-γ immunity. The elucidation of the genetic basis of FVH has important clinical implications (given the high mortality) for patients and their families, with the possibility of genetic diagnoses and counseling plus enabling the development of treatments based on the immunological mechanism involved.[13]

HYPOMORPHIC MUTATIONS OF *IKBKG*/NEMO

XR anhidrotic ectodermal dysplasia with immunodeficiency (XR-EDA-ID) is a rare IEI associated with a developmental disorder (Fig. 35.4) (OMIM# 300291). Patients with XR-EDA-ID carry hypomorphic mutations of *IKBKG*, which encodes IKK-γ/NEMO, a protein essential for NF-κB activation. More than 100 patients with such mutations have been reported.[36] The only immunological abnormality known to be common to all patients with NEMO deficiency is low levels of memory B cells, while an absence of serum antibodies generated to carbohydrate antigens has also been reported in the majority of patients. Some patients have high IgM levels, and a small number have NK-cell abnormalities.[36] The infectious phenotype of patients with *IKBKG* mutations patients is characterized by infections caused by encapsulated pyogenic bacteria, such as *Haemophilus influenzae* and *Streptococcus pneumoniae*. Infections caused by weakly pathogenic microorganisms, such as *M. avium* and *M. kansasii*, as well as other viral and bacterial infections have also been reported.[36] Infections are marked by a poor or delayed clinical and biological inflammatory response. Some patients (~20%) present with recurrent diarrhea and/or colitis. About 80% of the patients with *NEMO* mutations described have ectodermal dysplasia (EDA), which is characterized by hypohidrosis, widely spaced cone-shaped or peg-shaped teeth, and hypotrichosis (see Fig. 35.4).[36] These features result from defective signaling downstream from the ectodysplasin receptor. A small number of patients with more severe *IKBKG* mutations have been reported to display osteopetrosis and lymphedema associated with the EDA phenotype.[36] Some patients also have dysmorphia with mild frontal bossing. However, about 10% of patients with *IKBKG* mutations display none of the classic features of the EDA phenotype.[36,37]

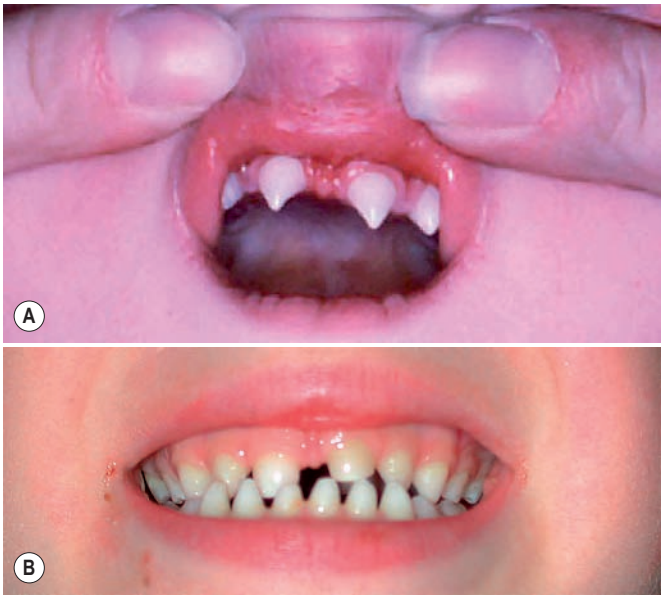


FIG. 35.4 Ectodermal Dysplasia With Immunodeficiency Patients. Two patients with ectodermal dysplasia with immunodeficiency, one with widely spaced cone- or peg-shaped teeth, the other having conical incisors.

GENETIC DISORDERS OF IKKα/IKKβ

βAR complete IKKβ deficiency (OMIM#603258) was first reported in 2013, in four infants.[38] All patients presented with early-onset, life-threatening bacterial, mycobacterial, fungal, and viral infections as well as failure to thrive, conditions consistent with a clinical diagnosis of severe combined immunodeficiency (SCID).[38] Other patients with biallelic loss-of-function (LOF) mutations of *IKBKB* and a similar infectious phenotype have since been reported.[16] Immunological and functional investigations revealed abnormal T-cell differentiation (excess of naïve T cells), very low levels or absent regulatory T cells (Tregs), and low or absence of γ/δ T cells. The B-cell differentiation is also impacted with low levels of class-switched memory B cells.[16] IKKβ-deficient patients display no signs of EDA, developmental or gene regulatory defects in other organs, except for two patients with conical teeth.[16] A heterozygous GOF mutation of *IKBKB* has also been reported.[39] The patients presented with recurrent otitis media and sinusitis. EDA was considered possible only in the proband of one kindred. These patients have immune dysregulation, combined T-cell and B-cell deficiency, with low levels of naïve T cells and memory B cells associated with hypogammaglobulinemia.[39] Finally, AR complete IKKα deficiency is very rare and, like NEMO deficiency, seems to lead to a spectrum of phenotypes extending from fetal death to severe EDA-ID-like phenotypes in newborns. The presence of a heterozygous stop-gain mutation in a public database and the absence of a phenotype in the heterozygous parents in the initial report of the AR condition excluded the possibility of dominance by haploinsufficiency.

HYPERMORPHIC MUTATIONS OF NUCLEAR FACTOR-κB INHIBITOR ALPHA (NFKBIA/IκBα)

Heterozygosity for a hypermorphic or GOF *NFKBIA* mutation (OMIM#164008) encoding IκBα was first reported in 2003, in a European patient.[40] Other unrelated patients with seven different hypermorphic mutations of *NFKBIA* have since been reported.[40] IκBα GOF mutations lead to an impairment of T-cell receptor (TCR) signaling in 70% of patients, while all patients described with these mutations have impaired TNFR and IL-1R/TLR responses.[40] Patients with *NFKBIA* GOF mutations have low levels of memory B cells, dysgammaglobulinemia, and low levels of specific antibodies. Some also have low proportions of memory CD4 or CD8 T cells, or both, an excess of naïve T cells, or an absence of TCRγ/δ T cells. With the exception of two patients with IκBα disorder, one with complex mosaicism and one with S36Y mutation, all patients have been found to have the features of EDA.[36,40] Patients with IκBα GOF mutations have developed recurrent bacterial infections: pneumonia, sepsis or meningitis, and arthritis. They are also prone to opportunistic infections, such as *Pneumocystis jiroveci* pneumonia and CMC. Finally, some patients have presented with recurrent diarrhea and/or colitis.[16]

LUBAC DEFICIENCY: HOIL1/RBCK1 AND HOIP/RNF31 DEFICIENCIES

AR complete HOIL-1 deficiency (HOIL1/*RBCK1*, OMIM#610924) due to biallelic mutations of the HOIL-1 gene was first reported in 2012.[41] Patients had an autoinflammatory syndrome with pyogenic bacterial diseases, which led to their death.[41] LUBAC deficiency results in an impaired response to TNF and IL-1β in fibroblasts as well as an impaired response to CD40L in B cells, but it also leads to hyperresponsiveness to IL-1β in monocytes. Patients presented with recurrent systemic inflammatory symptoms starting during the first few months of life due to monocyte overactivation. They also developed recurrent pyogenic bacterial infections caused by *S. pneumoniae, H. influenzae, Escherichia coli, Staphylococcus* spp., and *Enterococcus*, at least in part because of their inability to produce anti-polysaccharide antibodies. One patient also had chronic CMV infection and another had *Giardia intestinalis* infection.[41] All patients also had amyotrophy, muscle weakness, and failure to thrive, probably due to muscular amylopectinosis complicated by peripheral myopathy and cardiomyopathy.[16] Interestingly, 14 patients with HOIL1 deficiency from 10 unrelated kindreds were identified on the basis of neuromuscular and cardiac involvement secondary to amylopectinosis. Partial AR HOIP deficiency (RNF31, OMIM# 612487) has been identified in two patients. The first patient displayed an autoinflammatory syndrome, pyogenic bacterial diseases, amylopectinosis, and lymphangiectasia.[42] There are reports of a second patient diagnosed with common variable immune deficiency (CVID) who displayed autoinflammation in the absence of lymphangiectasia or amylopectinosis. At the cellular level, HOIP deficiency is similar to AR HOIL-1 deficiency (see above), presenting with recurrent episodes of amyotrophy, muscle weakness, and failure to thrive, partly because of muscle amylopectinosis. These two cases of HOIP deficiency displayed similar clinical features of immune dysregulation.[16] Both suffered from severe bacterial infections due to the low level of memory B cells and impaired antibody production.

IRAK4 AND MYD88 DEFICIENCY

AR complete IRAK-4 deficiency (OMIM#607676) was first described in 2003.[14] These patients have normal basic immunophenotyping results, normal antigen-specific T-cell and B-cell responses with two notable exceptions.[16] First, the glycan-specific IgG and IgM antibody responses to pneumococcal and AB glycans (allohemagglutinins of the ABO system) are impaired in up to one third of the cases explored.[43] Second, serum IgE and IgG4 concentrations are high in up to two thirds and one third of the patients, respectively.[43] The IRAK-4-deficient patients were found to have a specific deficiency of unswitched memory B cells ($CD19^+$, IgD^+, $CD27^+$) but normal levels of switched memory B cells. IRAK-4 deficiency confers a selective predisposition to severe bacterial infection (*S. pneumoniae*, *S. aureus*, and *Pseudomonas aeruginosa*), with an impairment of the ability to increase plasma C-reactive protein (CRP) concentrations and to mount a fever at the beginning of an infection. Patients with IRAK-4 deficiency suffered from their first bacterial infection before the age of 2 years. However, the clinical phenotype of IRAK-4 deficiency seems to improve with age, as none of the patients had invasive bacterial infection after the onset of adolescence.[43] However, patients with IRAK-4 deficiency suffer from skin and respiratory infections even during and after adolescence.[43] AR complete MyD88 deficiency has been described in more than 20 patients (OMIM#612260).[15,43] The clinical and immunological features of these patients are similar to those with AR complete IRAK-4 deficiency.[15,16,43]

THERAPEUTIC PRINCPLES

Principles for the Treatment of Inherited Disorders of TIR-Mediated Immunity

- Patients should receive conjugated and nonconjugated vaccines against encapsulated bacteria (*S. pneumoniae*, *H. influenzae*, *N. meningitidis*). All live vaccines (BCG, poliovirus, MMR) are contraindicated.
- A preventive treatment, including antibiotic prophylaxis with trimethoprim-sulfamethoxazole and/or penicillin V, should be administered throughout the life of the patient.
- Monthly prophylactic administrations of intravenous or subcutaneous immunoglobulins should be considered in selected patients.
- Empiric parenteral antibiotic treatment against *S. pneumoniae*, *S. aureus*, and *P. aeruginosa* should be initiated as soon as an infection is suspected or if the patient develops a moderate fever, without taking inflammatory parameters into account.
- Hematopoietic stem cell transplantation should be considered in selected patients with genetic disorders of NEMO, IκBα, and complete deficiency of IKBKB.
- Patients with HOIL-1 and HOIP deficiencies should be followed for possible muscle weakness and cardiomyopathy. Heart transplants may need to be considered for these patients.

TIRAP Deficiency

AR complete TIRAP deficiency (OMIM#606252) has been identified in eight individuals of a large consanguineous family.[44] The homozygous mutation (R121W) of the Toll-IL-1 receptor domain-containing adapter protein (TIRAP) identified in this family affects the TIR domain and impairs the binding of TIRAP to both the receptor and MyD88. All but one of the individuals had no clinical phenotype and this patient suffered from pneumonia and sepsis caused by Panton-Valentine leukocidin (PVL)-producing *S. aureus*. Specific deficiency of unswitched memory B cells ($CD19^+$, IgD^+, $CD27^+$) but normal levels of switched memory B cells were observed.[44] Antibody production was normal in all carriers of the mutation, except for an absence of antibodies directed against lipoteichoic acid (anti-LTA) in the patient with clinical manifestations. In humans, the leukocyte response to LTA is mediated principally by TLR2/6 heterodimers acting in concert with CD36, but anti-LTA antibodies enhance this response through the recognition of their invariant IgG domain by CD32.[44]

Genetic Disorders of Interleukin-17-Mediated Immunity and Chronic Mucocutaneous Candidiasis

CMC is characterized by persistent or recurrent infections of the skin (intertrigo, angular cheilitis), mucous membranes (oral, esophageal, genital), and nails (onychomycosis) caused by *Candida* (*C. albicans* in particular).[18] CMC was first described in the 1960s, and was reported to display AD heritance. Soon after, cases with a possible AR inheritance were reported. Inherited CMC usually begins early in infancy and can affect either otherwise healthy individuals (isolated CMC) or be associated with other clinical features (syndromic CMC). Superficial dermatophytosis, invasive fungal diseases, bacterial infections of the skin and respiratory tract, and viral cutaneous infections have been reported in some patients. In rare cases, cerebral aneurysms, oral/esophageal squamous cell carcinoma, or autoimmunity have also been described. These IEIs are caused by mutations of *IL17F*, *IL17RA*, *IL17RC*, *ACT1/TRAF3IP2*, *MAPK8*, *CARD9*, *STAT3*, *ZNF341*, *RORC*, *STAT1*, and *AIRE*, representing genes that disrupt the production of or the response to IL-17A and IL-17F (Fig. 35.5).[18] Some patients with invasive fungal diseases (e.g., CNS candidiasis) carry biallelic mutations of the *CARD9* gene.[18] The molecular and clinical features of CMC have been reviewed elsewhere.[18] An exception to these IEIs is the AR autoimmune polyendocrinopathy syndrome type 1 (APS-1 or autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy [APECED]) (OMIM 240300). APS-1 is caused by mutations of the *AIRE* gene resulting in impaired T-cell tolerance, with 88% of patients developing CMC. High levels of neutralizing autoantibodies against IL-17A, IL-17F, and/or IL-22 have been detected in the serum of many APS-1 patients possibly responsible for the CMC.[45]

AR complete IL-17RA deficiency (OMIM 613953) was first identified in 2011 and additional patients were reported later (see Fig. 35.5).[17,18] CMC was present in all patients, but some may also have developed staphylococcal cutaneous lesions and recurrent bacterial infections of respiratory tract.[18] The cellular phenotype was characterized by an absence of response to IL-17A and IL-17F homo- and heterodimers in the patients' fibroblasts, and to IL-17E (IL-25) in the patients' peripheral blood mononuclear cells.[18] AD IL-17F deficiency (OMIM 613956) was also first described in 2011 in a multiplex kindred from Argentina with isolated CMC.[17] A monoallelic private missense mutation of *IL17F* was detected and found to have no impact on protein production. However, this mutation greatly decreased the activity of homo- and heterodimers (IL-17F/IL-17F or IL-17A/IL-17F) containing the mutant protein, by affecting their binding to the receptor. Another patient with a monoallelic mutation of *IL-17F* has been reported, but no cellular characterization was carried out.[17] AR complete IL-17RC deficiency has been reported in three unrelated patients from Turkey and Argentina.[46] All patients suffered from CMC, and none presented invasive or recurrent bacterial infections. The clinical manifestations of infectious diseases in this group of patients resembled those in patients with AD IL-17 deficiency.[46] The homozygous mutations identified confer a loss of IL-17RC protein expression in transfected HEK293T cells, with complete abolition of the response to IL-17A and IL-17 homo- and heterodimers of patients' fibroblasts.[46] AR complete ACT1 deficiency has been reported in two patients from a consanguineous family from Algeria.[47] These two siblings developed CMC. However, one of them presented recurrent episodes of folliculitis decalvans and bilateral blepharitis caused by *S. aureus*. Their cellular phenotypes were characterized by impaired responses to IL-17A and IL-17F homo- and heterodimers in patients' fibroblasts, and impaired responses to IL-17E in T cells.[47] Recently, an AD deficiency of JNK1 has been identified in three patients from a French multiplex family. Patients presented with syndromic CMC that also included mucocutaneous infections with *S. aureus* and a connective tissue disorder. JNK1 is involved in several pathways, including the IL-17 signaling pathway. Patients' fibroblasts displayed impaired cellular responses to IL-17A and IL-17F homo- and heterodimers. In humans, JNK1 also acts downstream from transforming growth factor-β1 (TGF-β1), involved in Th17 differentiation in vitro, and, as expected, patients' peripheral blood mononuclear cells (PBMCs) displayed reduced proportions of Th17 cells.[48]

Genome-wide approaches led to the discovery of heterozygous *STAT1* missense mutations in patients with CMC (OMIM 614162).[49,50] These mutations, unlike the previously reported mono- or biallelic *STAT1* LOF mutations associated with susceptibility to mycobacterial, intracellular bacterial,

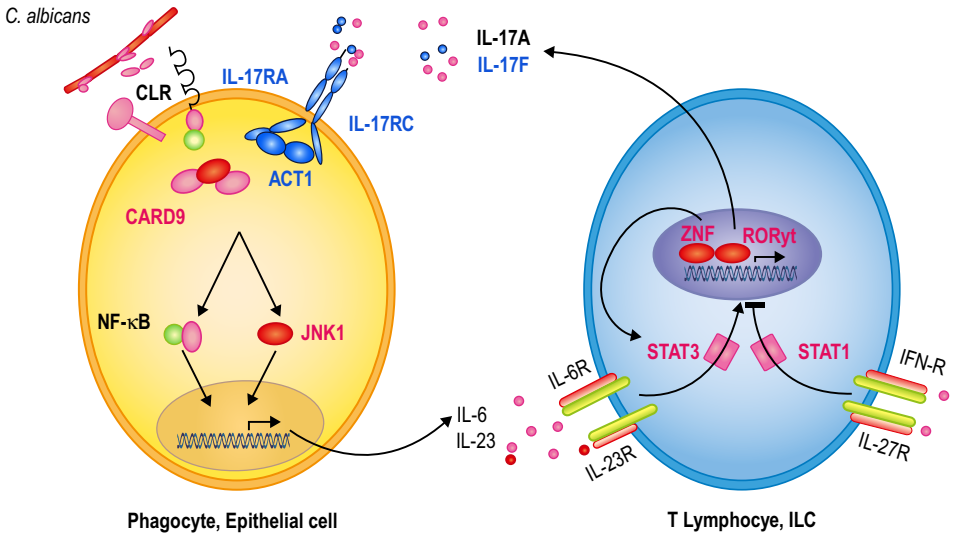


FIG. 35.5 Genetic Etiologies Affecting the Interleukin-17-Mediated Immunity. Schematic representation of interleukin-17 *(IL-17F)*-mediated immunity and cooperation between cells recognizing *Candida albicans* (phagocytes and epithelial cells) and cells producing IL-17 cytokines (T and innate lymphoid cells *[ILC]*). Proteins for which a mutation of the corresponding gene has been recognized to cause isolated chronic mucocutaneous candidiasis (CMC) are shown in *blue*, and those responsible for syndromic CMC are shown in *red*. CMC causing mutations of *IL17F*, *IL17RA*, *IL17RC*, *ACT1*, and *MAPK8* (encoding JNK1) impair responses to IL-17A/F response. CMC causing mutations of *IL12B*, *IL12RB1*, *STAT1* (GOF), *STAT3*, *ZNF341*, *RORC*, and *CARD9* genes impair the development of IL-17-producing T cells.

and viral infections, were shown to be GOF (see Fig. 35.5).[18] CMCD-causing *STAT1* mutations enhance the responses of STAT1 to IFN-α/β, IFN-γ, and IL-27, thereby repressing IL-17 T-cell development, likely accounting for the low IL-17 T-cell counts in these patients, resulting in CMCD.[18] Presently over 400 patients with STAT1 GOF have been identified worldwide. AR complete ROR-γ/ROR-γT deficiency has recently been identified in three unrelated consanguineous families (see Fig. 35.5).[22] RORC is a DNA-binding transcription factor that plays an important role in thymopoiesis. The patients suffered from an unusual combination of BCG-osis and mild CMC.[22] Mutations identified affect both IL-17 and IFN-γ immunity. The lack of functional ROR-γT protein, a transcription factor, prevents the development of IL-17-producing T cells, accounting for the CMC observed in these patients. Surprisingly, the patients also lacked mucosal-associated invariant T (MAIT) cells and invariant natural killer T cells (iNKT), which normally produce IFN-γ, and can inhibit intracellular mycobacterial replication. Moreover, their conventional CD8 α/β and γ/δ T cells, unlike their CD4 α/β T and NK cells, produced only very small amounts of IFN-γ. ROR-γ/ROR-γT, thus plays a critical role in MAIT and iNKT development, and in the capacity of γ/δ T cells and $CCR6^{+}CXCR3^{+}CD4^{+}$ α/β Th1 cells to make IFN-γ. This results in a profound impairment of IFN-γ production by lymphocytes, leading to susceptibility to mycobacterial disease and making vaccination with BCG contraindicated.[22] Antibiotic treatment can be combined with recombinant IFN-γ in cases of disseminated mycobacterial infection. Finally, patients with hyper-IgE syndrome (HIES), whether AD or AR, caused by monoallelic dominant negative mutations of *STAT3*[51,52] or biallelic loss of function mutations of *ZNF341* (see Fig. 35.5),[53,54] respectively, display syndromic CMC, with severe skin and pulmonary staphylococcal disease, severe eczema, high serum IgE levels, and some developmental abnormalities.[18] These patients were shown to display abnormally low proportions of ex vivo and in vitro differentiated Th17 cells, due to impaired signaling downstream of cytokines important in their development (e.g., IL-6, IL-23) for STAT3, or to the disruption of ZNF341-dependent STAT3 transcription and activity.[53,54] Overall, these studies strongly suggest that IL-17 immunity plays a crucial role in defense against CMC in humans. The careful clinical description of a large series of patients would be required to determine the role of these cytokines more precisely.

THERAPEUTIC PRINCIPLES

Management of Patients With Chronic Mucocutaneous Candidiasis

- Patients with *chronic mucocutaneous candidiasis* should be managed with preventive long-term antifungal therapy, principally fluconazole followed by other antifungal drugs such as itraconazole or posaconazole.
- Antibiotic prophylaxis should be considered in select patients with cutaneous staphylococcal disease.
- Immunoglobulin replacement therapy should be considered in patients with recurrent pneumonia.
- Genetically defining the defect can lead to specific immunomodulatory therapy, such as using a Janus kinase (JAK) inhibitor in patients with STAT1 gain of function mutation.

CONCLUSION

An understanding of the molecular basis of the IEIs affecting the innate and adaptive immune responses has provided detailed insight into the pathogenesis of infections in affected patients, paving the way for genetic counseling and rational treatment

design. These IEIs should be considered in patients with unexplained infectious diseases, whether caused by single or multiple infectious agents, even if all standard immunological testing has been unrevealing. Interestingly, even common infectious diseases, such as TB, invasive pneumococcal disease, and HSE, may be associated with monogenic immune disorders. The discovery of many novel IEIs opens up exciting new perspectives, not only in increasing our understanding of immunity to pathogens but also benefiting patients. It is thought that most patients with severe infectious diseases likely suffer from an underlying IEI, and they should therefore be investigated for known and potentially unknown immunodeficiency conditions.

ON THE HORIZON

- Unexplained infectious diseases occur in patients, particularly in children, as a result of an inborn error of immunity.
- Patients with severe infectious disease should be repeatedly investigated even if all standard immunological testing has been unrevealing.
- The exploration of these defects, by application of next-generation deep-sequencing technologies, provides tools for molecular diagnosis and genetic counseling and opens up exciting new perspectives benefiting patients.
- The identification of associated genetic defects will also guide rational treatment based on improvements in our understanding of specific disease pathogenesis.

ACKNOWLEDGMENTS

We thank our collaborators, our patients, and families for their assistance in making possible the studies contained in this chapter. We thank all the members of the Human Genetics of Infectious Diseases Laboratory for helpful discussions. The Laboratory of Human Genetics of Infectious Diseases is supported by grants from The Rockefeller University Center for Clinical and The Rockefeller University. The Laboratory of Human Genetics of Infectious Diseases is supported by ANR, ANRS, NIH grants; the INSERM and the Fondation pour la Recherche Médicale (FRM).

REFERENCES

1. Casanova JL. Severe infectious diseases of childhood as monogenic inborn errors of immunity. *Proc Natl Acad Sci U S A*. 2015;112(51):E7128–E7137.
2. Zhang SY, et al. Human inborn errors of immunity to infection affecting cells other than leukocytes: from the immune system to the whole organism. *Curr Opin Immunol*. 2019;59:88–100.
3. Bustamante J, et al. Mendelian susceptibility to mycobacterial disease: genetic, immunological, and clinical features of inborn errors of IFN-gamma immunity. *Semin Immunol*. 2014;26(6):454–470.
4. Bustamante J. Mendelian susceptibility to mycobacterial disease: recent discoveries. *Hum Genet*. 2020.
5. Boisson-Dupuis S, et al. Inherited and acquired immunodeficiencies underlying tuberculosis in childhood. *Immunol Rev*. 2015;264(1):103–120.
6. Boisson-Dupuis S. The monogenic basis of human tuberculosis. *Hum Genet*. 2020;139(6–7):1001–1009.

KEY CONCEPTS

- New IEIs should be sought in patients with unexplained infectious diseases.
- Children with severe infectious diseases should be repeatedly investigated for known and unknown immunodeficiency conditions.
- The exploration of idiopathic infections leads to the discovery of new IEIs and to a better understanding of immunity to pathogens.

Principles for the Treatment of Mendelian Susceptibility to Mycobacterial Disease Patients

- Vaccination with live BCG is contraindicated.
- Multiple ATBs against mycobacteria should be administered without interruption in patients with complete IFN-γR1, IFN-γR2, IRF8 or STAT1 deficiency.
- Prolonged and aggressive antimycobacterial ATBs may be associated with subcutaneous recombinant IFN-γ in selected patients with partial IFN-γR1, IFN-γR2, IRF8, CYBB, JAK1 or STAT1 deficiency, complete IL-12p40, IL12Rβ1, IL12Rβ2, IL-23R, SPPL2A, IFN-γ or ISG15 deficiency.
- HSCT should be considered in patients with complete IFN-γR1, IFN-γR2, IRF8, or STAT1 deficiency.

Principles for the Treatment of Herpes Simplex Virus-1 Encephalitis

- Treatment with IFN-α, in parallel with acyclovir, may help to improve disease outcome in patients with TLR3 pathway deficiencies, in particular.
- Serological monitoring for HSV-1 infection should be considered in individuals carrying mutations of TLR3 pathway genes but with no anti-HSV-1 antibody detectable in serum.
- In the absence of an effective vaccine against HSV-1, acyclovir may be considered an appropriate prophylactic treatment in individuals carrying mutations of TLR3 pathway genes, even if serologically negative for HSV-1.

Principles for the Treatment of Inherited Disorders of TIR-Mediated Immunity

- Patients should receive conjugated and nonconjugated vaccines against encapsulated bacteria (pneumococcus, *H. influenzae*, meningococcus). Live vaccines (BCG, poliovirus) are contraindicated.
- A preventive treatment, including ATB prophylaxis with trimethoprim-sulfamethoxazole and/or penicillin V, should be administered throughout the life of the patient.
- Monthly prophylactic administrations of intravenous or subcutaneous immunoglobulins should be considered in selected patients.
- Empiric parenteral ATB treatment against *S. pneumoniae*, *S. aureus*, and *P. aeruginosa* should be initiated as soon as an infection is suspected or if the patient develops a moderate fever, without taking inflammatory parameters into account.
- HSCT should be considered in selected patients with genetic disorders of NEMO, IκBα, and complete deficiency of IKBKB.
- Patients with HOIL1 and HOIP deficiencies should be followed for possible muscular weakness and cardiomyopathy. Heart transplants should be considered for these patients.

Principles for the Treatment of Chronic Mucocutaneous Candidiasis

- A preventive antifungal treatment, (principally fluconazole) should be administered in the long term, followed by other antifungal drugs, such as itraconazole or posaconazole.
- ATB prophylaxis should be considered in selected patients with cutaneous staphylococcal disease.
- Monthly prophylactic administrations of intravenous immunoglobulins should be considered in selected patients with recurrent pneumonia.
- G-CSF, the JAK inhibitor ruxolitinib or HSCT may also be considered.

7. Zhang SY. Herpes simplex virus encephalitis of childhood: inborn errors of central nervous system cell-intrinsic immunity. *Hum Genet*. 2020;139(6–7):911–918.
8. Ciancanelli MJ, et al. Infectious disease. Life-threatening influenza and impaired interferon amplification in human IRF7 deficiency. *Science*. 2015;348(6233):448–453.
9. Hernandez N, et al. Life-threatening influenza pneumonitis in a child with inherited IRF9 deficiency. *J Exp Med*. 2018;215(10):2567–2585.
10. Drutman SB, et al. Fatal cytomegalovirus infection in an adult with inherited NOS2 deficiency. *N Engl J Med*. 2020;382(5):437–445.
11. Drutman SB, et al. Homozygous NLRP1 gain-of-function mutation in siblings with a syndromic form of recurrent respiratory papillomatosis. *Proc Natl Acad Sci U S A*. 2019;116(38):19055–19063.
12. de Jong SJ, et al. The human CIB1-EVER1-EVER2 complex governs keratinocyte-intrinsic immunity to beta-papillomaviruses. *J Exp Med*. 2018;215(9):2289–2310.
13. Jouanguy E. Human genetic basis of fulminant viral hepatitis. *Hum Genet*. 2000:1–8.
14. Picard C, et al. Pyogenic bacterial infections in humans with IRAK-4 deficiency. *Science*. 2003;299(5615):2076–2079.
15. von Bernuth H, et al. Pyogenic bacterial infections in humans with MyD88 deficiency. *Science*. 2008;321(5889):691–696.
16. Boisson B. The genetic basis of pneumococcal and staphylococcal infections: inborn errors of human TLR and IL-1R immunity. *Hum Genet*. 2020;139(6–7):981–991.
17. Puel A, et al. Chronic mucocutaneous candidiasis in humans with inborn errors of interleukin-17 immunity. *Science*. 2011;332(6025):65–68.
18. Puel A. Human inborn errors of immunity underlying superficial or invasive candidiasis. *Hum Genet*. 2020;139(6–7):1011–1022.
19. Jouanguy E, et al. Interferon-gamma-receptor deficiency in an infant with fatal bacille Calmette-Guerin infection. *N Engl J Med*. 1996; 335(26):1956–1961.
20. Newport MJ, et al. A mutation in the interferon-gamma-receptor gene and susceptibility to mycobacterial infection. *N Engl J Med*. 1996;335(26):1941–1949.
21. Kerner G, et al. Inherited human IFNgamma deficiency underlies mycobacterial disease. *J Clin Invest*. 2020
22. Okada S, Markle JG, Deenick EK, et al. Impairment of immunity to *Candida* and *Mycobacterium* in humans with bi-allelic *RORC* mutations. *Science*. 2015;349(6248):606–613.
23. Boisson-Dupuis S, et al. Tuberculosis and impaired IL-23-dependent IFN-gamma immunity in humans homozygous for a common TYK2 missense variant. *Sci Immunol*. 2018;3(30)
24. Zhang SY, et al. TLR3 deficiency in patients with herpes simplex encephalitis. *Science*. 2007;317(5844):1522–1527.
25. Casrouge A, et al. Herpes simplex virus encephalitis in human UNC-93B deficiency. *Science*. 2006;314(5797):308–312.
26. Zimmer B, et al. Human iPSC-derived trigeminal neurons lack constitutive TLR3-dependent immunity that protects cortical neurons from HSV-1 infection. *Proc Natl Acad Sci U S A*. 2018;115(37):E8775–E8782.
27. Zhang Q. Human genetics of life-threatening influenza pneumonitis. *Hum Genet*. 2000:1–8.
28. Ogunjimi B, et al. Inborn errors in RNA polymerase III underlie severe varicella zoster virus infections. *J Clin Invest*. 2017;127(9):3543–3556.
29. Carter-Timofte ME, et al. Varicella-zoster virus CNS vasculitis and RNA polymerase III gene mutation in identical twins. *Neurol Neuroimmunol Neuroinflamm*. 2018;5(6):e500.
30. Lim HK, et al. Severe influenza pneumonitis in children with inherited TLR3 deficiency. *J Exp Med*. 2019;216(9):2038–2056.
31. Fernandes RA, et al. Complete multilineage CD4 expression defect associated with warts due to an inherited homozygous CD4 gene mutation. *Front Immunol*. 2019;10:2502.
32. Horev L, et al. Generalized verrucosis and HPV-3 susceptibility associated with CD4 T-cell lymphopenia caused by inherited human interleukin-7 deficiency. *J Am Acad Dermatol*. 2015;72(6): 1082–1084.
33. Kosumi H, et al. Two cases of interleukin-7 (IL-7)-deficient generalized verrucosis. *Clin Infect Dis*. 2020
34. Grandemange S, et al. A new autoinflammatory and autoimmune syndrome associated with NLRP1 mutations: NAIAD (NLRP1-associated autoinflammation with arthritis and dyskeratosis). *Ann Rheum Dis*. 2017;76(7):1191–1198.
35. Belkaya S, et al. Inherited IL-18BP deficiency in human fulminant viral hepatitis. *J Exp Med*. 2019;216(8):1777–1790.
36. Picard C, Casanova JL, Puel A. Infectious diseases in patients with IRAK-4, MyD88, NEMO, or IkappaBalpha deficiency. *Clin Microbiol Rev*. 2013;24(3):490–497.
37. Miot C, et al. Hematopoietic stem cell transplantation in 29 patients hemizygous for hypomorphic IKBKG/NEMO mutations. *Blood*. 2017;130(12):1456–1467.
38. Pannicke U, et al. Deficiency of innate and acquired immunity caused by an IKBKB mutation. *N Engl J Med*. 2013;369(26):2504–2514.
39. Cardinez C, et al. Gain-of-function IKBKB mutation causes human combined immune deficiency. *J Exp Med*. 2018;215(11):2715–2724.
40. Boisson B, et al. Human IkappaBalpha gain of function: a severe and syndromic immunodeficiency. *J Clin Immunol*. 2017;37(5): 397–412.
41. Boisson B, et al. Immunodeficiency, autoinflammation and amylopectinosis in humans with inherited HOIL-1 and LUBAC deficiency. *Nat Immunol*. 2012;13(12):1178–1186.
42. Boisson B, et al. Human HOIP and LUBAC deficiency underlies autoinflammation, immunodeficiency, amylopectinosis, and lymphangiectasia. *J Exp Med*. 2015;212(6):939–951.
43. Picard C, et al. Clinical features and outcome of patients with IRAK-4 and MyD88 deficiency. *Medicine (Baltimore)*. 2010;89(6):403–425.
44. Israel L, et al. Human adaptive immunity rescues an inborn error of innate immunity. *Cell*. 2017;168(5):789–800. e10.
45. Puel A, et al. Autoantibodies against IL-17A, IL-17F, and IL-22 in patients with chronic mucocutaneous candidiasis and autoimmune polyendocrine syndrome type I. *J Exp Med*. 2010;207(2):291–297.
46. Ling Y, et al. Inherited IL-17RC deficiency in patients with chronic mucocutaneous candidiasis. *J Exp Med*. 2015;212(5):619–631.
47. Boisson B, et al. An ACT1 mutation selectively abolishes interleukin-17 responses in humans with chronic mucocutaneous candidiasis. *Immunity*. 2013;39(4):676–686.
48. Li J, et al. Chronic mucocutaneous candidiasis and connective tissue disorder in humans with impaired JNK1-dependent responses to IL-17A/F and TGF-beta. *Sci Immunol*. 2019;4(41)
49. Liu L, et al. Gain-of-function human STAT1 mutations impair IL-17 immunity and underlie chronic mucocutaneous candidiasis. *J Exp Med*. 2011;208(8):1635–1648.
50. van de Veerdonk FL, et al. STAT1 mutations in autosomal dominant chronic mucocutaneous candidiasis. *N Engl J Med*. 2011;365(1): 54–61.
51. Minegishi Y, et al. Dominant-negative mutations in the DNA-binding domain of STAT3 cause hyper-IgE syndrome. *Nature*. 2007;448(7157):1058–1062.
52. Holland SM, et al. STAT3 mutations in the hyper-IgE syndrome. *N Engl J Med*. 2007;357(16):1608–1619.
53. Beziat V, et al. A recessive form of hyper-IgE syndrome by disruption of ZNF341-dependent STAT3 transcription and activity. *Sci Immunol*. 2018;3(24).
54. Frey-Jakobs S, et al. ZNF341 controls STAT3 expression and thereby immunocompetence. *Sci Immunol*. 2018;3(24).

Hemophagocytic Lymphohistiocytosis

Michael B. Jordan and Adi Zoref Lorenz

Hemophagocytic lymphohistiocytosis (HLH) is an increasingly recognized life-threatening hyperinflammatory syndrome. The syndrome of HLH describes both patients with familial HLH (due to a variety of genetic lesions) as well as patients without clear genetic causes, many of whom have other medical conditions, including infection, malignancy, and rheumatologic disorders that are thought to contribute to its development.

KEY CONCEPTS

- A hyperactive immune response, characterized by excessive activation of T cells, which recruit innate effector cells, leading to severe, often fatal, immune-mediated pathology
- Tissue damage is due to the toxic effects of immune activation, not self-reactivity ("collateral damage" instead of autoimmunity)
- "Hemophagocytosis" refers to the appearance of macrophages eating other cells in an apparent nonspecific fashion
- Although hemophagocytic lymphohistiocytosis (HLH) has been historically divided into familial (or "primary") forms and "secondary" forms, it should be viewed as a singular syndrome that is associated with a continuum of genetic and environmental risk factors.
- The syndrome of HLH encompasses HLH diseases (requiring immune suppression for treatment) and HLH disease mimickers—diseases that create the same clinical picture but do not involve immune activation or require immune suppression

HLH may present in a variety of clinical contexts. Patients in the familial HLH (fHLH) category are usually infants or young children with a positive family history or known genetic causes. These patients have a clear risk of HLH recurrence and will generally not survive long term without hematopoietic cell transplantation (HCT). Although HLH in these patients can be associated with infections or vaccination, the immunologic activator is often not apparent.

Older children or adults who present without a family history or a known genetic cause typically have concurrent infections or medical conditions that appear to activate their HLH. These patients have historically been referred to as "secondary HLH," as opposed to "primary" (familial) HLH. However, a variety of data indicate that this dichotomy is a false one; "triggers" are seen in many patients with fHLH. Experimental studies demonstrate that even severe genetic lesions still require a trigger—suggesting that some patients have unrecognized triggers. Finally, the range of identified genetic lesions now includes milder ones, which typically present at older ages.

In certain cases the term "secondary" is still apt, due to their distinctive features and treatments: malignancy (a unique sort of "acquired genetic lesion") and rheumatologic disorders, such as systemic onset juvenile idiopathic arthritis (SJIA). Although the clinical contexts and triggers for HLH are diverse, the immunopathologic syndrome that develops in response to excessive or poorly regulated adaptive immune responses is similar in all patients. Indeed, the severity of mutations appears to underlie an individual's risk for developing HLH in response to mild or extreme immune stimuli.[1]

EPIDEMIOLOGY

The exact incidence and prevalence of HLH are unknown and remain challenging to ascertain accurately. The diagnosis of HLH is challenging because of its variable presentation and the many nonspecific clinical features it shares with other disease processes. HLH is considered to be rare, but increasing awareness and recognition of the syndrome is leading to more frequent diagnoses.

Infants are most commonly affected, with the highest incidence in those less than 3 months of age. A report estimated the prevalence of fHLH in tertiary care pediatric hospitals as 1 case per 3000 inpatient admissions.[2] A comprehensive, population-based study in Sweden reported an incidence of 1.2/1,000,000 children per year. One child per 50,000 live born developed fHLH during the 15 years study period.[3] Although HLH is primarily a pediatric disease, it is diagnosed in patients of all ages. In a large cohort from Japan, 40% of the HLH patients were adults.[4,5] The incidence may be as high as 1 in 2000 adult admissions at tertiary medical centers.[6]

PATHOPHYSIOLOGY OF HEMOPHAGOCYTIC LYMPHOHISTIOCYTOSIS

- In genetic forms, a disorder of intrinsic immune regulation
- In syndromic forms not well understood
- T cells are the keys driver of disease
- Interferon gamma (IFN-γ) is the main mediator in genetic forms

PATHOPHYSIOLOGY

HLH is a syndrome of excessive inflammation and tissue destruction due to abnormal immune activation. For fHLH at least, experimental studies in animal models have demonstrated that HLH results from excessive activation of a normal CD8 T-cell reaction (an otherwise physiologic response to infection) (Chapter 12).[7] The lack of normal cytotoxic function deregulates antigen presentation, leading to excessive T-cell activation. Several key concepts are apparent from these studies. First, the immune response itself, not the infectious or other environmental activators, drives disease pathology. Second, unlike many other

immune disorders, immune activation, not self-reactivity (or autoimmunity), is the cause of disease pathology. Third, while macrophages are clearly involved in disease development, T cells (specifically CD8 T cells) are key upstream drivers of HLH disease.[8] Finally, across multiple genetic models, interferon gamma (IFN-γ) (Chapter 14) is a key mediator of disease development and is likely the principal connector between activated T cells and activation of macrophages. Furthermore, the clinical testing of anti-IFN-γ antibody (emapalumab) has proven that IFN-γ is a key driver of disease.[9]

GENETICS

A widening spectrum of genetic lesions associated with HLH is now recognized, in which the most severe ones inevitably lead to HLH at an early age, and milder ones are associated with older onset age and more potent environmental activators. Single allele mutations typically thought of as conferring only carrier status are frequently seen in patients with HLH associated with rheumatologic disorders or malignancy.[10] Thus, genetic contribution/risk does not fit neatly into the familial/nonfamilial dichotomy.

A nonexhaustive list of genetic lesions leading to HLH is described in Table 36.1. A recent whole-exome based study has expanded the list of potentially HLH-associated genes to include other genes associated with immune function and regulation, albeit with less clear causality. However, the overall theme of all of these genetic findings is that fHLH is caused by defective immune regulation, leading to excessive activation of T cells and macrophages (Fig. 36.1).

TABLE 36.1 Genetic Mutations Associated With Hemophagocytic Lymphohistiocytosis

Gene	Disease	Apparent Mechanism of HLH Predisposition
PRF1	FHL2	Defective granule-mediated cytotoxicity
UNC13D	FHL3	
STX11	FHL4	
STXBP2	FHL5	
RAB27A	Griscelli syndrome	
LYST	Chediak-Higashi syndrome	
AP3B1	Hermanski-Pudlak syndrome	
XIAP	XLP2/ X-linked HLH	Disorders of inflammasome activation
NLRC4	Auto-inflammation with infantile enterocolitis	
CDC42	NOCARH syndrome	
SH2D1A	Lymphoproliferative syndrome, X-linked, 1 (XLP1)	Disorders of T-cell signaling
ITK	Lymphoproliferative syndrome 1	
CD27	Lymphoproliferative syndrome 2	
MAGT1	Immunodeficiency, X-linked, with magnesium defect, Epstein-Barr virus infection, and neoplasia (X-MEN)	
SLC7A7	Lysinuric protein intolerance	Disorders of macrophage inflammatory signaling
HMOX1	Heme oxygenase deficiency	

HLH, Hemophagocytic lymphohistiocytosis.

DIAGNOSIS

HLH was defined in the HLH-2004 study as fulfilling five of eight criteria: fever, splenomegaly, cytopenias in two or more lineages (platelets <100 K/mL, hemoglobin <9 g/dL, neutrophils <1000 /mL), ferritin >500 ng/mL, sCD25 >2400 U/mL, low NK cell function, hemophagocytosis on biopsy, and either fibrinogen <150 mg/dL or triglycerides >265 mg/dL. Initial symptoms of HLH are nonspecific and may overlap with other inflammatory or hematopoietic diseases. Conditions fulfilling

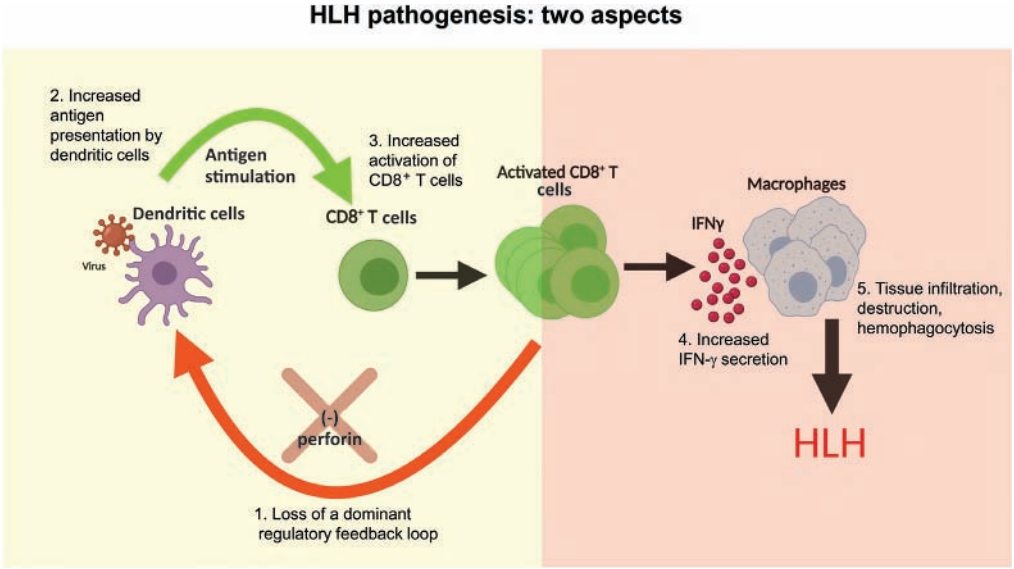


FIG. 36.1 The Pathogenesis of Familial Hemophagocytic Lymphohistiocytosis (*HLH*). T-cell responses are normally shaped by negative feedback involving the perforin pathway in which cytotoxic T cells (Chapter 12) eliminate antigen-presenting cells (Chapter 6). Patients with familial HLH (fHLH) often have defects of the perforin pathway. Experimental studies have demonstrated that defects of this feedback regulation result in excessive antigen presentation by dendritic cells, leading to excessive T-cell activation (Chapter 10), and overproduction of interferon-gamma (*IFN-γ*) (Chapter 14), leading to pathologic macrophage activation. Thus, clinical HLH results from a primary defect of immune regulation leading to severe immunopathology, which is largely driven by IFN-γ.

this syndromic definition may be conceptually split into *HLH disease* (fHLH and others) and *mimickers* of HLH disease. This distinction is important as HLH disease requires prompt and aggressive immune suppression, while other conditions may require different therapies or may be worsened by immune suppression. For example, a patient with a visceral *Leishmania* infection could present with this clinical pattern without having "inappropriate" immune activation and would be harmed by immune suppression.

Why is this concept valuable? HLH typically requires prompt recognition and treatment, often before all data can be gathered. However, this aggressive stance should be balanced by a cautious and careful exclusion of conditions that look like HLH but would not benefit from or even be harmed by immune suppression and may require different specific treatments. Undiagnosed malignancy is especially problematic as HLH-directed therapy often obscures diagnosis.[11] Moreover, the cases should be divided by the HLH activator because each has its characteristics and treatment.[12]

Since T-cell activation is central to HLH pathogenesis, sCD25 should always be elevated in untreated HLH.[13] If not elevated, then one should doubt a diagnosis of HLH. Similarly, though not as well established, because HLH appears to be primarily driven by IFN-γ, elevations of CXCL9 (Chapter 15) (a sensitive indicator of IFN-γ bioactivity) should be seen in untreated cases of HLH disease).[14] Ferritin levels above 10,000 ng/mL appear to be relatively specific for HLH in children but are not very sensitive.[15] While specialized immunologic testing may facilitate diagnosis, if a diagnosis can be made without them, then treatment should not be delayed pending these results. Likewise, treatment should not be delayed for assessment of central nervous system (CNS) involvement, though this should always be conducted (once a lumbar puncture can be safely performed).

Most features of HLH are unexpected in either their high levels (ferritin, sCD25, triglycerides), distinctiveness (splenomegaly and hemophagocytosis), or acute appearance (anemia). There are also "ironic" features that are the opposite of what we expect to see in a state of inflammation such as neutropenia, thrombocytopenia, low NK function, and fibrinogen level. Though many HLH diagnostic features are unique, their presence or absence should not automatically determine the diagnosis. Moreover, they have not been validated for secondary HLH and efforts to determine specific diagnostic criteria are ongoing.

OTHER MANIFESTATIONS

HLH may present in many forms, including fever of unknown origin (FUO); hepatitis or acute liver failure; and sepsis-like, Kawasaki-like, and familial neurologic abnormalities. Not all of the HLH diagnostic criteria may be present initially, so it is vital to follow clinical signs and laboratory markers of pathologic inflammation repeatedly to identify the trends. Typical clinical features seen in patients with HLH, grouped by organ system, are described below.

Prolonged Fever is a common manifestation. In patients with FUO, cytopenia, highly elevated ferritin (>3000 g/dL), or sCD25 significantly above age-adjusted normal ranges generally suggest that a complete HLH diagnostic evaluation should be pursued.

Liver Disease and Coagulopathy. Most patients with HLH have variable evidence of hepatitis at presentation. HLH should be considered in the differential diagnosis of acute liver failure, mainly if lymphocytic infiltrates are noted on biopsy. Most patients have evidence of disseminated intravascular coagulation (DIC) and are at high risk for acute bleeding.

Blood Lineages Cytopenia. Anemia and thrombocytopenia occur in more than 80% of patients at the time of presentation with HLH. Although hemophagocytosis in bone marrow is associated with HLH, the morphologic phenomenon may also be induced by blood transfusions, infection, an autoimmune disease, or other events. Despite the terminology of HLH, the diagnosis of HLH should never be made or excluded solely on the presence or absence of hemophagocytosis.

Neurologic Symptoms. More than one-third of patients present with neurologic symptoms. These include seizures, meningismus, decreased level of consciousness, cranial nerve palsy, psychomotor retardation, ataxia, irritability, and hypotonia. The cerebrospinal fluid (CSF) is abnormal in more than 50% of fHLH patients with findings of pleocytosis, elevated protein, or hemophagocytosis. Magnetic resonance imaging (MRI) findings are highly variable and include discrete lesions, leptomeningeal enhancement, or global edema, and images that correlate with neurologic symptoms.[12]

DISTINCTIVE CLINICAL CONTEXTS

Hemophagocytic Lymphohistiocytosis in the Context of a Rheumatologic Disease (R-HLH)

Macrophage activation syndrome (MAS) is the term most commonly used to refer to an HLH or HLH-like syndrome occurring in the context of rheumatologic disorders (Chapter 52 and 54). R-HLH is most commonly associated with juvenile idiopathic arthritis (SJIA) (Chapter 54), systemic lupus erythematosus (SLE) (Chapter 52), or adult-onset Still disease. While MAS and HLH are very similar and should be viewed as the same disease, there are notable differences in presentation. Consensus criteria for recognizing MAS in the context of SJIA have been developed, and, in general, these MAS patients are older than fHLH patients and present with substantially higher platelet and neutrophil counts as well as higher fibrinogen levels. These laboratory values are typically elevated in patients with SJIA, so normal levels may be viewed as "abnormally normal" in MAS.[16]

Hemophagocytic Lymphohistiocytosis in the Context of Malignancy (M-HLH)

The association of HLH with malignancy has been recognized for decades. Patients may present with the clinical syndrome of HLH associated with undiagnosed underlying malignancy or transformation of an indolent hematologic malignancy, or they may develop HLH during treatment for known malignancy, usually in the context of infection. The pathophysiology of M-HLH is not well defined, and the tumor itself may "mimic" the HLH diagnostic criteria. However, the prognosis of these patients is abysmal and significantly worse than other HLH etiologies.

Lymphoma deserves special mention (Chapter 78), as it is the most common malignancy associated with HLH at its initial presentation. Because of the difficulty of distinguishing lymphoma from fHLH or R-HLH, thorough imaging and aggressive biopsy, often guided by positron emission tomography/computed tomography (PET/CT), should be pursued or at least considered before starting corticosteroids and other therapies that may obscure diagnosis.[11] Malignancy may present with HLH at all ages (including infancy), but is increasingly likely at older ages and is associated

with perhaps a majority of cases in adults.[5,17] When considering M-HLH, it is essential to note that the presence of Epstein-Barr virus (EBV) viremia does not rule out malignancy (including B- or T-cell lymphomas), and sCD25 may be disproportionately elevated compared to other features of HLH in patients with occult lymphoma.[18]

CLINICAL PEARLS

Differential Diagnosis—an Hemophagocytic Lymphohistiocytosis Trigger or a Mimic of Hemophagocytic Lymphohistiocytosis Disease?

Infections

- Most atypical infections should be considered a "mimic" as hyperinflammation is not the central problem.
- Infection-specific therapy is needed, and immunosuppressive therapy is likely to be harmful.
- Common infections include visceral leishmaniasis, atypical/tuberculous mycobacteria, histoplasmosis, ehrlichia, bartonella and Brucella species, disseminated adenovirus, and disseminated herpes simplex.
- Many other (mostly viral) infections may be viewed as "triggers" of hemophagocytic lymphohistiocytosis (HLH) disease, including EBV and Cytomegalovirus CMV (which may also require infection-specific therapies).

Other Hematologic Disorders

- Mimics include Langerhans cell histiocytosis involving the marrow and/or visceral organs and multicentric Castleman disease, especially the TAFRO (thrombocytopenia, anasarca, myelofibrosis, renal dysfunction, and organomegaly) variant.
- Therapy should be directed towards the underlying disorder, +/– additional corticosteroids.

Drug Reactions

- Drug Rash with Eosinophilia and Systemic Symptoms, or DRESS syndrome, may present as HLH.
- It may be considered as both trigger and mimic as treatment requires both withdrawal of the offending agent and prolonged corticosteroids.

Storage Diseases

- Mimics include Wolman disease (infantile lysosomal acid lipase deficiency) and Gaucher disease.
- These disorders develop features of HLH (e.g., splenomegaly, cytopenias) due to processes not involving immune hyperactivation.

Metabolic Disorders

- Lysinuric protein intolerance (LPI) and others may be considered as mimics of HLH disease, as they require different, specific treatment.
- The inflammatory features of LPI in particular are notable and may be viewed as overlapping with fHLH.

Hemophagocytic Lymphohistiocytosis in the Context of Immune Compromise (IC-HLH)

HLH can occur in a variety of immune compromised patients, including primary immune deficiency (PID) or patients receiving immunosuppressive therapy, mostly in the context of unresolved infection. For instance, patients with inflammatory bowel disease (IBD) (Chapter 75)—usually treated with azathioprine or mercaptopurine—have been reported to develop (usually relatively mild) HLH after infection with EBV or cytomegalovirus (CMV). In these contexts, the pathophysiology is not well understood. While it may be related to the underlying disease, HLH appears to be a dysregulated response to infection in the context of immunosuppression. It is difficult to generalize about whether IC-HLH may benefit from significant immune suppression. For IBD patients, withdrawal of mercaptopurine, treatment of infection, supportive care, and moderate-dose corticosteroids often suffice. Thus, IC-HLH ambiguously straddles the categories of HLH disease and HLH disease mimics.

A variety of PIDs have been reported to present with HLH.[19] These include combined or selective T-cell disorders (Chapter 34), such as severe combined immunodeficiency (SCID), Omenns syndrome, severe DiGeorge syndrome, Wiskott-Aldrich syndrome, and autoimmune lymphoproliferative syndrome; selective B-cell disorders such as X-linked agammaglobulinemia (Chapter 33); and neutrophil disorders such as chronic granulomatous disease (CGD) (Chapter 39).

Patients with PID and HLH often have unresolved, severe infections. Patients with SCID most often have viral infections, while those with CGD present with bacterial infections. Thus, the presence of HLH associated with unusual or unusually severe infection should suggest undiagnosed immune deficiency.

For patients with SCID and infection, immunosuppressive therapy is generally not helpful, and this condition should be considered to be a mimic of HLH disease. HLH in CGD patients is less clear, though typical treatment for HLH beyond corticosteroids is usually not indicated. Thus, the constellation of HLH symptoms in the context of PID should typically be considered a mimic of HLH disease, though some patients may require immunosuppressive therapy.

Hemophagocytic Lymphohistiocytosis in the Context of Immune-activating Therapies (Rx-HLH)

The HLH syndrome develops in some patients receiving immune-activating therapies, such as T-cell engaging antibodies, CAR T cells (Chapter 81), or immune checkpoint inhibitors (Chapter 80). In this context, this syndrome is usually referred to as cytokine release syndrome (CRS). However, its pathophysiology appears quite similar to fHLH and should be recognized as iatrogenic HLH, or Rx-HLH.

TREATMENT

Prompt and aggressive initiation of treatment may be lifesaving.

HLH-94 Protocol

Currently, the standard of care for HLH should be considered to be treatment with etoposide and dexamethasone as per a slightly modified HLH-94 protocol.[20] In general, immediate treatment of HLH is warranted once a diagnosis is made. At the same time, it is essential to rule out HLH disease mimics or malignancies before starting therapy to avoid inappropriate treatment and obscuring the underlying diagnosis. The risk of patients deteriorating during prolonged diagnostic evaluation must be balanced against these concerns, especially in the face of limited data.

If treatment is initiated before a firm diagnosis is established, the uncertainty of the diagnosis should be revisited once the patient is stable. Though aggressive treatment is needed for most patients, initial treatment with dexamethasone alone with close inpatient monitoring may be appropriate in patients who are not infants and not severely ill.

Salvage Therapy

Emapalumab, an interferon-gamma (IFN-γ) blocking monoclonal antibody, was approved by the US Food and Drug Administration for refractory or recurrent HLH.[9] Another well-supported salvage therapy is alemtuzumab (anti CD52), which can be used as a bridge to bone marrow transplant (BMT).

Targeted Therapy

As mentioned, anti IFN-γ (emapalumab) was the first targeted therapy approved and is now being examined for frontline use in HLH and in specific subsets (R-HLH, M-HLH). A JAK1/2 inhibitor (ruxolitinib) (Chapter 86) is being investigated currently in several clinical trials. Anti-cytokine treatment is common in R-HLH (IL-1 or IL-6 blocking agents), though these patients are increasingly diagnosed with chronic complications, such as lung disease. Blockade of IL-18 (Chapter 14) is being investigated for patients with HLH driven by the inflammasome.[21]

Allogeneic Hematopoietic Stem Cell Transplantation

Allogeneic hematopoietic stem cell transplantation (HSCT) is performed to prevent potentially fatal HLH disease recurrence. Patients with clear family histories and/or genetic etiologies for HLH, as well as those with significant CNS involvement or who have recurrent disease, should proceed to BMT as soon as the disease is reasonably well controlled. Moreover, patients with a hematologic malignancy that cannot be cured with conventional agents (but may be treated with HSCT) should also proceed to transplant (Chapter 92).

Relapsing HLH, however, does not always reflect an underlying genetic predisposition; sometimes, it may indicate undiagnosed underlying infection or malignancy rather than a genetic etiology. Since genetic risks are usually not known when patients present, HLA typing (Chapter 5), and preliminary donor searches should be performed early after diagnosis. Sibling donors should be screened for HLH-associated genetic lesions. Decisions regarding whether or not to proceed to transplantation hinge upon an assessment of HLH recurrence risk, balanced against the risks of transplantation.

ON THE HORIZON

- Clarification of the cellular and molecular mechanisms underlying nonfamilial hemophagocytic lymphohistiocytosis (HLH) types, development of new diagnostic tools, and improvement of therapeutic strategies.
- Use of large cohorts to validate and optimize diagnostic criteria for both familial and secondary forms of HLH.
- Clinical trials of novel therapeutic agents. Examples at the time this chapter was written include:
  - Emapalumab for fHLH, for adults with HLH, or in SJIA-MAS/HLH.
  - Recombinant IL-18-binding protein (tadekinig α) for HLH driven by XIAP or NLRC4 (see Table 36.1).
  - Ruxolitinib for HLH.

HLH, Hemophagocytic lymphohistiocytosis; *MAS*, macrophage activation syndrome; *SJIA*, systemic onset juvenile idiopathic arthritis.

SUMMARY

HLH has many forms, but it is always a disorder of excessive immune activation. fHLH is a unique PID involving excessive T-cell activation and recruitment of innate effector cells. Without prompt diagnosis and treatment, the natural history of HLH is almost uniformly fatal. Rising awareness of HLH in recent years has improved recognition, but also increased the risk of misdiagnosis and inappropriate treatment of HLH-mimicking conditions. An evolving understanding of HLH pathophysiology is beginning to alter therapy for this disease (e.g., emapalumab for second-line treatment of HLH). The recognition of clinical patterns is essential for diagnosis. However, because of the nonspecific nature of the clinical symptoms, the differential diagnoses of other inflammatory conditions must be considered to avoid unnecessary or harmful immune suppression induced by HLH therapy.

REFERENCES

1. Risma K, Jordan MB. Hemophagocytic lymphohistiocytosis: updates and evolving concepts. *Curr Opin Pediatr*. 2012;24(1):9–15.
2. Jordan MB, Allen CE, Weitzman S, et al. How I treat hemophagocytic lymphohistiocytosis. *Blood*. 2011;118(15):4041–4052.
3. Henter JI, Elinder G, Soder O, Ost A. Incidence in Sweden and clinical features of familial hemophagocytic lymphohistiocytosis. *Acta Paediatr Scand*. 1991;80(4):428–435.
4. Ishii E, Ohga S, Imashuku S, et al. Nationwide survey of hemophagocytic lymphohistiocytosis in Japan. *Int J Hematol*. 2007;86(1):58–65.
5. Ramos-Casals M, Brito-Zeron P, Lopez-Guillermo A, et al. Adult haemophagocytic syndrome. *Lancet*. 2014;383(9927):1503–1516.
6. Parikh SA, Kapoor P, Letendre L, et al. Prognostic factors and outcomes of adults with hemophagocytic lymphohistiocytosis. *Mayo Clin Proc*. 2014;89(4):484–492.
7. Jordan MB, Hildeman D, Kappler J, Marrack P. An animal model of hemophagocytic lymphohistiocytosis (HLH): $CD8^+$ T cells and interferon gamma are essential for the disorder. *Blood*. 2004;104(3):735–743.
8. Zoller EE, Lykens JE, Terrell CE, et al. Hemophagocytosis causes a consumptive anemia of inflammation. *J Exp Med*. 2011;208(6):1203–1214.
9. Locatelli F, Jordan MB, Allen C, et al. Emapalumab in children with primary hemophagocytic lymphohistiocytosis. *N Engl J Med*. 2020;382(19):1811–1822.
10. Kaufman KM, Linghu B, Szustakowski JD, et al. Whole-exome sequencing reveals overlap between macrophage activation syndrome in systemic juvenile idiopathic arthritis and familial hemophagocytic lymphohistiocytosis. *Arthritis Rheumatol*. 2014;66(12):3486–3495.
11. Gurunathan A, Boucher AA, Mark M, et al. Limitations of HLH-2004 criteria in distinguishing malignancy-associated hemophagocytic lymphohistiocytosis. *Pediatr Blood Cancer*. 2018;65(12):e27400.
12. Jordan MB, Allen CE, Greenberg J, et al. Challenges in the diagnosis of hemophagocytic lymphohistiocytosis: recommendations from the North American Consortium for Histiocytosis (NACHO). *Pediatr Blood Cancer*. 2019;66(11):e27929.
13. Lin M, Park S, Hayden A, et al. Clinical utility of soluble interleukin-2 receptor in hemophagocytic syndromes: a systematic scoping review. *Ann Hematol*. 2017;96(8):1241–1251.
14. Bracaglia C, de Graaf K, Pires Marafon D, et al. Elevated circulating levels of interferon-gamma and interferon-gamma-induced chemokines characterise patients with macrophage activation syndrome complicating systemic juvenile idiopathic arthritis. *Ann Rheum Dis*. 2017;76(1):166–172.
15. Allen CE, Yu X, Kozinetz CA, et al. Highly elevated ferritin levels and the diagnosis of hemophagocytic lymphohistiocytosis. *Pediatr Blood Cancer*. 2008;50(6):1227–1235.
16. Gansner JM, Berliner N. The rheumatology/hematology interface: CAPS and MAS diagnosis and management. *Hematology Am Soc Hematol Educ Program*. 2018;2018(1):313–317.
17. Lehmberg K, Nichols KE, Henter JI, et al. Consensus recommendations for the diagnosis and management of hemophagocytic lymphohistiocytosis associated with malignancies. *Haematologica*. 2015;100(8):997–1004.
18. Tabata C, Tabata R. Possible prediction of underlying lymphoma by high sIL-2R/ferritin ratio in hemophagocytic syndrome. *Ann Hematol*. 2012;91(1):63–71.
19. Bode SF, Ammann S, Al-Herz W, et al. The syndrome of hemophagocytic lymphohistiocytosis in primary immunodeficiencies: implications for differential diagnosis and pathogenesis. *Haematologica*. 2015;100(7):978–988.
20. Ehl S, Astigarraga I, von Bahr Greenwood T, et al. Recommendations for the use of etoposide-based therapy and bone marrow transplantation for the treatment of HLH: Consensus statements by the HLH Steering Committee of the Histiocyte Society. *J Allergy Clin Immunol Pract*. 2018;6(5):1508–1517.
21. Broglie L, Pommert L, Rao S, et al. Ruxolitinib for treatment of refractory hemophagocytic lymphohistiocytosis. *Blood Adv*. 2017;1(19):1533–1536.

37

Autoinflammatory Syndromes

Catharina M. Mulders-Manders, Jeroen C.H. van der Hilst, Jos W.M. van der Meer and Anna Simon

Autoinflammatory diseases, which are also known as *periodic fever syndromes*, encompass a group of rare disorders characterized by recurrent or persistent inflammation. *Autoinflammation* is a term that has been used since the late 1990s to illustrate the difference between autoimmune disorders and diseases characterized by exuberant inflammation. Typically, autoinflammatory diseases do not show features of excess adaptive immune system activation, and autoantigens or autoantigen-specific T cells are not present in these diseases. It is now recognized that autoinflammation and autoimmunity form two ends of a spectrum of inappropriate immune system activation and share several common features.

Located at the autoinflammatory end of this spectrum are the classic monogenic autoinflammatory diseases: familial Mediterranean fever (FMF), cryopyrin-associated periodic syndrome (CAPS), mevalonate kinase deficiency (MKD; also known as *hyperimmunoglobulin D and periodic fever syndrome* [HIDS]), and tumor necrosis factor (TNF) receptor–associated periodic syndrome (TRAPS). The number of autoinflammatory diseases is increasing rapidly. New monogenic autoinflammatory diseases, as well as new autoinflammatory diseases without a clear genetic background, have been identified in the past decades. It has also become clear that autoinflammation is at least partially involved in the pathogenesis of other, more common diseases, such as gout, Crohn disease, and ulcerative colitis.

Because it is impossible to discuss all autoinflammatory diseases in detail here, the classic monogenic diseases FMF, CAPS, TRAPS, and MKD have been selected as the main focus of this chapter. Their pathophysiologic mechanisms are understood to a much higher degree than in many newer autoinflammatory diseases, and their clinical presentations have been described precisely.

In addition, two other autoinflammatory diseases are discussed, one with relatively high prevalence and the other because of its interesting pathophysiologic mechanism: (i) periodic fever, aphthous stomatitis, pharyngitis, and adenitis (PFAPA) syndrome and (ii) Schnitzler syndrome.

The cornerstone of diagnosing an autoinflammatory disease is the clinical assessment of the patient. This includes a detailed medical and family history and direct observation of an inflammatory episode. The first step in the diagnostic process is to exclude other more common causes of recurrent inflammation, including infections, malignancy and paraneoplastic phenomena, and autoimmune disease.[1] A first differential diagnosis can be made on the basis of age of onset, associated signs and symptoms, duration of inflammation, family history, and ethnic background, (Table 37.1), and this can guide targeted diagnostic testing.

KEY CONCEPTS

Autoinflammation Versus Autoimmunity

- Common features:
  - Inflammation due to excessive immune activation
  - Phenotypes characterized by exacerbations and remissions.
- Distinctive features:
  - Autoinflammation: dysregulation of innate immunity, no high-titer autoantibodies or autoantigen-specific T cells
  - Autoimmunity: dysregulation of adaptive immunity, defect in lymphocyte function, autoantibodies may be present.
- Autoinflammation and autoimmunity form two ends of a continuous spectrum of excessive immune system activation.
- A number of diseases show overlapping features between autoinflammation and autoimmunity.

EPIDEMIOLOGY

It is important to realize that the incidence of specific diseases varies widely among ethnic groups. With more than 100,000 patients worldwide, FMF is the most prevalent monogenic autoinflammatory disease. It is most common in individuals originating from around the Mediterranean basin, such as Turks, Jews (primarily non-Ashkenazi), Arabs, and Armenians. In these selected populations, the carrier frequency of mutations in the Mediterranean fever gene *(MEFV)* can be as high as one in three individuals.[1] This may indicate a survival benefit for carriers of heterozygous mutations, possibly through protection against infectious agents.

The first patients with MKD were described in 1984 in the Netherlands[2] (then referred to as HIDS). More than 200 patients have now been identified, most of Western European and Caucasian ancestry. This could be partly explained by increased awareness for this disease among physicians in that part of the world. An alternative explanation is a common founder effect with clustering of carriers, illustrated by a carrier rate of 1:153 for the most common mutation in the mevalonate kinase gene *(MVK)* (V377I) in Dutch newborns.[3,4]

TRAPS is seen in patients from around the world, although most patients originate from northwestern Europe. A few dozen families and more than 200 sporadic cases have been reported.

The exact prevalence of the CAPS is unknown, but more than 200 cases have been reported. Disease awareness and recognition among clinicians have improved because of the availability of effective treatment for this disease.

PFAPA syndrome was first reported at the end of the 1980s.[5] It is difficult to estimate the incidence of PFAPA, because the level

TABLE 37.1 Classic Monogenetic Autoinflammatory Diseases

	FMF	CAPS	TRAPS	MKD
Mode of inheritance	Autosomal recessive	Autosomal dominant	Autosomal dominant	Autosomal recessive
Age of onset (years)	<20	Generally <1, in MWS/FCAS <20 possible	Variable, most <10	<1
Main ethnic distribution	Turks, Arabs, Jews, Armenians	Europeans	All	Northwestern Europeans (Dutch, French)
Gene involved	*MEFV*	*NLRP3*	*TNFRSF1A*	*MVK*
Protein involved	Pyrin	NLRP3	TNF receptor type 1	Mevalonate kinase
Duration of typical attack	2–3 days	Variable; hours–days or continuous inflammation	Days–weeks	HIDS: 4–6 days; MA: continuous, flares possible
Distinguishing symptoms	Peritonitis, pleuritis, erysipelas-like skin lesions	Aseptic meningitis; sensorineural deafness; bone lesions, dysmorphic features May be cold induced	Severe myalgia, periorbital edema	HIDS: lymphadenopathy, attacks induced by vaccination MA: joint contractions, growth and developmental delay
Risk of amyloidosis[a]	Up to 75%	Up to 33%	25%	<5%
Treatment	Colchicine, combination with IL-1 inhibition when resistant	IL-1 inhibition	Mild disease: NSAIDs; Severe disease: IL-1 inhibition	IL-1 inhibition

[a]In patients with long-term uncontrolled inflammation.
CAPS, Cryopyrin-associated periodic syndrome; *FCAS*, familial cold autoinflammatory syndrome; *FMF*, familial Mediterranean fever; *HIDS*, hyperimmunoglobulin D and periodic fever syndrome; *IL*, interleukin; *MA*, mevalonic aciduria; *MEFV*, Mediterranean fever gene; *MKD*, mevalonate kinase deficiency; *MWS*, Muckle-Wells syndrome; *NSAIDs*, nonsteroidal antiinflammatory drugs; *TRAPS*, tumor necrosis factor (TNF) receptor–associated periodic syndrome.

of awareness of this disease seems to vary among clinicians. A single pediatric center in the United States reported 122 patients fulfilling the criteria for PFAPA in 10 years,[6] making it more common than any of the monogenic autoinflammatory diseases (with the exception of FMF in certain populations). In most patients with PFAPA, symptoms cease before or during adolescence. The cause of this spontaneous resolution is unknown. The characteristic combination of symptoms of PFAPA has been described in adults, but it remains a matter of debate whether these patients suffer from true PFAPA syndrome.

Schnitzler syndrome, first described by the French dermatologist Schnitzler in 1972,[7] is an acquired autoinflammatory disorder with a median age of onset of 51 years. More than 160 cases have been reported worldwide.

SIGNS AND SYMPTOMS

Familial Mediterranean Fever

FMF is an autosomal recessive disease. More than 90% of patients become symptomatic within the first two decades of life. Typical attacks are characterized by abrupt onset of high fever, peaking soon after onset and lasting for 12 hours to 3 days. Subsequently, the fever subsides rapidly. Painful serositis accompanies the fever. Serositis can also be present without fever. More than 95% of patients experience abdominal pain, which lasts up to 3 days. The pain, which is caused by sterile peritonitis, may initially be focal and then progresses to more diffuse pain. Before being diagnosed with FMF, a significant proportion of patients will have undergone exploratory abdominal surgery under suspicion of appendicitis. At surgery, intra-abdominal adhesions, a result of recurrent peritonitis, may be found. Pelvic adhesions can reduce fertility in female patients. Pleuritis, presenting as thoracic pain, is experienced by approximately 40% of patients. Synovitis with monoarthritis of knee, ankle, or wrist occurs in one-half to three-quarters of patients. An arthritic attack may have a more protracted course compared with nonarthritic FMF, with fever lasting up to a week. Joint pain may persist when fever has already subsided. Synovitis usually resolves completely without joint destruction. The skin can be affected. Erysipelas-like skin lesions overlying the shins are very characteristic of FMF but are seen only in 30% of patients. Less frequent symptoms of FMF include vasculitis, orchitis, aseptic meningitis, and myalgia. Pericarditis is rare in FMF.

There are no consistent triggers for FMF attacks. Emotional stress or menstruation may increase the frequency of attacks; some patients are able to report specific triggers for their attacks. Attack frequency varies greatly among patients and during an individual patient's life. Some patients suffer from continuous inflammation, whereas others have attacks once a year or even less frequently.

The literature on FMF is replete with genotype-phenotype studies. The most consistent finding is that carriers of the M694V/M694V genotype have more severe disease, with earlier onset and higher frequency of arthritis and long-term complications.

Life expectancy of patients with FMF depends on timely initiation of appropriate treatment to prevent amyloidosis. Without amyloidosis, FMF patients have normal life expectancy.

Cryopyrin-Associated Periodic Syndrome

CAPS is autosomal dominantly inherited. Originally, three separate clinical syndromes, all with their own clinical features, were distinguished: familial cold autoinflammatory syndrome (FCAS), Muckle-Wells syndrome (MWS), and neonatal-onset multisystem inflammatory disease (NOMID), which is also known as chronic infantile neurologic, cutaneous, and articular (CINCA) syndrome. With the discovery of *NLRP3* mutations in all three diseases, it has become clear that the clinical phenotype of CAPS is a continuous spectrum of severity, instead of distinct diseases. There is no genotype-phenotype association, suggesting a role for other yet undiscovered disease-modifying factors.

CAPS often manifests clinically soon after birth or in early childhood. It is characterized by recurrent urticaria-like rash, arthralgias, myalgias, headaches, and fever. Ocular symptoms, including conjunctivitis and uveitis, are common. Some

patients develop sensorineural hearing loss during adolescence or adulthood. At the severe end of the clinical spectrum, central nervous system (CNS) symptoms, including chronic aseptic meningitis that is characterized by chronic headache, increased intracranial pressure, hydrocephalus, mental retardation, and seizures, are common. Papilledema with optic nerve atrophy can lead to loss of vision. In severely affected patients, arthropathy, with distinct radiographic findings of premature patellar and epiphyseal long-bone ossification and osseous overgrowth, develops early in life. If left untreated, this leads to growth retardation, severe joint contractures, and persisting disability.

The duration of attacks is variable and ranges from hours to days. At the severe end of the spectrum, patients have continuous inflammation. Attacks may be triggered by exposure to cold, minor trauma, or emotional stress.

In the past, patients with severe CAPS often died in childhood. This changed after the introduction of anti–interleukin-1 (IL-1) therapy, which is very effective in treating CAPS. Overall, patients without neurologic involvement are now believed to have a normal life expectancy.

Tumor Necrosis Factor Receptor–Associated Periodic Syndrome

TRAPS is inherited in an autosomal dominant fashion. Age of onset varies widely. Many patients become symptomatic within the first years of life, with a median age of onset of 3 years, but adult onset is also possible. The usual duration of fever in TRAPS is considerably longer than in the other classic autoinflammatory syndromes: attacks persist for a minimum of 3 days but can last for several weeks. The interval between attacks in a single patient can vary substantially.

Localized myalgia, a deep cramping, and often severely disabling pain in a single limb resulting from monocytic fasciitis and associated with fever are found in virtually all patients. The affected limb may show local erythema, which may migrate to the distal part of the extremity (Fig. 37.1). Almost all patients have abdominal pain, often accompanied by vomiting, constipation, and bowel obstruction. Arthralgia and monoarthritis involving hips, knees, or ankles are present in 25% of patients at some point. Chest pain is frequent and can be caused by pleuritis or may be musculoskeletal in origin. Ocular symptoms range from conjunctivitis and periorbital pain to severe uveitis and iritis. Periorbital edema with conjunctival injection is a distinctive, but infrequent, feature of TRAPS. Other less frequently observed symptoms are pericarditis and lymphadenopathy.

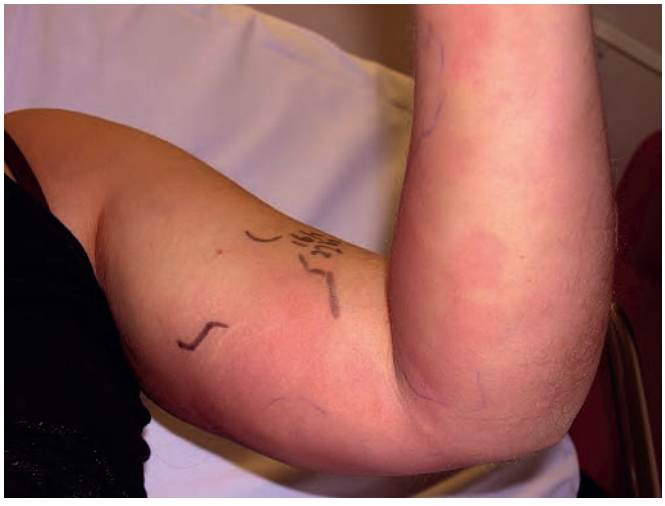


FIG. 37.1 Migratory erythematous macular rash during an inflammatory attack in a patient with tumor necrosis factor receptor–associated periodic syndrome.

Mevalonate Kinase Deficiency

Before the discovery of the underlying genetic defect in MVK, two distinct autosomal recessive diseases were distinguished, which are now known to form two ends of a continuous spectrum: HIDS at the less severe end and mevalonic aciduria (MA) at the most severe end.

HIDS is characterized by recurrent fever attacks that last for 4 to 6 days, starting in early childhood. The inflammatory attacks occur on average every 4 to 6 weeks. Attack frequency varies in a single patient and among patients and tends to decrease later in life. Attacks often start with chills, followed by a rapid rise in temperature. Factors that can provoke an attack are infections, trauma, vaccination, and both physical and emotional stress, although a clear trigger is often absent. Characteristic for HIDS is the first attack being triggered by childhood vaccination.

Fever is accompanied by cervical lymphadenopathy and abdominal pain with vomiting and diarrhea. The skin may show erythema, papules, urticarial rash, or exanthema. The majority of patients suffer from large-joint arthralgia or arthritis. Oral or genital aphthous ulcers may be present during attacks. Hepatosplenomegaly has been reported. Patients with HIDS appear to have normal life expectancy and experience no complications.

MA is located at the severe end of the MKD spectrum. This severe disease is present from birth and is characterized by psychomotor retardation, ataxia, failure to thrive, cataracts, and facial dysmorphia. Episodic fever or inflammation are present in MA. Many patients die in early childhood.

In recent years, it has become clear that the spectrum of MKD comprises more than only these two diseases. Mutations in *MVK* have been found in the absence of typical MKD features in patients with retinitis pigmentosa and early-onset ulcerative colitis. Mutations in *MVK* have also been found in patients with the skin diseases disseminated superficial actinic porokeratosis and porokeratosis of Mibelli, cyclic neutropenia, and macrophage activation syndrome, but there is no evidence of decreased mevalonate kinase activity in these patients. Mutations in these diseases may overlap with mutations that may cause MA and HIDS.

Periodic Fever, Aphthous Stomatitis, Pharyngitis, and Adenitis Syndrome

PFAPA is primarily a childhood disease with a usual onset before the age of 5 years. Patients suffer from recurring episodes of fever that generally last for 3 to 6 days and recur with great regularity. Additional symptoms include pharyngitis, cervical adenitis, and aphthous stomatitis. Other symptoms may include headache, vomiting and mild abdominal pain, arthralgia, and myalgia. Between fever episodes, patients are symptom free. In most patients, attacks cease after several years, often before or during adolescence.

Schnitzler Syndrome

A typical feature of Schnitzler syndrome is its late onset, at a median age of 51 years. Patients typically present with chronic

TABLE 37.2 Diagnostic Criteria for Schnitzler Syndrome[a] and PFAPA[a]

Schnitzler Syndrome

Major Criteria (≥1 present)

(Chronic) urticarial rash

Monoclonal immunoglobulin M (IgM; or IgG: variant type)

Minor Criteria (≥2 present)

Intermittent fever

Arthralgia or arthritis

Bone pain

Lymphadenopathy

Hepatomegaly and/or splenomegaly

Elevated erythrocyte sedimentation rate (ESR) and/or leukocytosis

Bone abnormalities (on radiologic or histologic examination)

Periodic Fever, Aphthous Stomatitis, Pharyngitis, and Adenitis Syndrome

I Regularly recurring fevers with an early age of onset (<5 years of age)

II Constitutional symptoms in the absence of upper respiratory infection, which include at least one of the following:
- (a) Aphthous stomatitis
- (b) Cervical lymphadenitis
- (c) Pharyngitis

III Exclusion of cyclic neutropenia

IV Completely asymptomatic interval between episodes

V Normal growth and development

[a]Schnitzler syndrome and PFAPA can be diagnosed only after exclusion of other causes.

PFAPA, Periodic fever, aphthous ulcers, pharyngitis, and adenitis syndrome.

de Koning HD, Bodar EJ, van der Meer JW, et al. Schnitzler syndrome: beyond the case reports: review and follow-up of 94 patients with an emphasis on prognosis and treatment. Semin Arthritis Rheum. 2007 Dec;37(3):137–48 and Thomas KT, Feder jr HM, Lawton AR et al. Periodic fever syndrome in children. J Pediatr. 1999 Jul;135(1):15–21

recurrent and mostly nonpruritic urticarial rash. This can be accompanied by fever, arthralgia or arthritis, and bone pain (Table 37.2). Symptoms progress over years. The presence of monoclonal paraproteinemia, typically of immunoglobulin M (IgM), is characteristic for Schnitzler syndrome. Presence of monoclonal IgG is less common and is sometimes referred to as *variant Schnitzler syndrome*. Onset of symptoms can precede paraproteinemia for years, but only little is known about this. Symptom severity is unrelated to the level and type of paraproteinemia.

An important long-term complication of Schnitzler syndrome is Waldenström macroglobulinemia, which, in a single study, had an incidence of 15% 10 years after diagnosis. Patients with Schnitzler syndrome have a normal life expectancy.

PATHOGENESIS

The common pathophysiologic feature of most autoinflammatory diseases is overproduction of the proinflammatory cytokine IL-1β. This protein is produced as an inactive proform (pro–IL-1β), which must be cleaved to become activated. The most common pathway of cleavage is by caspase-1. Like IL-1β, caspase-1 is transcribed as an inactive proform (procaspase-1), and it, too, must be cleaved by a multiprotein complex called the *inflammasome*. Several inflammasomes have been identified, of which the nucleotide-binding oligomerization domain (NOD)-like receptor P3 (NLRP3) inflammasome has been studied in greatest detail.

The NLRP3 inflammasome consists of the central protein NLRP3, the adapter protein apoptosis-associated speck-like protein containing a C-terminal caspase recruitment domain (ASC), and the effector protein procaspase-1. Upon activation of the inflammasome, procaspase-1 is converted into mature caspase-1, which is then able to cleave inactive pro–IL-1β to its active form (Fig. 37.2A).

Familial Mediterranean Fever

FMF is caused by mutations in *MEFV*, which encodes the protein pyrin, primarily expressed in peripheral blood leukocytes, especially neutrophils and monocytes. Pyrin is a member of the pyrin-domain (PYD)-containing proteins, which are able to bind to the PYD domain of other proteins, including apoptosis-associated speck-like protein containing a C-terminal caspase recruitment domain. Binding of pyrin to ASC leads to activation of ASC, with consequent recruitment and activation of procaspase-1. Pyrin can also bind to the PYD domains of other proteins that are able to initiate apoptosis or activate nuclear factor (NF)-κB, including caspase-8 (Fig. 37.2B). These complexes are called *pyrin inflammasomes*.

So far, more than 370 sequence variants in the *MEFV* gene have been reported in the central online Infevers registry (https://infevers.umai-montpellier.fr/web/), most of which are clustered in exon 10 of the gene. The six most prevalent mutations (M694V, V726A, M680I, M694I, V694I, E148Q) cause approximately 80% of cases. Changes in *MEFV* can also be found in other inflammatory diseases including periodic fever with autoinflammation and neutrophilic dermatosis, chronic nonbacterial osteomyelitis, and livedoid ulcerative dermatitis. It has been proposed that *pyrin-associated autoinflammatory diseases* may be used as the common term to refer to all *MEFV*-related diseases.[8]

Cryopyrin-Associated Periodic Syndrome

CAPS is caused by mutations in the gene encoding NLRP3. Previous to its discovery at the beginning of this century, the protein was unknown. It was named *cryopyrin*, in analogy with pyrin in FMF and to illustrate the influence of cold exposure in some patients with CAPS. Literature on this gene can be confusing because the gene previously has also been referred to as *NALP3, PYPAF1*, and *CIAS1*. Mutations associated with CAPS are gain-of-function (GOF) mutations, leading to increased NLRP3 activity. To better reflect the current genetic nomenclature and because periodic fever may not always be present, an international expert group has proposed to rename CAPS to *NLRP3-associated autoinflammatory disease* (NLRP3-AID).[8]

Tumor Necrosis Factor Receptor–Associated Periodic Syndrome

Mutations in the gene *TNFRSF1A* are responsible for TRAPS. This gene encodes the TNF-receptor superfamily 1A (TNFRSF1A), the main cell surface receptor for TNF. This receptor consists of three domains: an extracellular ligand-binding domain, a transmembrane domain, and an intracellular effector domain. So far, more than 170 TNFRSF1A sequence variants have been described, and all TRAPS-associated mutations are located within the extracellular domain of the protein. Upon ligand binding by the extracellular receptor domain, the TNFR forms trimers, triggering the recruitment of intracellular adaptor proteins, which initiate a downstream signaling cascade, leading to NF-κB and mitogen-activated protein kinase (MAPK) activation and caspase-induced apoptosis. When the receptor is activated, the extracellular domain of the TNFR is shed from the membrane. These shed receptors form

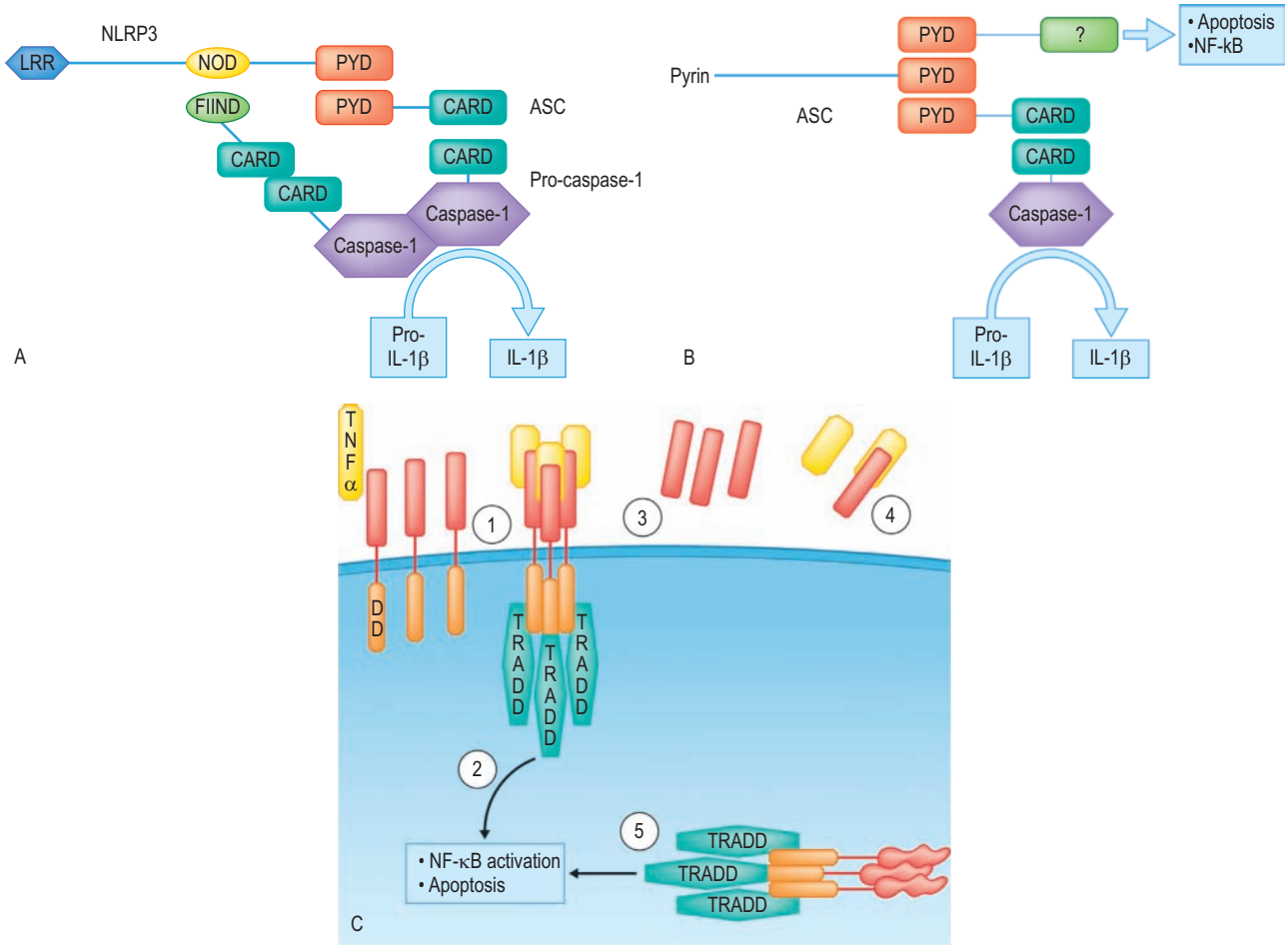


FIG. 37.2 (A) Pathophysiologic mechanisms of classic monogenic autoinflammatory diseases. The nucleotide-binding oligomerization domain *(NOD)*-like receptor P3 *(NLRP3)* Inflammasome. NLRP3 is the central component of this inflammasome. NLRP3 contains three domains: a pyrin domain *(PYD)*, NOD, and a domain of leucine-rich repeats *(LRR)*. Cryopyrin binds apoptosis-associated speck-like protein containing a C-terminal caspase recruitment domain *(ASC)* through its PYD domain and through the NOD domain. The association of these proteins ultimately leads to the release of active caspase-1, which, in turn, activates interleukin-1β *(IL-1β)* through the cleavage of pro–IL-1β. (B) Mechanism of action of pyrin. Pyrin contains a pyrin domain *(PYD)* that is able to bind to apoptosis-associated speck-like protein containing a C-terminal caspase recruitment domain *(ASC)*. ASC can recruit caspase-1 via its *CARD* domain, leading to production of mature interleukin-1β *(IL-1β)*. Pyrin is also able to bind to the PYD domain of other proteins involved in inflammation and apoptosis. (C) Pathophysiology of tumor necrosis factor (*TNF)* receptor–associated periodic syndrome (TRAPS). (1) TNF binds to the TNF receptor on the surface of inflammatory cells (2). After receptor triggering, TNF receptor type 1–associated DEATH domain *(TRADD)* is recruited, inducing a signaling cascade leading to apoptosis and production of proinflammatory cytokines (3). Receptors are shed from the surface, leading to a pool of receptors that dampen immune responses (4). Mutated TNF receptors form aggregates and are retained intracellularly. These aggregated receptors are capable of binding TRADD (5) and stimulate ligand-independent cytokine production.

an extracellular pool of soluble TNFRs, retain their affinity for binding TNF, and are therefore able to mitigate the immune response. Initially, it was hypothesized that TRAPS-associated mutations would lead to defective shedding of TNFR1 receptors, but this hypothesis was discarded because the major pathogenetic mechanism for TRAPS after in vitro experiments showed misfolding and intracellular accumulation of mutated proteins. These aggregated receptors retain their normal signaling function and can induce ligand-independent MAPK signaling and production of reactive oxygen species (ROS), resulting in inflammation (Fig. 37.2C).

There are two exceptional *TNFRSF1A* mutations: R121Q (previously referred to as R92Q) and P75L (previously P46L). These mutations do not lead to receptor misfolding and are present in low frequency in the general population. They may cause a mild inflammatory phenotype.

Mevalonate Kinase Deficiency

The genetic defect in MKD is located in *MVK*. Mevalonate kinase is a key enzyme in the isoprenoid pathway and is located directly downstream from 3-hydroxy-3-methylglutaryl-coenzyme A reductase (HMG-CoA-reductase). The end products of the mevalonate kinase pathway are cholesterol and a number of nonsterol isoprenoids, which are essential compounds in various cellular functions. Mutations in *MVK* lead to reduced mevalonate kinase enzyme activity. In patients with mild disease, residual mevalonate kinase activity is generally 5% to 15% of healthy controls and is even lower in patients with the severe phenotypes.

The mechanistic link between reduced mevalonate kinase activity and autoinflammation is thought to be defective protein prenylation. Prenylation is a posttranscriptional modification, in which nonsterol isoprenoids are coupled to proteins, influencing protein-protein and protein-membrane interactions. Several hypotheses on the effect of defective prenylation in the pathogenesis of MKD exist. Defective prenylation of RhoA, with consequent activation of Rac1 and PKB, may lead to IL-1β secretion with formation of an NLRP3-dependent inflammasome, or to instable mitochondria, which are inadequately cleared from the cytosol with consequent production of ROS, activating NLRP3.[9–11] The MVK pathway is involved in the metabolic introduction of trained immunity and this may explain, at least in part, the constitutively enhanced cytokine production by monocytes of these patients.[12]

Periodic Fever, Aphthous Stomatitis, Pharyngitis, and Adenitis Syndrome

Little is known about the pathophysiology of PFAPA. No genetic defect for PFAPA has been discovered, and this is in agreement with the absence of a clear hereditary pattern. It may be linked to a complex genetic trait. A positive family history has been described, although not all of the patients with a family history were screened for other autoinflammatory diseases.

During PFAPA flare-ups, upregulation of complement genes and genes in the interferon (IFN)–IL-1 pathway are seen. Isolated peripheral blood mononuclear cells and monocytes of patients with PFAPA show increased IL-1β production without induction of transcription of IL-1β RNA or caspase-1 induction upon lipopolysaccharide stimulation. This increased inflammatory response can be abolished by a pan-caspase inhibitor, indicating the important role of the inflammasome in this disease.[13]

Schnitzler Syndrome

The etiology of Schnitzler syndrome remains unknown. A central role for IL-1β is illustrated by the high efficacy of anti–IL-1β therapy in patients with Schnitzler syndrome.[14] No causative genetic defect for Schnitzler syndrome has been found. Somatic mosaicism may be an explanation for the late onset of Schnitzler syndrome, and low-grade mosaicism may not be picked up by routine gene sequencing.[15] Several genetic defects have also been described in very small numbers of patients, but their role in the pathogenesis of Schnitzler syndrome remains unclear.

LABORATORY TESTS

In classic monogenetic autoinflammatory diseases, an explicit acute phase response, with elevated inflammatory markers (C-reactive protein [CRP], erythrocyte sedimentation rate [ESR], serum amyloid A [SAA]) and leukocytosis, is invariably present during symptomatic periods. In other autoinflammatory diseases, this may be less evident.

During clinical disease remission, persistent subclinical inflammation can be found. Cold agglutinins, antinuclear autoantibodies (ANAs), or cryoglobulins are usually not present but may be found in low titers. Proteinuria (>0.5 g of protein per 24 hours) is highly suggestive for secondary amyloid A (AA) amyloidosis.

Elevated serum IgD can be present in MKD. It is discussed in detail in the respective section on the diagnosis of this disease. Similarly, the paraproteins encountered in Schnitzler syndrome are discussed in the section on the symptoms of Schnitzler syndrome.

TABLE 37.3 Tel Hashomer Criteria for the Diagnosis of Familial Mediterranean Fever

Major Criteria (≥1 present)
Typical attack[a] with abdominal symptoms
Typical attack[a] with pleural symptoms
Typical attack[a] with monoarthritis
Typical attack[a] with only fever
Incomplete attack[b] with abdominal symptoms
Minor Criteria (≥2 present)
Favorable response to colchicine
Incomplete attack with monoarthritis
Exertional leg pain

[a]Typical attacks are defined as at least three attacks with fever greater than 38°C.
[b]Incomplete attacks are painful and recurrent attacks not meeting the criteria for typical.
The sensitivity and specificity of these criteria for the diagnosis of FMF are greater than 95% and greater than 97%, respectively.
From Livneh A, Langevitz P, Zemer D, et al. Criteria for the diagnosis of familial Mediterranean fever. *Arthritis Rheum.* 1997;40:1879–1885.

DIAGNOSIS

Familial Mediterranean Fever

FMF is a clinical diagnosis. A validated set of clinical criteria for the diagnosis of FMF has been called: the Tel Hashomer criteria (Table 37.3). These criteria have high positive predictive value and negative predictive value in populations with high pretest probability, but their diagnostic accuracy is lower in other populations. In 2015 it was proposed to replace these criteria, which have been in use since 1997, by a new set of criteria (Eurofever criteria, Table 37.4).

For experienced physicians, it is not difficult to diagnose FMF in patients who have a medical history compatible with FMF and who come from an ethnic group with high prevalence of FMF. In countries with a low incidence of FMF, several years of diagnostic delay are not unusual, as typical attacks remain unrecognized.

When FMF is suspected on clinical grounds, treatment with colchicine should be initiated immediately. A positive effect of colchicine is confirmatory for the diagnosis FMF.

In populations with a low prevalence of FMF and in atypical cases, sequencing of the *MEFV* gene can be helpful in the diagnostic work-up.

Cryopyrin-Associated Periodic Syndrome

CAPS is diagnosed on the basis of typical clinical features (Table 37.5), sometimes supported by a positive family history reflecting autosomal dominant inheritance. Detection of mutations in the *NLRP3* gene will confirm the diagnosis in most cases, but cases with "mutation-negative" CAPS have been described. Some of these patients may have somatic mosaicism. Reaction to treatment is the same for these patients as for patients with a proven *NLRP3* mutation.

Tumor Necrosis Factor Receptor–Associated Periodic Syndrome

A new set of clinical criteria for TRAPS has recently been proposed (see Table 37.4). The cornerstone of the diagnosis of TRAPS is detection of mutations in the *TNFRSF1A* gene.

TABLE 37.4 Diagnostic Criteria for Familial Mediterranean Fever, Tumor Necrosis Factor–Associated Periodic Syndrome, and Mevalonate Kinase Deficiency

FMF		TRAPS		MKD	
Presence	**Score**	**Presence**	**Score**	**Presence**	**Score**
Episode duration <2 days	9	Periorbital edema	21	Age <2 years at onset	10
Chest pain	13	Duration of episodes >6 days	19	Aphthous stomatitis	11
Abdominal pain	9	Centrifugal migratory erythematous patches, most typically overlying a local area of myalgia, usually on limbs of trunk	18	Generalized lymphadenopathy or splenomegaly	8
Turkish, Armenian, non-Ashkenazi Jewish, Arabian ancestry	22	Myalgia	6	Painful lymph nodes	13
Spanish, Italian or Greek ancestry	7	Relatives affected	7	Intermittent diarrhea (sometimes/often)	20
				Persistent diarrhea (always)	37
Absence	**Score**	**Absence**	**Score**	**Absence**	**Score**
Aphthous stomatitis	9	Vomiting	14	Chest pain	11
Urticarial rash	15	Aphthous stomatitis	15		
Cervical lymphadenopathy	10				
Episode duration >6 days	13				
Cutoff value	≥60	Cutoff value	≥43	Cutoff value	≥42

For FMF: A cutoff of ≥60 points using these criteria has a sensitivity of 94% to 97% and a specificity of 91% to 98%.
For TRAPS: A cutoff of ≥43 has a sensitivity of 80% to 85% and a specificity of 87% to 91%.
For MKD: A cutoff of ≥42 has a sensitivity of 89% to 93% and a specificity of 89% to 92%.
FMF, Familial Mediterranean fever; *MKD*, mevalonate kinase deficiency; *TRAPS*, tumor necrosis factor–associated periodic syndrome.
Adapted from Federici S, Sormani MP, Ozen S, et al. Evidence-based provisional clinical classification criteria for autoinflammatory periodic fevers. *Ann Rheum Dis.* 2015;74(5):799–805.

TABLE 37.5 Diagnostic Criteria for Cryopyrin-Associated Periodic Syndrome

Mandatory criterion: Recurrent elevated inflammatory markers (CRP/SAA)
AND
≥2 of the following:
- Urticaria-like rash
- Episodes triggered by cold or stress
- Sensorineural hearing loss
- Musculoskeletal symptoms (arthralgia/arthritis/myalgia)
- Chronic aseptic meningitis
- Bone abnormalities (epiphyseal overgrowth/frontal bossing)

CRP, C-reactive protein; *SAA*, serum amyloid A.
From Kuemmerle-Deschner JB, Ozen S, Tyrrell PN, et al. Diagnostic criteria for cryopyrin-associated periodic syndrome (CAPS). *Ann Rheum Dis.* 2017;76(6):942–947.

Mevalonate Kinase Deficiency

MKD can be suspected when characteristic clinical findings are present (see Table 37.4) in combination with persistent elevated serum levels of IgD greater than 100 international unit per milliliter (IU/mL). Elevation of serum IgD is not pathognomonic for MKD because it may also occur in other inflammatory conditions, including FMF and PFAPA. Furthermore, in the very young, IgD may be normal, and in some of the affected individuals, IgD is never elevated. Elevated IgD is accompanied by elevated IgA in 80% of patients. The serum level of IgD does not correlate with disease severity, and elevated concentrations are present during asymptomatic periods.

During attacks, traces of mevalonic acid may be found both in urine and serum and can be measured with special techniques, which are, however, not commonly available. Measurement of mevalonate kinase enzyme activity is usually only done in the research setting and requires cell cultures.

Clinical suspicion of MKD can be confirmed by sequencing of the *MVK* gene.

Periodic Fever, Aphthous Stomatitis, Pharyngitis, and Adenitis Syndrome

There is currently no diagnostic test to prove PFAPA. It is diagnosed based on clinical signs and symptoms. The modified criteria by Thomas et al. are used (see Table 37.2) and may be supplemented by numeric limits, such as minimum number of attacks and duration of fever. Exclusion of other causes, including other autoinflammatory syndromes, is important.

Schnitzler Syndrome

Schnitzler syndrome is diagnosed on the basis of clinical criteria (see Table 37.2). Exclusion of other causes, particularly monoclonal gammopathy of undetermined significance (MGUS) and chronic idiopathic urticaria, is of paramount importance.

Autoinflammation of Unknown Origin

Regardless of the increasing number of autoinflammatory diseases recognized and the increasing insights into the mechanisms of autoinflammation, an increasing number of patients presenting with an autoinflammatory phenotype cannot be assigned to one of the known autoinflammatory diseases. These patients are considered to suffer from *autoinflammation of unknown origin*. It is likely that further research and newly developed diagnostic techniques will identify new proteins, genetic defects, and pathways in these patients, leading to recognition of new diseases. In patients with autoinflammation of unknown origin, anti–IL-1 therapy can be tried both diagnostically and therapeutically.[16]

TREATMENT

Colchicine

Colchicine is the treatment of first choice for FMF. Although it has been used since the 1970s, the mechanism of action of colchicine in FMF is still unknown. It is highly effective in

preventing attacks. Response to colchicine has been used as a diagnostic criterion for FMF.

The average dose used is 1.0 to 1.5 mg/day. If tolerated, the dose can be increased up to 3 mg/day in patients with insufficient response. There are few patients with FMF who are unresponsive to colchicine. Others may not be able to tolerate an effective dose of colchicine because of side effects. These patients may benefit from IL-1 inhibition.[17]

All patients with FMF should be prescribed colchicine, regardless of disease severity and attack frequency. When anti–IL-1 treatment is started, expert opinion recommends continuation of colchicine in the highest tolerable dose for the prevention of amyloidosis, as it has not been proven that anti–IL-1 monotherapy fully protects against secondary amyloidosis, even when there is complete suppression of inflammation.

The most common side effects of colchicine are diarrhea and abdominal pain. These are dose dependent. In patients with persistent diarrhea, dose reduction can be tried to reduce severity. Myopathy, neuropathy, and leukopenia are very rare, with severe side effects occurring primarily in patients with abnormal kidney or liver function or because of interaction with other drugs (*e.g.*, CYP3A4 inhibitors).

High-dose colchicine has been shown to be teratogenic in animals. However, multiple cohort studies have shown that colchicine can be used safely during pregnancy and breastfeeding. In therapeutic doses, colchicine does not have a negative effect on sperm number and quality, and it has no negative effect on male or female fertility.

As a general rule, there is no place for colchicine in the treatment of autoinflammatory syndromes other than FMF. As an exception to this rule, patients with autoinflammation of unknown origin may experience a favorable effect, especially if their disease shares characteristics with FMF or Behçet disease. A small trial has demonstrated decreased attack frequency during colchicine treatment for PFAPA.[18]

Inhibition of Interleukin-1

KEY CONCEPTS

Interleukin-1β

- Interleukin-1β (IL-1β) is a very potent proinflammatory cytokine.
- Many autoinflammatory diseases are caused by dysregulation of IL-1β.
- Measurement of IL-1 β serum levels has no value in the diagnosis of autoinflammation or in assessing disease severity.
- IL-1 inhibition is the treatment of first choice for many autoinflammatory diseases, and it has greatly improved quality of life in patients.

The detection of *NLRP3* mutations in CAPS has illustrated the importance of IL-1β in the pathogenesis of autoinflammation.

The first inhibitor of IL-1 developed was the recombinant human IL-1 receptor antagonist anakinra, which is still the most commonly available IL-1 inhibitor. It competitively binds to the IL-1 receptor, completely inhibiting the actions of both IL-1α and IL-1β. Anakinra has a short half-life and needs to be given as a once-daily subcutaneous injection.

The selective anti–IL-1β monoclonal antibody canakinumab has a longer half-life and is also injected subcutaneously. Standard injection frequency is once per 8 weeks, but shorter intervals may be necessary in severe disease. In cases with adequate disease control, longer intervals may suffice.

Rilonacept is a construct of two extracellular chains of the IL-1 receptor complex fused to the Fc-portion of IgG. It is given as a weekly subcutaneous injection.

Currently, there is evidence for the effectiveness of anti–IL-1 therapy in many autoinflammatory diseases, including FMF, CAPS, TRAPS, MKD, PFAPA, and Schnitzler syndrome. Typically, anti–IL-1 treatment leads to instant abortion of inflammation with clinical response within the first hours to days after the first injection. This vivid response to treatment is so characteristic that IL-1 inhibition can serve as a diagnostic tool in the diagnosis of these autoinflammatory disorders. Anakinra and canakinumab have both been approved for the treatment of CAPS by the US Food and Drug Administration (FDA) and the European Medical Association (EMA). These agents were given "orphan" status for the treatment of TRAPS by the EMA. Canakinumab is effective in HIDS, TRAPS, and colchicine-resistant FMF.[19] In 2016 the FDA and EMA approved the use of canakinumab in patients with these diseases. Canakinumab is also effective in, but not approved for, the treatment of Schnitzler syndrome.[20] Rilonacept is approved for the treatment of CAPS by the FDA but is not commonly used in Europe.

In patients with mild MKD and periodic symptoms with long symptom-free intervals, anakinra can be used on demand. In these cases it can be started at the first signs of an attack and only be continued for a few days.[21]

Side effects of IL-1 inhibition include painful injection-site reactions, which are most commonly seen with anakinra, and increased frequency of infections, mostly mild upper respiratory tract infections. More severe infections are rarely seen.

Inhibition of Interleukin-6

Tocilizumab, a monoclonal antibody against the IL-6 receptor, has been registered by the FDA and EMA for the treatment of rheumatoid arthritis and systemic-onset juvenile idiopathic arthritis (SoJIA), giant cell arteritis, and chimeric antigen receptor (CAR)-T-cell–induced cytokine release syndrome and is used with increasing frequency in autoinflammatory diseases. It is available as an intravenous infusion or subcutaneous injection. The most common dose is 8 mg/kg in children and adults, or 10 to 12 mg/kg in children with body weight less than 30 kg, with an interval of 2 to 4 weeks. Side effects of tocilizumab are increased susceptibility to infections, most commonly upper respiratory tract infections, elevated liver enzymes, and hematologic abnormalities. Bowel perforations have been reported. As IL-6 induces the production of CRP in the liver, anti–IL-6 therapy always normalizes CRP, making it impossible to use it as a marker for disease activity. Tocilizumab has been shown to be effective in patients with anakinra-resistant Schnitzler syndrome, MKD, and TRAPS.

Inhibition of Tumor Necrosis Factor

Three widely used inhibitors of TNF are infliximab, adalimumab, and etanercept. The most common side effect of TNF inhibition is increased risk of serious infections.

Originally, TNF inhibition was regarded the treatment of first choice in patients with TRAPS unresponsive to nonsteroidal antiinflammatory drugs (NSAIDs). However, TNF inhibition induces complete response only in a minority of patients with TRAPS and is therefore far less effective than anti–IL-1 treatment.[22]

Anti-TNF treatment may also be effective in MKD, but responses are mostly partial. It may be tried in patients who show unsatisfactory response to anti–IL-1 treatment.[22]

Corticosteroids

Corticosteroids are very effective in PFAPA when initiated early in an attack. Prednisone 1 mg/kg is most commonly used during attacks, although lower doses may also be effective. Use of corticosteroids may increase attack frequency in PFAPA.[23] Mild TRAPS can be treated with a short course of steroids (30 mg daily for 7 days), and patients experiencing more severe attacks may respond to higher doses. The beneficial effect of corticosteroids may decrease over time, necessitating dose escalation.[22] Short-term corticosteroids may also be effective in patients with mild MKD.[22]

Simvastatin

The finding of dysregulation of the mevalonate kinase pathway, with cholesterol as its major end product, has led to the hypothesis that blockage of this pathway by HMG-CoA reductase inhibitors (statins) would possible be beneficial in patients with MKD. Small trials and case reports showed a statistically significant, but clinically negligible, effect in some patients. Therefore statins have been abandoned as treatment for MKD in clinical practice.

Other Immunosuppressive Drugs

In the past, numerous immunosuppressive drugs were tried in a trial-and-error approach to find an effective treatment for the disabling symptoms of autoinflammation. The results have been mostly disappointing, and there is no strong evidence to support the use of other immunosuppressants in autoinflammation. When other treatments are ineffective, some patients may, however, benefit from immunosuppressive drugs.

Other Treatments

It remains a matter of debate whether adenotonsillectomy effectively resolves symptoms in PFAPA. A Cochrane review of two unblinded randomized controlled trials showed that children with PFAPA undergoing adenotonsillectomy had a four-time higher chance of becoming symptom free.[24]

Several case reports on the positive effects of hematopoietic stem cell transplantation (HSCT) in patients with severe MKD have been published.[25] Because of the severe side effects, HSCT should be considered the last resort.

AMYLOIDOSIS

Secondary or type AA amyloidosis is a serious complication of all autoinflammatory syndromes. It is caused by tissue deposition of insoluble degradation products of the inflammatory protein SAA. Kidneys are most commonly affected. Because SAA is an acute phase reactant, there is a close relationship between the duration and level of inflammation and the development of amyloidosis. The incidence of AA amyloidosis varies among autoinflammatory diseases. Patients with FMF are at highest risk, with an incidence of up to 75% before the introduction of colchicine treatment for this disease. There is a strong correlation between ethnicity and risk of amyloidosis in FMF, with increased risk among Sephardic Jews.

Up to 25% of patients with TRAPS will develop amyloidosis if left untreated. There seems to be a strong familial predilection. In CAPS, approximately one-third of patients develop amyloidosis in the absence of treatment. Patients with MKD and Schnitzler syndrome have a relatively small risk of amyloidosis, with only a few patients with these diseases and amyloidosis known worldwide.[26]

It is unclear why some patients with the same level of inflammation may develop or never develop amyloidosis. This may be related to single nucleotide polymorphisms (SNPs) in the *SAA* gene or other genotypes.

As proteinuria is often the first sign of AA amyloidosis, patients with autoinflammation should be screened for it with regular urine sampling. Amyloidosis can be confirmed by Congo red staining of the biopsy specimen of affected tissue. This will show an apple-green birefringence under polarized light microscopy (Fig. 37.3). Progression of amyloidosis is strongly dependent on the ability to control the underlying inflammation. If the SAA concentration can be maintained less than 10 mg/L, progression of amyloidosis can be halted in many cases. Some patients even show regression of amyloidosis during treatment.[27] A cohort mainly consisting of patients

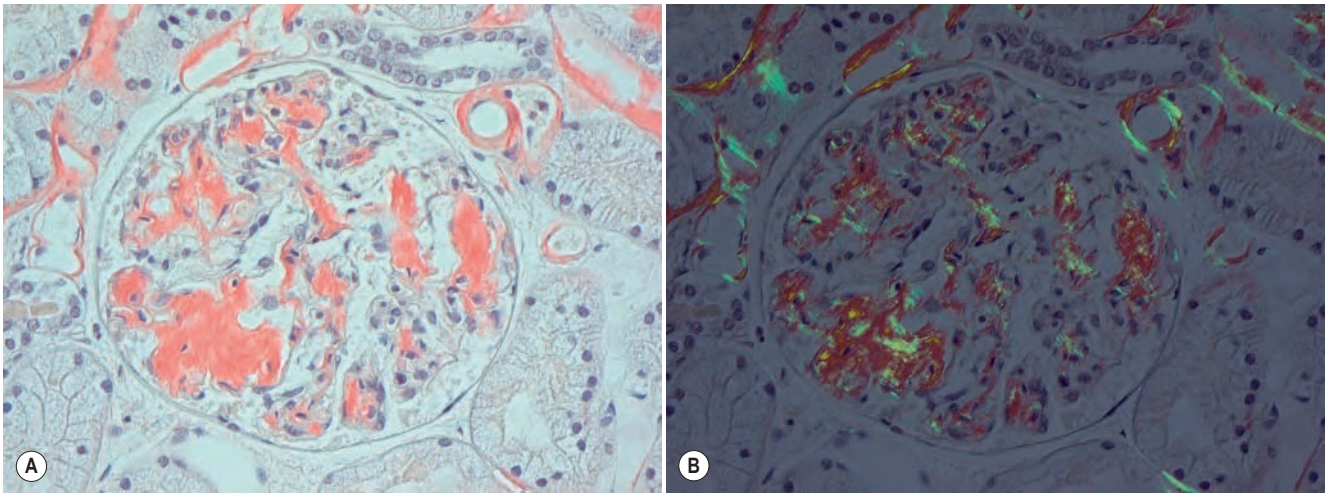


FIG. 37.3 Renal Biopsy from a Patient With Amyloid A Amyloidosis. Amyloid deposits are visualized by staining with Congo *red* (A). Under polarized light microscopy amyloid deposits show typical *apple-green* birefringence (B).

diagnosed with amyloidosis before the availability of targeted therapy showed a median survival of 19 years after the diagnosis of amyloidosis.[28]

CLINICAL PEARLS

A Typical Case of Autoinflammation

- An 18-year-old Western European woman was seen because of recurrent episodes of fever and skin rash. These attacks started 2 hours after birth when a systemic maculopapular skin rash appeared. The rash remained present on a daily basis but was not triggered or exacerbated by cold exposure.
- At the age of 2 years, the patient developed episodes of fever, approximately 3 days each week. Typically, she had a single daily fever spike up to 39°C, which usually occurred in the evening. There was no hearing loss, and arthralgia/arthritis, dysmorphia, meningitis, myalgias, abdominal pain, and lymphadenopathy were also absent. The family history was negative for autoinflammatory diseases.
- On the basis of the clinical presentation, cryopyrin-associated periodic syndrome (CAPS) was suspected as a cause, and targeted gene analysis was performed. This showed an R260W mutation of the *NLRP3* gene. Thus this patient was diagnosed with CAPS.
- After the diagnosis was made, the patient was successfully treated with anakinra, 100 mg once daily. When canakinumab, the long-acting interleukin-1 (IL-1) inhibitor, became available, she was switched to 150 mg canakinumab once every 2 months; she remains symptom-free.

CONCLUSIONS

ON THE HORIZON

- Identification of new (hereditary) autoinflammatory syndromes in patients with episodic inflammation of unknown origin
- Wider availability and application of interleukin-1 (IL-1) inhibitors and development of new drugs to improve treatment.

Since the introduction of the term *autoinflammation* at the end of the 20th century, the clinical characteristics of the classic monogenic autoinflammatory diseases FMF, CAPS, MKD, and TRAPS have been described in more detail. Many new diseases have been identified. Research focused on their pathophysiologic mechanism has revealed new genes and pathways and provided further insight into the mechanism of inflammation in general. The development of IL-1–targeting drugs has led to better quality of life for patients and to increased life expectancy in some of these diseases by prevention of complications, such as secondary amyloidosis. Nevertheless, many patients with autoinflammation remain undiagnosed.

For the near future, further study of the mechanisms of inflammation in patients with autoinflammatory diseases will likely provide a deeper insight into the workings of innate immunity and will lead to the identification of even more diseases in the autoinflammatory spectrum. The search for new therapies, preferably oral drugs, for the treatment of autoinflammation will continue.

REFERENCES

1. Bodar EJ, Drenth JP, van der Meer JW, Simon A. Dysregulation of innate immunity: hereditary periodic fever syndromes. *Br J Haematol*. 2009;144(3):279–302.
2. van der Meer JW, Vossen JM, Radl J, et al. Hyperimmunoglobulinaemia D and periodic fever: a new syndrome. *Lancet*. 1984;1(8386):1087–1090.
3. Houten SM, van Woerden CS, Wijburg FA, et al. Carrier frequency of the V377I (1129G>A) MVK mutation, associated with Hyper-IgD and periodic fever syndrome, in the Netherlands. *Eur J Hum Genet*. 2003;11(2):196–200.
4. Simon A, Mariman EC, van der Meer JW, Drenth JP. A founder effect in the hyperimmunoglobulinemia D and periodic fever syndrome. *Am J Med*. 2003;114(2):148–152.
5. Marshall GS, Edwards KM, Butler J, Lawton AR. Syndrome of periodic fever, pharyngitis, and aphthous stomatitis. *J Pediatr*. 1987;110(1):43–46.
6. Feder HM, Salazar JC. A clinical review of 105 patients with PFAPA (a periodic fever syndrome). *Acta Paediatr*. 2010;99(2):178–184.
7. Schnitzler L. Lésions urticariennes chroniques permanentes (érythème pétaloïde?) Cas cliniques, n° 46 B. *Journée Dermatologique d'Angers*. 1972
8. Ben-Chetrit E, Gattorno M, Gul A, et al. Consensus proposal for taxonomy and definition of the autoinflammatory diseases (AIDs): a Delphi study. *Ann Rheumatic Dis*. 2018;77(11):1558–1565.
9. Mandey SHL, Kuijk LM, Frenkel J, Waterham HR. A role for geranylgeranylation in interleukin-1 beta secretion. *Arthritis Rheum*. 2006;54(11):3690–3695.
10. van der Burgh R, Nijhuis L, Pervolaraki K, et al. Defects in mitochondrial clearance predispose human monocytes to interleukin-1beta hypersecretion. *J Biol Chem*. 2014;289(8):5000–5012.
11. Tricarico PM, Kleiner G, Valencic E, et al. Block of the mevalonate pathway triggers oxidative and inflammatory molecular mechanisms modulated by exogenous isoprenoid compounds. *Int J Mol Sci*. 2014;15(4):6843–6856.
12. Bekkering S, Arts RJW, Novakovic B, et al. Metabolic induction of trained immunity through the mevalonate pathway. *Cell*. 2018;172(1–2) 135–146 e9.
13. Kolly L, Busso N, von Scheven-Gete A, et al. Periodic fever, aphthous stomatitis, pharyngitis, cervical adenitis syndrome is linked to dysregulated monocyte IL-1beta production. *J Allergy Clin Immunol*. 2013;131(6):1635–1643.
14. de Koning HD, Schalkwijk J, van der Meer JW, et al. Successful canakinumab treatment identifies IL-1beta as a pivotal mediator in Schnitzler syndrome. *J Allergy Clin Immunol*. 2011;128(6):1352–1354.
15. de Koning HD, van Gijn ME, Stoffels M, et al. Myeloid lineage-restricted somatic mosaicism of NLRP3 mutations in patients with variant Schnitzler syndrome. *J Allergy Clin Immunol*. 2015;135(2):561–564.
16. Harrison SR, McGonagle D, Nizam S, et al. Anakinra as a diagnostic challenge and treatment option for systemic autoinflammatory disorders of undefined etiology. *JCI Insight*. 2016;1(6):e86336.
17. vd Hilst J, Moutschen M, Messiaen P, et al. Efficacy of anti-IL-1 treatment in familial Mediterranean fever: a systematic review of the literature. *Biol Targets Ther*. 2016;10:75–80.
18. Butbul Aviel Y, Tatour S, Gershoni Baruch R, et al. Colchicine as a therapeutic option in periodic fever, aphthous stomatitis, pharyngitis, cervical adenitis (PFAPA) syndrome. *Sem Arthritis Rheumatism*. 2016;45(4):471–474.
19. De Benedetti F, Gattorno M, Anton J, et al. Canakinumab for the treatment of autoinflammatory recurrent fever syndromes. *N Engl J Med*. 2018;378(20):1908–1919.
20. de Koning HD, Schalkwijk J, van der Ven-Jongekrijg J, et al. Sustained efficacy of the monoclonal anti-interleukin-1 beta antibody canakinumab in a 9-month trial in Schnitzler's syndrome. *Ann Rheumatic Dis*. 2013;72(10):1634–1638.
21. Bodar EJ, Kuijk LM, Drenth JP, et al. On-demand anakinra treatment is effective in mevalonate kinase deficiency. *Ann Rheumatic Dis*. 2011;70(12):2155–2158.
22. ter Haar NM, Oswald M, Jeyaratnam J, et al. Recommendations for the management of autoinflammatory diseases. *Ann Rheumatic Dis*. 2015;74(9):1636–1644.
23. Ter Haar N, Lachmann H, Ozen S, et al. Treatment of autoinflammatory diseases: results from the Eurofever Registry and a literature review. *Ann Rheumatic Dis*. 2013;72(5):678–685.
24. Burton MJ, Pollard AJ, Ramsden JD, et al. Tonsillectomy for periodic fever, aphthous stomatitis, pharyngitis and cervical adenitis syndrome (PFAPA). *Cochrane Database Syst Rev*. 2019;12:CD008669.
25. Mulders-Manders CM, Simon A. Hyper-IgD syndrome/mevalonate kinase deficiency: what is new? *Sem Immunopathol*. 2015;37(4):371–376.
26. de Koning HD, Bodar EJ, van der Meer JW, et al. Schnitzler syndrome: beyond the case reports: review and follow-up of 94 patients with an emphasis on prognosis and treatment. *Sem Arthritis Rheumatism*. 2007;37(3):137–148.
27. van der Hilst JC. Recent insights into the pathogenesis of type AA amyloidosis. *Sci World J*. 2011;11:641–650.
28. Lane T, Loeffler JM, Rowczenio DM, et al. AA amyloidosis complicating the hereditary periodic fever syndromes. *Arthritis Rheum*. 2013;65(4):1116–1121.

38

Immune Regulatory Disorders

Troy R. Torgerson and Alice Y. Chan

Normal immunity requires a delicate balance between immune effector mechanisms and immune regulatory mechanisms. Effector mechanisms are required to prevent, control, and eradicate external threats to the body, while regulatory mechanisms are required to modulate and control the establishment, initiation, intensity, and longevity of those immune effector responses. The first identified inborn errors of immunity (IEI) were associated principally with susceptibility to recurrent, severe, or unusual infections; the genes mutated in these disorders play key roles in the development or function of essential immune effector mechanisms. In contrast, approximately 20 years ago, a second major class of IEI was described in which patients suffer primarily from severe autoimmunity, inflammatory disease, or nonmalignant lymphoproliferation. Some of these patients have recurrent or unusual infections, but most of their clinical features were the result of immune-mediated pathology. Defective genes identified in these patients typically encode proteins that play key roles in immune regulatory mechanisms; consequently, these are considered to be disorders of immune regulation.

There are now over 430 genes that have been associated with IEI.[1] Of these, approximately 30% have clinical features that would classify them predominantly as immune regulatory disorders. These disorders cluster into groups characterized by similar clinical presentations and, in most cases, by genetic defects that often impact one or more related immune regulatory proteins or signaling pathways (see Table 38.1 for a list of these groups and specific examples). Among the disease groups listed in Table 38.1, hemophagocytic lymphohistiocytosis (HLH) and autoinflammatory disorders are covered in Chapters 36 and 37, respectively.

CLINICAL FEATURES OF IMMUNE REGULATORY DISORDERS

Autoimmunity, inflammatory disease, and nonmalignant lymphoproliferation are common clinical features of immune regulatory disorders. Each genetic disorder tends to have a characteristic combination of clinical symptoms affecting particular organ systems, but there are common types of organ-specific manifestations that tend to occur in multiple disorders. By organ system, clinical symptoms that are frequently observed, beginning with the most common, are: (i) *hematologic*—autoimmune cytopenias, anemia, thrombocytopenia, and neutropenia, often with detectable autoantibodies; (ii) *gastrointestinal* (GI; autoantibodies directed toward gut or liver antigens may or may not be present)—autoimmune or inflammatory enteropathy, lymphoproliferation in the gastrointestinal (GI) tract (nodular lymphoid hyperplasia or lymphocytic colitis), and autoimmune or inflammatory hepatitis; (iii) *skin*—various types of dermatitis, most commonly eczema and psoriasis; (iv) *lungs*—interstitial lung disease, follicular bronchiolitis, and non-caseating granulomas; (v) *endocrine*—thyroiditis, type 1 diabetes, adrenal insufficiency, gonadal insufficiency, and other endocrine disorders, commonly with autoantibodies against thyroid (thyroglobulin, thyroid peroxidase), pancreas (GAD65, insulin, ZnT8, islet cells), adrenal (21-hydroxylase), with other endocrine autoantibodies often present and even detectable well before development of clinical disease; (vi) *renal*—nephritic, nephrotic, or tubulointerstitial disease; (vii) *cardiovascular*—vasculitis; and (vii) *musculoskeletal*—arthritis or myositis.

Many of these clinical manifestations may be present in more common autoimmune or inflammatory diseases like systemic lupus or inflammatory bowel disease (IBD). There are, however, features that should raise suspicion that a patient may have an IEI with immune dysregulation, including: (1) *Early-onset disease*. Onset of clinically significant autoimmune, inflammatory, or lymphoproliferative disease early in life (i.e., under the age of 5 years) should raise suspicion for an IEI disorder. (2) *Unusual combinations of clinical features*: for example, type 1 diabetes or interstitial lung disease paired with IBD. While the incidence of type 1 diabetes is modestly higher in patients with IBD, this combination is not common so should raise suspicion for a potential underlying IEI disorder. Similarly, the combination of infections with unusual pathogens or atypically severe infections in a patient with immune dysregulation, particularly one not receiving significant immunosuppressive therapy, should raise suspicion for an underlying IEI. (3) *Disease that is unusually extensive*: for example, the presence of interstitial lung disease with follicular bronchiolitis, enteropathy with nodular lymphoid hyperplasia throughout the bowel, type 1 diabetes, psoriasis, psoriatic arthritis, and hypogammaglobulinemia in a single patient.

THE GENETICS OF IMMUNE REGULATORY DISORDERS

The number of different gene defects associated with immune regulatory disorders has grown steadily since 2010. All modes of inheritance, including X-linked recessive, autosomal recessive, and autosomal dominant have been observed, but many of the single gene disorders identified since 2015 demonstrate autosomal dominant inheritance. Despite segregating as dominant traits, the mechanisms of disease among these autosomal

TABLE 38.1 Immune Regulatory Disease Phenotypes and Genes

Disease Class	Disease Entities	Common Clinical Manifestations	Example Genes (Protein)[a]
Treg/Teff Cell Axis	IPEX IPEX-like APECED	Enteropathy (diarrhea) Dermatitis (eczema, psoriasis) Cytopenias Endocrinopathies (T1DM, thyroiditis, etc.)	FOXP3, CD25, CTLA4, LRBA, DEF6, IL2RB, STAT1(GOF), STAT3(GOF), STAT5B (LOF or GOF)
Nonmalignant Lymphoproliferation	ALPS ALPS-like ALPS-U RALD	Lymphadenopathy Splenomegaly Recurrent cytopenias Leukoproliferation Fever Malignancy (lymphoma)	FAS, FASL, CTLA4, LRBA, DEF6, PIK3CD, PIK3R1 STAT3(GOF), TNFRSF9, TET2, CASP10, KRAS, NRAS
Autoinflammatory Syndromes	TRAPS CAPS FMF DADA2 DIRA CRIA	Recurrent fevers Rash (various types) Abdominal and/or chest pain Joint/muscle pain and/or swelling	TNFRSF1A, TNFRSF11A, CDC42, NLRP3, MEFV, ADA2, IL1RN, RIPK1
Debris Defects	Complement deficiency Interferonopathies PRAAS	Glomerulonephritis Lupus (systemic, chilblain. etc.) Aicardi-Goutieres syndrome (basal ganglia calcifications) Vasculitis	C1q, C2, COPA, DNASE1, IFIH1, STING, TREX1, PSMB8
Hyperinflammatory Disorders	fHLH	Fever Cytopenias Hepatosplenomegaly, lymphadenopathy Rash Elevated ferritin, triglycerides, and sCD25 Low fibrinogen, DIC	LYST, PRF1, RAB27A, UNC13D, CDC42, ITK, MAGT1, SH2D1A, STAT1(GOF), NLRC4, XIAP
Hematopoietic Malignancies	SPTCL LGL DLBCL AML / MDS	Susceptibility to malignancy (various)	HAVCR2 (TIM3) STAT3-GOF CARD11, BCL10, MALT1 PIK3CD GATA2 TET2
Congenital Atopic Hypersensitivity		Allergic disease	JAK1(GOF), PGM3, STAT5B(GOF)
Inflammatory Bowel Disease (IBD)	Infantile-onset IBD VEO-IBD		CYBB, IL10, IL10RA, IL10RB, SKIV2L, TTC7A, CYBB, RIPK1, NEMO
Rheumatologic Diseases	Bechet disease Lupus JIA		TNFAIP3, WDR1 C1q, C2 FAS, STAT3-GOF, NCKAP1L (HEM1), LACC1

[a]Genes listed together function in the same pathway or complex.

ALPS, Autoimmune lymphoproliferative syndrome, *AML*, acute myeloid leukemia, *CAPS*, cryopyrin-associated autoinflammatory syndromes, *CRIA*, cleavage-resistant RIPK1-induced autoinflammatory, *DADA2*, deficiency of adenosine deaminase 2, *DIRA*, deficiency of the interleukin-1 receptor antagonist, *DLBCL*, diffuse large B-cell lymphoma, *Fhlh*, familial hemophagocytic lymphohistiocytosis, *FMF*, familial Mediterranean fever, *GOF*, gain of function, *IPEX*, immune dysregulation, polyendocrinopathy, enteropathy, X-linked, *JIA*, juvenile idiopathic arthritis, *LGL*, large granular lymphocytic leukemia, *LOF*, loss-of-function, *MDS*, myelodysplastic syndrome, *PRAAS*, proteasome-associated autoinflammatory syndrome, *RALD*, RAS-associated autoimmune leukoproliferative disease, *SPTCL*, subcutaneous panniculitis-like T-cell lymphoma, *TRAPS*, tumor necrosis factor receptor associated periodic syndrome, *VEO-IBD*, very early onset inflammatory bowel disease.

dominant disorders vary from haploinsufficiency to dominant loss-of-function (LOF) to dominant gain-of-function (GOF). To complicate matters more, many autosomal dominant immune dysregulation disorders exhibit variable expressivity of the associated clinical phenotype, meaning that some patients with the same mutation may have little to no clinical disease while others have severe disease. This creates challenges when trying to counsel individuals carrying disease-associated mutations about their risks of the disease when making decisions about treatment.

Somatic or Mosaic Genetic Defects and Immune Dysregulation

One unique aspect of immune dysregulation disorders is that acquired (somatic) mutations can cause clinical disease due to the fact that these mutations often activate or enhance the growth, survival, or reactivity of the cells in which they occur. In many cases, these somatic mutations occur in only a subset of cells but can lead to the same phenotype and disease severity as when the mutation is present in the germ line. A growing number of disorders have been described in which a somatic mutation in as few as 5% of cells of a particular cell subset can lead to disease.[1] Depending on when the mutation occurs, the somatic variant can cause either germline or somatic mosaicism. *Germline mosaicism* usually occurs as a result of a mutation arising de novo in one or more cells relatively early after fertilization. As a result, most or all cells arising from that initial cell (including germ cells/gametes) carry the mutation, meaning that the mutation can be passed down to offspring. *Somatic mosaicism* typically occurs in a terminally differentiated cell type at any time during its life span.

Often, somatic mutations lead to a selective growth advantage for the cells that harbor the mutation, leading to their expansion. Mutations that lead to abnormal immune cell function, abnormal growth, resistance to apoptosis, resistance to regulation, etc., can lead to those cells secreting large amounts

of cytokine, growing in an unregulated manner, or inappropriately attacking cellular or protein targets. Because cells that acquire somatic mutations tend to gain enhanced or altered function, they lead almost exclusively to immune dysregulation and almost never to immune deficiency. A growing number of gene defects cause similar clinical disease, regardless of whether they occur as germline or somatic mutations (e.g., *FAS*, *NLRP3*, *NLRC4*, *TNFRSF1A*, *NOD2*, *TMEM173*, *TLR8*).[2,3] Some disorders, such as the RAS-associated autoimmune leukoproliferative disorder (RALD), are caused by mutations in the *NRAS* or *KRAS* genes, and the inflammatory disorder caused by mutations in *STAT5B* have been described exclusively in the setting of somatic mutations. Finally, immune regulatory disorders can also arise when mutations take place in only a subset of immune cells, including cryopyrin-associated periodic syndrome (CAPS) occurring when somatic *NLRP3* mutations arise in only a subset of myeloid cells, or pure red cell aplasia occurring when somatic *STAT3-GOF* mutations take place only in CD8 T cells.[4,5] Unfortunately, identification of somatic mutations can be very difficult, particularly when the mutation is present in only a subset of cells. In these cases, either ultra-deep next-generation sequencing (i.e., sequencing to a depth of 500× coverage or higher) or enrichment of the affected cell subset prior to sequencing may be required.

GENERAL DIAGNOSTIC TESTING APPROACH TO IMMUNE REGULATORY DISORDERS

One of the most challenging aspects of immune regulatory disorders is that there is often significant phenotypic overlap between patients with different genetic defects. Similarities tend to be highest among disorders in the same disease group; the underlying molecular defects often involve related immune mechanisms or signaling pathways (see Table 38.1), but there can also be significant overlap of clinical phenotypes between different disease groups. Previously, the approach to diagnosis varied, but typically depended on clinical laboratory testing paired with flow-cytometry-based testing in order to define an immunophenotype before embarking on targeted genetic testing to try to confirm a diagnosis. Because of the dramatic decline in the cost of genetic sequencing, the most direct way to obtain a definitive diagnosis in most cases is to perform broad-based genetic testing using a large gene panel, whole exome sequencing, or whole genome sequencing. Depending on the cohort, published data suggests that this "genetics first" approach can yield an answer in 30% to 50% of cases.[6–8]

Adjunctive Diagnostic Tests

At the present time, the most useful application of adjunctive diagnostic testing is to supplement or confirm pathogenicity of a genetic variant. Adjunctive testing generally falls into two major categories: protein expression studies and functional assays. A variety of tests have been developed to evaluate the presence and phenotype of various immune cell subsets or the expression of key proteins in cells; examples include B-cell immunophenotyping to assess naïve and memory B-cell populations or tests to evaluate FOXP3 or CTLA4 protein expression in regulatory T cells. Functional tests can be useful to determine whether novel gene variants identified by genetic testing have a functional consequence that may suggest pathogenicity. Examples include evaluation of STAT1 phosphorylation in patients with a suspected STAT1-GOF mutation, evaluation of CD80 transendocytosis/uptake in patients with suspected CTLA4 haploinsufficiency or LRBA deficiency, or assessment of muramyl dipeptide—stimulated TNF production by monocytes in patients with suspected XIAP deficiency.[9,10] Only a handful of these tests are available in clinical labs, while others are available only as research assays. Consequently, these tests are generally not good for screening patients for the presence of disease, but instead are used for follow-up assessments after genetic testing.

Additional Workup

In addition to genetic and adjunctive testing performed to arrive at a definitive diagnosis, assessments to screen for organ involvement are recommended. These include assessments of the hematopoietic system, immune system, GI tract and liver, endocrine organs, lungs, and kidneys. In each case, the goal is to screen for active organ-specific autoimmune disease, which should be considered in all patients with immune regulatory disorders. These should be tailored to each patient's clinical presentation.

A variety of autoantibodies have been identified in patients with immune regulatory disorders; some of these may be pathogenic, so screening should be considered in all cases. Autoantibodies against endocrine organs are quite common, making screening for thyroid autoantibodies recommended; screening for diabetes-associated autoantibodies should be considered, particularly if there are any signs of glucose instability. Investigation of autoantibodies targeting other organs may be guided by the organ-specific clinical manifestations in each patient.

SPECIFIC IMMUNE REGULATORY DISORDERS

Regulatory-T-Cell (Treg)/Effector-T-Cell (Teff) Axis Defects

Regulatory T cells (Tregs) play a major role in maintaining immune tolerance and returning the immune system to homeostasis after an immune activation event. Thymically derived regulatory T cells (tTregs) constitute the majority of Tregs in the resting state. Cells with a tTreg phenotype are first identifiable in the thymic medulla after progenitor T cells have undergone the initial rounds of positive selection in response to engagement of their T-cell receptor by antigen. The developing Tregs are exposed to tissue-specific antigens presented on AIRE-expressing thymic medullary epithelial cells, and their characteristic genetic program is set in place (although the precise mechanisms by which this happens is unknown) (Fig. 38.1). It is known that the transcription factor FOXP3 is required for the development of functional Treg cells. tTregs eventually leave the thymus and enter the periphery, where they require persistent interleukin (IL)-2 stimulation to maintain their FOXP3 expression, expand, and acquire maximal regulatory function. tTregs utilize a variety of mechanisms to regulate target cells, including direct cell-cell interactions via key regulatory molecules like CTLA4, secretion of regulatory cytokines like transforming growth factor-β (TGF-β) and IL-10, competition for IL-2 in the local microenvironment, and others (see Fig. 38.1). Absence of key molecules required for the development or function of Tregs can lead to an imbalance between regulatory and effector T-cell function, thereby causing immune dysregulation and autoimmunity.

Autoimmune regulator (AIRE) deficiency (autosomal recessive/autosomal dominant) causes the syndrome of autoimmune

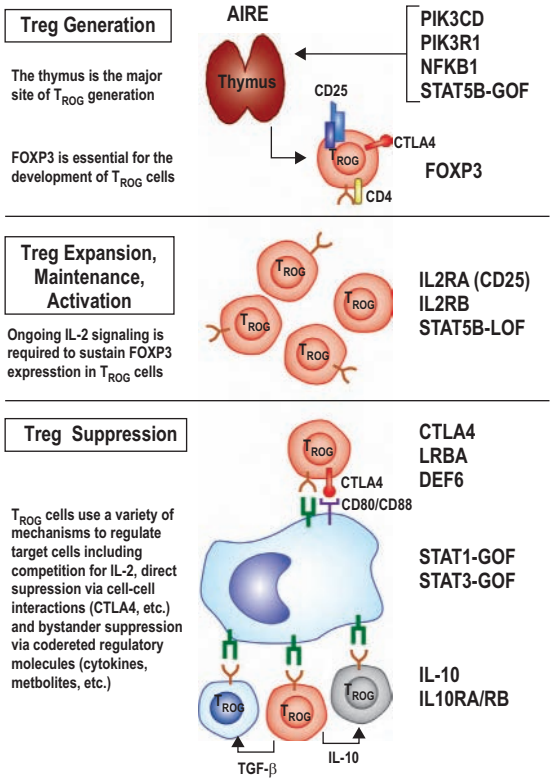


FIG. 38.1 The Normal Life Cycle of Thymically Derived Regulatory T Cells. Cells initially require selection in the thymus and expression of FOXP3 to become functional regulatory T cells (*Tregs*). In the periphery, they require ongoing interleukin (*IL*)-2 stimulation to sustain FOXP3 expression; they expand and become activated to suppress target cells. Functionally active Tregs exert suppressive function on target cells in various ways, including direct cell-cell interaction and secretion of immunoregulatory cytokines. Certain activating mutations in key immune signaling pathways (STAT1, STAT3, etc.) may cause effector cells to be resistant to the suppressive activity of Tregs. *AIRE*, Autoimmune regulator; *TGF-β*, transforming growth factor-β.

polyendocrinopathy, candidiasis, and ectodermal dystrophy (APECED), which is characterized by extensive organ-specific autoimmunity mediated by both cellular and humoral immune mechanisms. Key target tissues include endocrine organs (parathyroid dysfunction, adrenal insufficiency, gonadal failure, type 1 diabetes, etc.), lung (interstitial lung disease), GI tract (enteropathy, pernicious anemia), liver, and skin. Many patients also have chronic mucocutaneous candidiasis (CMC). Recent studies have shown that the CMC is caused primarily by dysregulated type 1 immunity with overexpression of interferon-gamma, leading to compromised mucosal barrier function.[11] Since AIRE is primarily expressed in thymic medullary epithelial cells where it plays a role in thymic T-cell selection, this may be related to defective negative selection of autoreactive T cells. Decreased Treg function may also contribute to the excessive type 1 immunity, as APECED patients demonstrate both a decreased percentage of $CD4^+CD25^{high}$ tTreg cells as well as decreased FOXP3 protein expression in these cells compared to normal individuals.[12] Furthermore, isolated tTreg cells from APECED patients have a decreased ability to suppress proliferation of effector T cells in vitro compared to those from healthy controls.[12] Murine studies suggest that Janus kinase (JAK) inhibitors may be effective in controlling the excessive type 1 immunity in APECED.[11] Since AIRE is expressed primarily in the thymus, hematopoietic cell transplant (HCT) is unlikely to yield significant benefit.

FOXP3 deficiency (X-linked recessive) causes the syndrome of immune dysregulation, polyendocrinopathy, enteropathy, and X-linked (IPEX). This occurs in males with a single mutated X chromosome. Most patients with IPEX present with a basic triad of clinical features, including autoimmune enteropathy (watery diarrhea), dermatitis (usually eczema), and endocrinopathy (type I diabetes and thyroiditis). In addition, most patients have other autoimmune features, including cytopenias, liver disease, nephritis/nephropathy, etc. Since FOXP3 is required for the development of functional Tregs, deficiency causes a lack of effective peripheral immune tolerance. Murine studies have demonstrated that in this setting, effector T cells act as the primary drivers of autoimmunity. The most effective therapies have therefore been those directed toward controlling effector T-cell function, including tacrolimus (FK506), cyclosporine, and sirolimus (rapapmycin). HCT has been shown to be an effective therapy for IPEX, but no particular transplant regimen has proven definitively to be superior as long as the chosen transplant regimen leads to stable donor cell engraftment.[13,14]

IL2RA, IL2RB, and STAT5B deficiency (autosomal recessive) cause defects in IL-2 signaling, either as a result of absent IL-2 receptor chains (IL2RA and IL2RB) or absence of the dominant IL-2 driven transcription factor, STAT5B. Ongoing IL-2 stimulation of Tregs is required to sustain FOXP3 expression, thereby preserving their suppressive function and maintaining their competitive fitness. As a result of Treg dysfunction, humans with defects in IL-2 signaling develop immune dysregulation with autoimmune enteropathy (diarrhea), dermatitis (eczema, pemphigus nodularis, psoriasiform dermatitis), and early-onset endocrinopathies (type 1 diabetes, thyroiditis). Patients with STAT5B deficiency also develop severe, interstitial pneumonitis that can be life threatening, and due to the separate role of STAT5B in signaling from the human growth hormone receptor, patients also exhibit dwarfism. Development of pathogenic autoantibodies is common, and many patients also have hepatosplenomegaly, lymphadenopathy, and lymphocytic infiltrates in various organs. Because IL-2 is essential for expansion of effector T cells, patients with defects in this signaling pathway often have moderate T-cell lymphopenia, and in cases of IL2RB and STAT5B deficiency, may also have low NK cell numbers due to the role that these molecules play in IL-15 signaling. As a result, patients are typically susceptible to viral and fungal pathogens, with cytomegalovirus (CMV) pneumonitis being reported in several patients and thrush and *candida* esophagitis being common. Various immunosuppressive therapies have been tried in patients with IL-2 signaling defects, but treatment with aggressive immunosuppression is complicated due to infectious susceptibilities. HCT has been successful in both IL2RA and IL2RB deficiency, has not been reported in STAT5B deficiency.[15–18]

STAT1-GOF defects (autosomal dominant) are most commonly associated with CMC, but there is a wide range of phenotypic heterogeneity among patients, with some having only intermittent CMC episodes and others with extensive immune dysregulation and autoimmunity. Common features among patients with more extensive autoimmunity include enteropathy (diarrhea), dermatitis (eczema, psoriasiform dermatitis), endocrinopathy

(thyroiditis, type 1 diabetes), and cytopenias. From an infection standpoint, some STAT1-GOF patients have susceptibility to recurrent or severe herpes viral infections, and a subset have hypogammaglobulinemia, poor vaccine responses, and recurrent bacterial sinopulmonary infections, sometimes misdiagnosed as common variable immune deficiency (CVID) before a molecular diagnosis is obtained. Vascular aneurysms have been described in a small number of patients, and many patients with a severe immune dysregulation phenotype have failure to thrive with growth delay. A growing number of case reports have described significant clinical improvement in response to JAK kinase inhibitors (JAKinibs), although there have also been reports of invasive fungal and herpes viral infections in patients receiving this treatment. HCT can be successful, but in the largest published case series, survival was only 40%, with high rates of primary and secondary donor graft failure, suggesting that transplant should be approached with caution until more data are available regarding the best HCT approach.[15]

STAT3-GOF defects (autosomal dominant) are characterized by severe, systemic autoimmunity and marked growth failure in at least 50% of patients. Common autoimmune features include autoimmune enteropathy (diarrhea), early-onset type 1 diabetes, cytopenias, inflammatory liver disease, splenomegaly, and lymphadenopathy, sometimes mistaken for autoimmune lymphoproliferative syndrome (ALPS) prior to obtaining a molecular diagnosis. Other findings that are less common but still serious include hypogammaglobulinemia and recurrent infections, arthritis, and vasculitis. Like patients with STAT1-GOF disease, there are case reports describing significant clinical improvement with the use of JAKinib therapy. Since IL-6 is a major driver of STAT3 activation, IL-6 blockade using monoclonal antibody therapy has also been used, although the clinical effectiveness of this is less clear. HCT has been used successfully, but an early case series suggests that survival may be just over 50%, with many patients dying of severe graft-versus-host disease or infections after transplant.[15]

CTLA4 haploinsufficiency (autosomal dominant) is usually associated with both an immune deficiency (hypogammaglobulinemia and recurrent infections) and profound, systemic autoimmunity. The most common autoimmune features include recurrent, autoimmune cytopenias (hemolytic anemia, thrombocytopenia, neutropenia) and nonmalignant lymphoproliferation in nonlymphoid organs. The most common sites of lymphoid proliferation are the GI tract (nodular lymphoid hyperplasia) and lungs (follicular bronchiolitis, granulomatous lymphocytic interstitial lung disease [GLILD], etc.). Nonmalignant lymphoproliferative lesions in the central nervous system occur in 25% to 30% of patients and are quite likely to involve CTLA4 pathway defects. Less common autoimmune features include endocrinopathies (thyroiditis, type 1 diabetes, etc.), arthritis, and aplastic anemia. In patients with lymphoproliferation and autoimmunity, sirolimus (rapamycin) has been used successfully but may not fully control symptoms. Targeted treatment with CTLA4-Ig (immunoglobulin) fusion protein to "add back" functional CTLA4 has been used in an increasing number of cases and has shown significant therapeutic benefit, although questions remain about how long patients can be maintained on this therapy. HCT has shown encouraging results in patients who still have severe or progressive disease despite treatment.[19]

LRBA deficiency (autosomal recessive) has many features that overlap with CTLA4 haploinsufficiency, and early therapeutic attempts demonstrated an excellent disease response to CTLA4-Ig fusion protein. This observation led to the finding that LRBA is involved in recycling of CTLA4 back to the surface of T cells so that in its absence, T cells are partially deficient in their expression of CTLA4. *LRBA* is a large gene, and a substantial number of identified mutations are copy number variants of various sizes that can be missed by traditional next-generation sequencing approaches, which can make confirmation of a diagnosis difficult. Like CTLA4 haploinsufficiency, HCT has also been successful in patients with LRBA deficiency, although some patients do not achieve a complete remission of symptoms. LRBA is broadly expressed, and given its role in recycling CTLA4, it is likely that LRBA carries out additional cellular functions that may not be corrected by transplantation of only the hematopoietic cells.[20]

NONMALIGNANT LYMPHOPROLIFERATION

The best-described clinical disorder associated with nonmalignant lymphoproliferation is ALPS. ALPS typically becomes manifest within the first 5 years of life with nonmalignant, chronic peripheral lymphadenopathy and/or splenomegaly. Hepatomegaly develops in approximately 50% of cases. Patients typically have recurrent episodes of autoimmune cytopenias (hemolytic anemia and thrombocytopenia) and may develop more extensive autoimmunity, including autoimmune liver disease, glomerulonephritis, aplastic anemia, alopecia areata, and other features. Over time the lymphoproliferation tends to lessen, but often the autoimmune issues persist long term. ALPS patients have a high lifetime risk of lymphoma.

FAS deficiency (autosomal dominant) can occur as either a germline or somatically acquired defect. In addition to the clinical autoimmune features noted above, patients can have hypergammaglobulinemia with autoantibodies targeting hematopoietic cells (red cells, platelets, and neutrophils) as well as anti-phospholipid antibodies. Patients were found to have elevated α/β double-negative ($CD4^{neg}CD8^{neg}$) T cells in the peripheral circulation, although a number of other useful biomarkers have been identified which can aid in confirming a diagnosis of FAS deficiency. These include increased circulating levels of soluble FAS ligand (FASLG), vitamin B_{12}, IL-10, and IL-18. Because of the increased risk of lymphoma (most highly associated with mutations in the intracellular domain of FAS), patients should be counseled to watch for changes in the degree or character of their lymphadenopathy and for "B symptoms" (fever, night sweats, weight loss). Positron emission tomography (PET) scans and biopsies may be required to evaluate suspicious cases. Traditionally, first-line treatment to manage the lymphoproliferation and autoimmunity has been steroids, but steroid-sparing agents, including mycophenolate mofetil (MMF) and sirolimus (rapamycin), have been used for long-term therapy. Among these, sirolimus has proven to have a particularly beneficial effect on the lymphoproliferation, which has encouraged some to use this as a first-line therapy for FAS deficient ALPS. In patients who fail pharmacologic therapy, splenectomy was often considered, but the risks and benefits should be carefully weighed since post-splenectomy infection and sepsis are the most lethal complication of the disease. Since the disease of FAS-deficient patients tends to improve with age, HCT is not usually considered; there are, however, a small number of patients who have been treated with HCT with reportedly good survival and disease response.[21]

CASP10 deficiency (autosomal recessive/autosomal dominant) is associated with diffuse lymphadenopathy, splenomegaly,

autoimmune cytopenias, hypergammaglobulinemia, and development of pathogenic autoantibodies. Patients may have expanded double-negative T cells in peripheral blood.

PIK3CD and PIK3R1–related defects (autosomal dominant) lead to hyperactivity of a key phosphatidyl inositol-3 kinase (PI-3 kinase Δ) that is present in hematopoietic cells. Mutations cause either a GOF in the kinase subunit (PIK3CD) or a LOF in the regulatory subunit (PIK3R1) of this kinase, leading to an overall increase in activity. The resultant dysregulated growth of B and T cells is associated with lymphoproliferation manifested by lymphadenopathy, splenomegaly, and lymphocytic infiltrates in the lung and GI tract, causing interstitial pneumonitis and enteropathy. Most patients have hypogammaglobulinemia with recurrent bacterial sinopulmonary infections and a susceptibility to herpes group viruses, particularly Epstein-Barr virus (EBV), which often leads to persistent EBV viremia. Some patients develop autoimmune cytopenias and other organ-specific autoimmunity (kidney, liver, etc.). The combination of lymphoproliferative drive paired with a particular susceptibility to EBV leads to a high rate of malignant transformation (>10%) in these patients, with Hodgkin or non-Hodgkin lymphomas being the most common.

Treatment of PIK3CD and PIK3R1 disease typically involves a combination of immunoglobulin replacement therapy and antibiotics to control the recurrent bacterial infections as well as immunomodulatory therapy to control the lymphoproliferative and autoimmune aspects of the disease. Sirolimus (rapamycin) and B-cell-depletion therapy (rituximab, etc.) have shown significant benefits in treating the lymphoproliferative aspects of disease, and phase I/II clinical trials of targeted PI-3 kinase inhibitors (e.g., leniolisib) have shown promise as potential future treatment options. Patients who develop malignancies often undergo HCT as part of their treatment regimen, and their overall outcomes have been reasonably good but are confounded by the underlying malignancy and risks for disease relapse. HCT has not yet been broadly employed prophylactically in patients with PIK3CD or PIK3R1 defects, but is a consideration.[22,23]

NRAS and KRAS GOF defects (autosomal dominant) are somatic mutations that cause hematopoietic cells to become resistant to apoptosis. As a result, patients typically have leukocytosis with a persistent monocytosis that can be difficult to differentiate from juvenile myelomonocytic leukemia (JMML). As a group, these disorders are often referred to as RALD. Most patients have persistent splenomegaly, modest lymphadenopathy, and autoantibodies to a variety of targets, some of which may be pathogenic. These can cause cytopenias and other autoimmune disorders, including systemic lupus, which has been reported in a small number of cases. Like patients with PI-3 kinase pathway defects, patients with RALD are at higher risk of developing hematopoietic malignancies, particularly JMML, and should be counseled accordingly. Despite having a clinical phenotype that overlaps with FAS deficiency, RALD patients typically have normal to mildly elevated double-negative T cells and normal serum FAS ligand and vitamin B_{12} levels. A variety of immunomodulatory therapies have been tried for RALD, but responses vary. For patients who develop JMML, HCT has become standard of care; HCT is not typically considered prophylactically in patients with RALD.[24]

PRKCD deficiency (autosomal recessive) has only been described in a handful of patients. Lymphadenopathy, splenomegaly, hypergammaglobulinemia, and the presence of autoantibodies were common features. In addition, some patients developed lupus with nephritis, rash, serositis, and arthritis. These patients were treated with steroids and MMF, while another patient without lupus was reported to have a good clinical response to sirolimus (rapamycin) with a decrease in splenomegaly and hypergammaglobulinemia.[25]

CONCLUSIONS

There are a growing number of IEI characterized by immune dysregulation. In general, these have been linked either to activating mutations in molecules or pathways that promote immune effector cell growth, activation, or survival or to a loss of regulatory molecules or cells that control these processes. Patients with IEI defects generally present with autoimmunity, inflammatory disease, or lymphoproliferation. Since these processes may target multiple organs, patients often seek care in a variety of different subspecialty clinics. As a consequence, it is important for providers in virtually all medical subspecialties to have an awareness of these disorders and the fact that genetic testing can provide both a diagnosis and significant insight into potential targeted therapies that may be particularly effective in treating these patients.

KEY CONCEPTS

- Immune regulatory disorders should be suspected in the setting of autoimmune, inflammatory, or nonmalignant lymphoproliferative disease that is:
  - **Too much**—Multiple organ systems involved, multiple cell lines affected, multiple autoantibodies to diverse targets.
  - **Too severe**—Immune dysregulation is unusually severe and/or life-threatening.
  - **Too weird**—Unusual autoimmune disease, unusual or severe infections in addition to immune dysregulation.
  - **Too young**—Onset of disease in infancy or early childhood.
- Immune regulatory disorders have significant phenotypic overlap, so making a specific diagnosis can be challenging.
- Broad-based genetic testing (i.e., whole exome or whole genome sequencing, large gene panel) is the diagnostic approach most likely to lead to a specific diagnosis.

CLINICAL PEARLS

Common autoimmune, inflammatory, and lymphoproliferative disease manifestations:

- **Hematologic**—autoimmune cytopenias (anemia, thrombocytopenia, neutropenia)
- **GI tract**—autoimmune or inflammatory enteropathy, nodular lymphoid hyperplasia, autoimmune hepatitis, nodular regenerative hyperplasia (liver)
- **Lungs**—**interstitial** lung disease, follicular bronchiolitis, non-caseating granulomas
- **Skin**—**eczema**, psoriasis, other dermatitides
- **Endocrine**—autoimmune thyroiditis, type I diabetes, gonadal insufficiency, etc.
- **Cardiovascular**—vasculitis

Less common autoimmune, inflammatory, and lymphoproliferative disease manifestations:

- **Musculoskeletal**—arthritis, myositis
- **Cardiovascular**—vasculitis
- **Neurologic**—lymphoproliferation, strokes

THERAPEUTIC PRINCIPLES

- The primary goal of treatment is to control autoimmunity, inflammation, or nonmalignant lymphoproliferation in order to ameliorate acute symptoms and prevent permanent tissue and organ damage.
- Broadly active immunomodulators (e.g., corticosteroids) may be required acutely to stabilize acute illness, but for long-term therapy, efforts should be made to target treatment as much as possible to avoid undesired side effects.
- Identification of a genetic cause of disease can often provide guidance for targeted treatment (e.g., CTLA4-Ig in CTLA4 haploinsufficiency, JAK inhibitors in JAK or STAT gain-of-function disorders).
- Biopsies of affected tissue followed by staining of biopsies for immune cells (B cells, T cells, eosinophils, etc.) or immune proteins (complement, IgA, etc.) can provide useful information about the components of the immune system mediating disease at the level of the tissue, allowing therapy to be targeted accordingly.
- Use of immunomodulators may cause increased risk for infection, so antimicrobial prophylaxis should be considered in all cases, taking into consideration the patient's underlying infectious susceptibilities.
- In patients with pre-existing infections, treatment needs to address both immune dysregulation and infections in order to optimize outcomes.

JAK, Janus kinase.

ON THE HORIZON

- The number of genetically defined immune dysregulation disorders is likely to increase in the coming years.
- Newly identified genetic defects associated with a phenotype of immune dysregulation will highlight molecular pathways and molecules that play key roles in regulating immunity.
- A growing number of targeted therapies will increase opportunities to provide precision treatments to patients with a genetic diagnosis.
- There is a significant need for development of new tools and approaches to identify disease causing somatic variants in patients.

REFERENCES

1. Bousfiha A, Jeddane L, Picard C, et al. Human inborn errors of immunity: 2019 update of the IUIS phenotypical classification. *J Clin Immunol*. 2020;40(1):66–81.
2. Van Horebeek L, Dubois B, Goris A. Somatic variants: new kids on the block in human immunogenetics. *Trends Genet*. 2019;35(12):935–947.
3. Aluri J, Bach A, Kaviany S, et al. Immunodeficiency and bone marrow failure with mosaic and germline TLR8 gain-of-function. *Blood*. 2021;137(18):2450–2462.
4. Kawakami T, Sekiguchi N, Kobayashi J, et al. Frequent STAT3 mutations in $CD8^+$ T cells from patients with pure red cell aplasia. *Blood Adv*. 2018;2(20):2704–2712.
5. Louvrier C, Assrawi E, El Khouri E, et al. NLRP3-associated autoinflammatory diseases: phenotypic and molecular characteristics of germline versus somatic mutations. *J Allergy Clin Immunol*. 2020;145(4):1254–1261.
6. Simon AJ, Golan AC, Lev A, et al. Whole exome sequencing (WES) approach for diagnosing primary immunodeficiencies (PIDs) in a highly consanguineous community. *Clin Immunol*. 2020;214:108376.
7. Okano T, Imai K, Naruto T, et al. Whole-exome sequencing-based approach for germline mutations in patients with inborn errors of immunity. *J Clin Immunol*. 2020;40(5):729–740.
8. Rudilla F, Franco-Jarava C, Martínez-Gallo M, et al. Expanding the clinical and genetic spectra of primary immunodeficiency-related disorders with clinical exome sequencing: expected and unexpected findings. *Front Immunol*. 2019;10:2325.
9. Ammann S, Elling R, Gyrd-Hansen M, et al. A new functional assay for the diagnosis of X-linked inhibitor of apoptosis (XIAP) deficiency. *Clin Exp Immunol*. 2014;176(3):394–400.
10. Hou TZ, Qureshi OS, Wang CJ, et al. A transendocytosis model of CTLA-4 function predicts its suppressive behavior on regulatory T cells. *J Immunol*. 2015;194(5):2148–2159.
11. Break TJ, Oikonomou V, Dutzan N, et al. Aberrant type 1 immunity drives susceptibility to mucosal fungal infections. *Science*. 2021;371(6526).
12. Kekäläinen E, Tuovinen H, Joensuu J, et al. A defect of regulatory T cells in patients with autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy. *J Immunol*. 2007;178(2):1208–1215.
13. Gambineri E, Ciullini Mannurita S, Hagin D, et al. Clinical, immunological, and molecular heterogeneity of 173 patients with the phenotype of immune dysregulation, polyendocrinopathy, enteropathy, X-linked (IPEX) syndrome. *Front Immunol*. 2018;9:2411.
14. Barzaghi F, Amaya Hernandez LC, Neven B, et al. Long-term follow-up of IPEX syndrome patients after different therapeutic strategies: an international multicenter retrospective study. *J Allergy Clin Immunol*. 2018;141(3):1036–1049. e5.
15. Lorenzini T, Dotta L, Giacomelli M, et al. STAT mutations as program switchers: turning primary immunodeficiencies into autoimmune diseases. *J Leukoc Biol*. 2017;101(1):29–38.
16. Caudy AA, Reddy ST, Chatila T, et al. CD25 deficiency causes an immune dysregulation, polyendocrinopathy, enteropathy, X-linked-like syndrome, and defective IL-10 expression from CD4 lymphocytes. *J Allergy Clin Immunol*. 2007;119(2):482–487.
17. Fernandez IZ, Baxter RM, Garcia-Perez JE, et al. A novel human IL2RB mutation results in T and NK cell-driven immune dysregulation. *J Exp Med*. 2019;216(6):1255–1267.
18. Zhang Z, Gothe F, Pennamen P, et al. Human interleukin-2 receptor β mutations associated with defects in immunity and peripheral tolerance. *J Exp Med*. 2019;216(6):1311–1327.
19. Schwab C, Gabrysch A, Olbrich P, et al. Phenotype, penetrance, and treatment of 133 cytotoxic T-lymphocyte antigen 4-insufficient subjects. *J Allergy Clin Immunol*. 2018;142(6):1932–1946.
20. Tesch VK, Abolhassani H, Shadur B, et al. Long-term outcome of LRBA deficiency in 76 patients after various treatment modalities as evaluated by the immune deficiency and dysregulation activity (IDDA) score. *J Allergy Clin Immunol*. 2020;145(5):1452–1463.
21. Rieux-Laucat F, Magérus-Chatinet A, Neven B. The autoimmune lymphoproliferative syndrome with defective FAS or FAS-ligand functions. *J Clin Immunol*. 2018;38(5):558–568.
22. Jamee M, Moniri S, Zaki-Dizaji M, et al. Clinical, immunological, and genetic features in patients with activated PI3Kδ syndrome (APDS): a systematic review. *Clin Rev Allergy Immunol*. 2020;59(3):323–333.
23. Nunes-Santos CJ, Uzel G, Rosenzweig SD. PI3K pathway defects leading to immunodeficiency and immune dysregulation. *J Allergy Clin Immunol*. 2019;143(5):1676–1687.
24. Calvo KR, Price S, Braylan RC, et al. JMML and RALD (Ras-associated autoimmune leukoproliferative disorder): common genetic etiology yet clinically distinct entities. *Blood*. 2015;125(18):2753–2758.
25. Salzer E, Santos-Valente E, et al. Protein kinase C δ: a gatekeeper of immune homeostasis. *J Clin Immunol*. 2016;36(7):631–640.

Neutrophils and Neutrophil Disorders

Steven M. Holland and Gülbü Uzel

We have learned a great deal about phagocytes since their discovery by Metchnikoff in 1905: neutrophils, monocytes, macrophages, and eosinophils traffic to sites of infection or inflammation and engulf microorganisms and apoptotic cells as the lead players in the innate immune response.

NEUTROPHILS

Neutrophils, also known as *granulocytes* because of their numerous cytoplasmic granules, are crucial for the host defense against bacteria and fungi. They are bone marrow–derived, terminally differentiated cells incapable of further cellular division. They have a short life span in the circulation ($t_{1/2}$ ≈7 hours) but survive an additional 1 to 2 days in tissue. In peripheral blood, they are normally maintained at 3000–6000 cells/mm^3 and represent 30% to 50% of the circulating leukocytes. There are four pools of neutrophils in vivo: (i) the bone marrow pool (≈90% of the total); (ii) the circulating pool (≈3% of the total); (iii) the marginated pool (adherent to the endothelium, ≈4% of the total); and (iv) those located in the tissues as extravasated or exudative neutrophils. About 55% to 60% of bone marrow is dedicated to the production of neutrophils, producing around 10^{11} cells daily, but production increases in times of stress.

Myeloid cell differentiation is a complex process that typically extends over 2 weeks in bone marrow. The pluripotent stem cell, the precursor for all hematopoiesis, develops into lineage-committed progenitors, which proceed to terminally differentiated distinct cells, all the while preserving and regenerating more pluripotent stem cells.[1]

PRODUCTION OF MACROPHAGES AND GRANULOCYTES

The pluripotent stem cell gives rise to the myeloid stem cell from which the colony-forming unit granulocyte/erythrocyte/macrophage/megakaryocyte (CFU-GEMM) is derived. Among the growth factors that are influential at this step are stem cell factor (SCF), interleukin-3 (IL-3), and granulocyte/macrophage–colony-stimulating factor (GM-CSF).[2] The CFU-GEMM further differentiates into the colony-forming unit–granulocyte/macrophage (CFU-GM) under the continuing influence of these growth factors. The colony-forming unit–granulocyte (CFU-G), a neutrophil lineage committed precursor, is derived from CFU-GM under the control of IL-3, GM-CSF, and granulocyte–colony-stimulating factor (G-CSF). The myeloblast is formed from the CFU-G under the influence of GM-CSF and G-CSF and is the first morphologically distinct cell of the neutrophil lineage. Promyelocyte, myelocyte, metamyelocyte, band form, and mature neutrophil formation follow consecutively under the ongoing control of G-CSF and GM-CSF. The maturation process from stem cell to myelocyte takes 4 to 6 days, with an additional 5 to 7 days for the myelocyte to form the mature neutrophil, all in bone marrow.

Macrophage differentiation is similar to granulocyte differentiation in many respects. The CFU-GM differentiates into the colony-forming unit–macrophage (CFU-M) followed by the formation of the monoblast, promonocyte, and monocyte under the influence of macrophage colony-stimulating factor (M-CSF).[3] After monocytes are released into blood, they circulate for 1 to 4 days before entering tissues, where they further differentiate into macrophages.

EVOLUTION OF NEUTROPHIL GRANULES

During myelopoiesis in bone marrow, the first granules forming at the promyelocyte stage, stain blue with the Wright or Romanowsky stain, and are called *primary granules* or *azurophilic granules*. Their formation ceases at the myelocyte stage, and they are distributed among the daughter cells. These primary granules contain microbicidal enzymes, including defensins, hydrolases, and proteases (Table 39.1). As the granulocyte precursors mature and divide, the number of primary granules per cell decreases. After the promyelocyte stage, secondary or specific granules form. In the mature neutrophil, they comprise about two-thirds of the granules. The secondary granules are less dense and contain cytochrome b_{558}, lysozyme, lactoferrin, and collagenase. The gelatinase-containing tertiary granule forms after the metamyelocyte stage and can be detected in the band form and the mature granulocyte.

DISORDERS OF NEUTROPHIL PRODUCTION

Chronic neutropenia refers to conditions lasting more than 6 months with an absolute neutrophil count (ANC) of less than 500 cells/μL. Chronic neutropenia has many etiologies, as listed in Table 39.2.

Severe Congenital Neutropenia and Cyclic Neutropenia

Kostmann originally described an extensive northern Swedish kindred with both recessive and dominant neutropenia, but subsequently sporadic cases were added, making Kostmann neutropenia a confusing mélange of syndromes.[4] Severe congenital neutropenia (SCN) is now known to be a heterogeneous

TABLE 39.1 Neutrophil Granule Components

Granule	Contents	Properties
Primary (azurophilic) granules	*Lysosomal hydrolases* Myeloperoxidase Defensins Lysozyme Elastase Cathepsin G Azurocidin Proteinase 3 Bacterial—permeability increasing protein (BPI) *Acid hydrolases* Cathepsin B Cathepsin D β-Glycerophosphatase granulocyte β-Glucuronidase *N*-acetyl-β-glucosaminidase α-Mannosidase *Other* Collagenase	• First formed during myelopoiesis at promyelocyte stage • Appearing blue when stained with Wright stain • Least mobilizable of all granules • Measuring ≈ 0.8 μm in diameter • Containing defensins making up 30%–50% of granule contents • Augmenting the microbial damage initiated by reactive oxidants • Helping to digest dead microbes and host cells • Neutralizing gram-negative bacteria by means of BPI
Secondary (specific) granules	*Lysosomal hydrolases* Lysozyme *Other* Collagenase Gelatinase Lactoferrin Vitamin B_{12}–binding proteins Cytochrome b_{558} Histaminidase FMLF receptors C3bi receptors	• First formed at the myelocyte stage • Specifically found specific in phagocytes • Measuring ≈0.5 μm in diameter • Binding proteins to deprive microorganisms of nutrients • Having positively charged, enhancing cell surface
Tertiary (smaller) granules	*Acid hydrolases* Cathepsin B Cathepsin D β-Glycerophosphatase granulocyte β-Glucuronidase *N*-acetyl-β-glucosaminidase α-Mannosidase *Other* Gelatinase	• Heterogeneous population of organelles, including C-particles and secretory vesicles • Detected in the band form and mature neutrophils

TABLE 39.2 Causes of Neutropenia

Classification	Etiology
Hematological	Kostmann syndrome Severe congenital neutropenia Cyclic neutropenia Myelodysplastic syndrome Aplastic anemia Leukemia
Immunological/inflammatory disorders	Severe combined immunodeficiency (SCID), certain types Hyper-IgM syndrome (CD40L deficiency) Chediak-Higashi syndrome Cartilage–hair hypoplasia Reticular dysgenesis Dyskeratosis congenita Autoimmune neutropenia Isoimmune neutropenia
Infections	Human immunodeficiency virus (HIV) Parvovirus Epstein-Barr virus Malaria Cytomegalovirus
Inborn errors of metabolism/nutritional disorders	Gaucher disease Glycogen storage disease, type Ib Transcobalamin deficiency Vitamin B_{12}, folate deficiency
Other	Schwachman-Bodian-Diamond syndrome Idiopathic neutropenia Chemotherapy Radiation therapy Drugs (e.g., vancomycin, chloramphenicol, sulfamethoxazole, clozapine) Toxins (e.g., benzene) Dialysis Reticuloendothelial sequestration

group of disorders that present similarly. The genes recognized as mendelian causes of neutropenia are neutrophil elastase (*ELANE* or *ELA2*), *HAX1*, *G6PC3*, *GFI1*, *GATA2, JAGN1*, *VPS45* genes, and activating mutations in the Wiskott–Aldrich syndrome (*WAS*) gene.

SCN is usually diagnosed in early infancy because of the occurrence of life-threatening pyogenic infections, cellulitis, stomatitis, peritonitis, perirectal abscess, or meningitis. The most common bacteria isolated are *Staphylococcus aureus, Escherichia coli*, and *Pseudomonas aeruginosa*. Patients usually have an ANC ≤200 cells/μL, mild anemia, and hypergammaglobulinemia, sometimes with eosinophilia and monocytosis. SCN represents a maturational arrest of neutrophil precursors at the level of promyelocytes or myelocytes in bone marrow. A subset of patients with SCN (7.5% to 10%) subsequently develops a myelodysplastic syndrome (MDS) or acute myeloid leukemia (AML), which has been associated with acquired truncation mutations of the G-CSF receptor (G-CSFR).

The majority of patients with SCN have heterozygous mutations in the neutrophil elastase (*ELANE*).[5] The fascinating cyclic form of this disorder has neutrophil counts oscillating with 21-day cycles: hence the name cyclic neutropenia (CN). These typically missense mutations are transmitted as autosomal dominant mutations but also occur spontaneously. There is no clear genotype–phenotype correlation between specific *ELANE* mutations that lead to CN as opposed to SCN. They lead to intracellular accumulation of mutant proteins that are inappropriately trafficked into azurophilic granules.[5] The mutated, aberrantly folded elastase is thought to contribute to neutrophil precursor apoptosis and the clinical phenotype of neutropenia, but the mechanisms by which this occurs are still obscure. Treatment with subcutaneous G-CSF can increase the ANC above 1000 cells/μL, with a decrease in the frequency of infections and significant clinical improvement.[6] However, patients with SCN who have received long-term G-CSF therapy are at an increased risk of developing AML or MDS, which correlates with overall G-CSF responsiveness.

Homozygous loss-of-function mutations in *HAX1* account for the majority of recessive cases of SCN,[7] some of which were in original pedigree described by Kostmann. Patients may have isolated SCN or associated neurological problems (cognitive impairment, developmental delay, or epilepsy), depending on which isoform of *HAX1* is mutated. Patients with mutations affecting isoform A have only SCN as opposed to patients with

mutations affecting both isoforms (A and B), who develop neurological problems in addition.[8]

Dominant zinc finger mutations disabling transcriptional repressor activity of the *growth factor independent 1* (*GFI1*) gene have been described in a few patients with SCN.[4] *GFI1* encodes a transcriptional repressor prooncogene controlling normal hematopoietic cell differentiation and also regulating *ELANE* as well as several genes encoding CAAT enhancer binding proteins (C/EBP). Mutations in *GFI1* are also associated with aberrations in lymphoid and myeloid cells, leading to a circulating population of immature myeloid cells. *Gfi1* knock-out mice have impaired regulation of T-helper type 2 (Th2) cells and differentiation of B, Th17, and dendritic cells (DC).

Mutations in the glucose-6-phosphatase catalytic subunit 3 (*G6PC3*) complex cause another form of SCN along with developmental and somatic problems.[9] *G6PC3* encodes glucose-6-phosphatase-β, which hydrolyzes glucose-6-phosphate (G6P) in the final step of gluconeogenesis and glycogenolysis. It is coupled to a glucose transporter (G6PT) that facilitates G6P transport from the cytoplasm to the endoplasmic reticulum. Mutations in the *G6PT* gene lead to glycogen storage disease type Ib, which has variable neutropenia and infections and other complications, such as liver adenomas, growth retardation, osteoporosis, polycystic ovaries, and inflammatory bowel disease (IBD). Children with these complications have increased susceptibility to bacterial infections and cardiovascular abnormalities, including prominent ectatic superficial veins.

Shwachman-Bodian-Diamond Syndrome

Shwachman-Bodian-Diamond syndrome (SBDS) was first described in 1964 as a disorder with pancreatic exocrine insufficiency and bone marrow dysfunction. Currently, it is recognized as the second most common cause of inherited exocrine pancreatic insufficiency after cystic fibrosis. It is autosomal recessive with an estimated incidence of 0.5 to 1/100,000.[10] The SDBS protein belongs to a highly conserved protein family involved in RNA metabolism. Mutations cause defects in the development of the exocrine pancreas, hematopoiesis, and chondrogenesis. Recurring mutations result from gene conversion caused by recombination with a pseudogene in 89% of unrelated patients; 60% carry two converted alleles. (Pseudogene conversion is also the cause of the majority of cases of $p47^{phox}$-deficient chronic granulomatous disease [CGD].)

Patients present with recurrent infections, failure to thrive, hematopoietic dysfunction, metaphyseal dysostosis, growth retardation, and fatty replacement of the pancreas. Most patients have mild neutropenia, and a few have intermittent or chronic neutrophil counts ≤500 cells/μL.[11] Anemia and thrombocytopenia are associated with neutropenia. Congenital aplastic anemia with anemia, thrombocytopenia, and neutropenia is an unusual presentation of SBDS. Upper and lower respiratory tract pyogenic infections are common and related to neutropenia. Short ribs with broadened anterior ends are common radiological findings, along with metaphyseal dyschondroplasia of the femoral head. The diagnosis is suggested by neutropenia, radiological findings, and abnormal pancreatic exocrine function. It is confirmed by gene sequencing.

Autoimmune Neutropenia

Autoimmune neutropenia (AIN) is caused by peripheral destruction of neutrophils as a result of granulocyte-specific autoantibodies.[12]

Primary Autoimmune Neutropenia

Primary AIN is seen in infancy unassociated with other systemic immune-mediated disorders and is the most common form of neutropenia, equally affecting boys and girls at around 1/100,000.[12] The average age at diagnosis is 8 months. The majority present with mild skin and upper respiratory tract infections; some patients remain asymptomatic despite low ANCs. The majority of patients have a neutrophil count ≥500 cells/μL at diagnosis, but ANCs may transiently increase two- to threefold during severe infection. Bone marrow shows normal to increased cellularity. Myeloid precursors typically reach the myelocyte/metamyelocyte stage. Phagocytosed granulocytes in bone marrow may indicate removal of sensitized granulocytes there. Granulocyte-specific antibodies are detected by direct granulocyte immunofluorescence testing (D-GIFT), the vast majority of which are immunoglobulin G (IgG) against glycoproteins of the granulocyte membrane designated neutrophil antigens (NAs). NAs are located on IgG receptor IIA or IIIB (FcγRIIa and FcγRIIIb).

AIN is generally self-limiting. Disappearance of the antibodies from the circulation precedes normalization of neutrophil counts. Prophylactic antibiotic treatment should be reserved for those with recurrent infections. Alternative treatment strategies for severe infections and in the setting of emergency surgical interventions include high-dose intravenous immunoglobulin (IVIG), corticosteroids, and G-CSF—with the latter being the most effective at increasing the ANC.

Secondary Autoimmune Neutropenia

Secondary AIN can be seen at any age and has a more variable clinical course. Hepatitis, systemic lupus erythematosus (SLE), or Hodgkin disease may underlie it and cause other autoimmune problems as well. These antineutrophil antibodies (ANAs) have pan-FcγRIII specificity. CD18/CD11b antibodies have been detected in a subset of patients with secondary AIN that responds poorly to most therapies.

Alloimmune Neonatal Neutropenia

First described by Lalezari in 1966, alloimmune neonatal neutropenia (ANN) is caused by the transplacental transfer of maternal antibodies against the fetal NAs NA1, NA2, and NB1, leading to immune destruction of neonatal neutrophils.[13] These complement-activating antineutrophil IgG antibodies can be detected in about 1/500 live births. Antibody-coated neutrophils in ANN are phagocytosed by the reticuloendothelial system and removed from the circulation, leaving the neutropenic neonate at risk for infections. Omphalitis, cellulitis, and pneumonia typically occur within the first 2 weeks of life in association with neutropenia. Diagnosis can be made by detection of neutrophil-specific alloantibodies in the maternal serum. ANN responds to G-CSF or high-dose IVIG, but most infants improve without specific treatment in a few weeks to 6 months with waning of maternal antibody.

DEFECTS OF LEUKOCYTE ADHESION

Migration of circulating leukocytes from the bloodstream into tissues depends on complex bidirectional interactions between leukocytes and endothelial cells (Chapter 16). The initial steps involve the activation of circulating leukocytes by signal molecules released from inflamed tissues or from the bacteria

themselves. After activation by chemotactic factors—such as the complement fragment C5a, IL-8, leukotriene B4 (LTB4), or the bacterial product formyl-methionyl-leucyl-phenylalanine (fMLF)—leukocytes rapidly become adhesive to the endothelium, other leukocytes, or laboratory surfaces. The activation process involves translocation of subcellular granules containing adhesion molecules (CD18/CD11b) to the polymorphonuclear leukocyte (PMN) surface and qualitative alterations in the adhesion molecules constitutively expressed on the plasma membrane. Adhesion and transmigration of leukocytes occur as a result of interactions between three groups of molecules: leukocyte integrins, endothelial intercellular adhesion molecules (ICAMs, members of the immunoglobulin supergene family), and glycosaminoglycans or selectins (Fig. 39.1). The first step in targeting PMNs to inflamed tissues is the rolling or tethering of PMNs on the endothelium of postcapillary venules.[14] This is attributed to the interactions between CD15s (sialyl Lewis^X or SLe^X) expressed on the leukocyte surface and P-selectin or E-selectin, members of the selectin family of adhesion molecules expressed on the vascular endothelium. In addition, L-selectin on the leukocyte surface interacts with its counterligands P-selectin, CD34, glyCAM-1, and other glycoproteins located on the endothelial surface. *Rolling*—a relatively low-affinity interaction mediated by selectins—is followed by *firm adhesion*, which is a high-affinity interaction between integrins on the neutrophil and ICAMs on the endothelium. Adhesion is followed by the transmigration of neutrophils between endothelial cells out to the extracellular matrix (ECM).

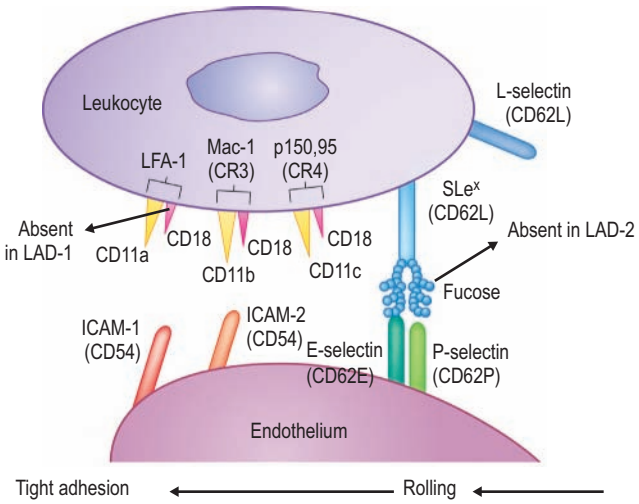


FIG. 39.1 Leukocyte Adhesion to Nonlymphoid Endothelium. Selectins (*L-selectin/CD62L, P-selectin/CD62P,* and *E-selectin/CD62E*), integrins (*CD18/CD11a* or *LFA-1, CD18/CD11b* or *Mac-1*, and *CD18/CD11c* or *p150,95*), and intercellular adhesion molecules (*ICAMs*) are involved in leukocyte adhesion to the nonlymphoid endothelium. Rolling or low-affinity tethering, the initial step of leukocyte adhesion, is mediated by the interactions of E- and P-selectin on endothelial surfaces with the sialyl Lewis^x (*SLe^x* or CD15s) of leukocytes as well as L-selectin on the leukocyte surfaces with its counterligands CD34 or glyCAM-1. Rolling facilitates tight adhesion as a result of the interactions of leukocyte function–associated antigen-1 (*LFA-1*) with ICAM-1 or ICAM-2 and Mac-1 with ICAM-2. CD18 is missing or dysfunctional in leukocyte adhesion defect-1 (*LAD-1*); SLe^x is missing in LAD-2.

Leukocyte Adhesion Defect-1

In the 1970s, infants and children were recognized with severe, recurrent life-threatening bacterial infections affecting the skin, gingiva, and lungs, often with delayed separation of the umbilical stump with severe omphalitis. These patients were shown to have defects in membrane expression of the leukocyte adhesion glycoproteins of the integrin superfamily.[15,16]

Integrins are noncovalently associated, heterodimeric cell surface receptors, comprising one α subunit (CD11a, CD11b, or CD11c) and a common β chain (CD18), the latter required for surface expression of the CD11 chains. These proteins mediate leukocyte adhesion to the endothelium and other leukocytes. Leukocyte adhesion defect-1 (LAD-1) results from mutations in the CD18 gene (*ITGB2*), located on chromosome 21q22. Patients with LAD-1 have defective polymorphonuclear cell adherence, leading to defective chemotaxis and trafficking as well as low natural killer (NK) and cytotoxic T-lymphocyte (CTL) activity. The absence of CR3 leads to loss of complement-mediated phagocytosis and bacterial killing. LAD-1 is often manifested by delayed umbilical cord separation, omphalitis, persistent leukocytosis, destructive periodontitis, and recurrent infections with *S. aureus*, *P. aeruginosa*, and *Klebsiella* spp. Patients with some residual CD18 expression and function (i.e., hypomorphic mutations) live beyond childhood with less frequent or severe infections and do not typically have delayed umbilical cord separation. Persistent neutrophil leukocytosis (usually >15.000 cells/μL) in the absence of infection is common in all patients, driven by both low-level ongoing infection and impaired exit of neutrophils from the circulation. Oral ulcers, severe periodontitis, gingivitis with apical bone loss (Fig. 39.2), and eventual loss of permanent teeth are major problems in LAD-1 and reflect excessive IL-17 expression by CD4 T cells as a result of uninhibited IL-23 production by tissue macrophages.[17] Necrotizing cutaneous ulcers with delayed wound healing and lingering eschar formation are common (Fig. 39.3). Defective chemotaxis and adhesion mean that leukocytes fail to migrate to sites of infection, accounting for the inability to form pus and erythema at the site of infection. Biopsies of the ulcers characteristically show poorly formed granulation tissue and scant fibrinous exudate without neutrophils. Ulcerative gastrointestinal (GI) disorders resembling IBD are also recognized in LAD-1, especially as patients grow older.

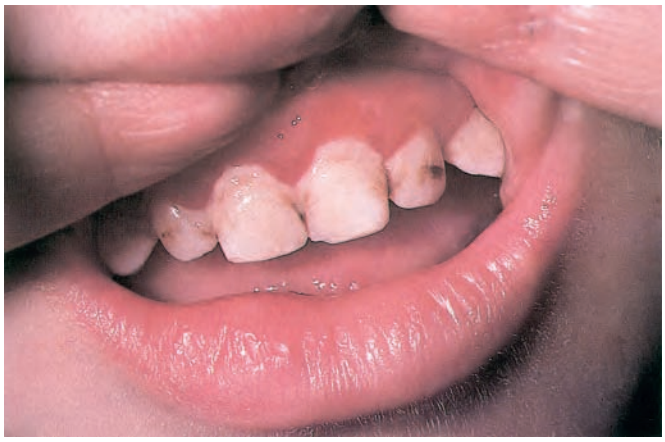


FIG. 39.2 Oral Pathology in Leukocyte Adhesion Defect-1 (LAD-1). Gingivitis and severe periodontitis are hallmarks of LAD-1.

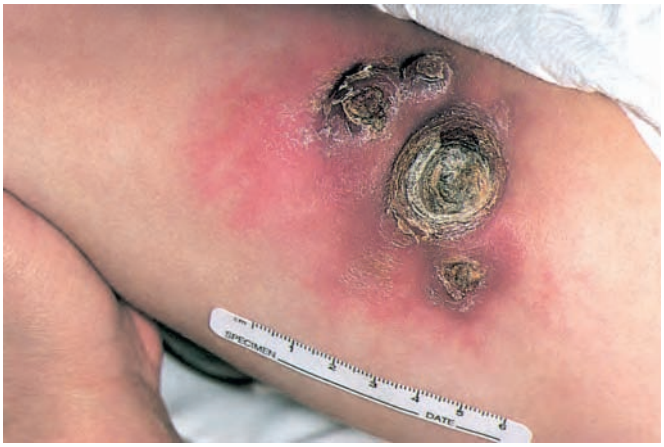


FIG. 39.3 Skin Infection in Leukocyte Adhesion Defect-1 (LAD-1). Failure to form pus, inability to demarcate the fibrotic skin debris, and limited inflammation.

Although most cases of CD18 deficiency are homozygous, compound heterozygotes also occur.[18] The diagnosis is usually made by flow cytometric analysis of neutrophils showing decreased or absent CD18 and its associated heterodimers—CD11a, CD11b, and CD11c—and confirmed by mutational analysis of *ITGB2*. More subtle phenotypes can be detected by testing for mobilization of CD18 complexes, such as CD18/CD11b from neutrophils upon cellular stimulation. Definitive therapy of LAD-1 is bone marrow transplantation. Infections must be managed aggressively, since inflammatory responses and clinical signs are unreliable in these patients with profoundly impaired innate immune responses. Surgery is often essential for debridement of nonhealing ulcers, which frequently need tissue grafts, but immunomodulation of the affected cytokine pathways may also be helpful. Although not correcting the underlying gene defect, but reflecting the excessive production of IL-23 and IL-17 at sites of inflammation that is associated with severe oral ulcers, periodontitis and bone loss, a patient with moderate LAD-1 (34% of control CD18 activity) was treated with ustekinumab, which blocks IL-23-dependent production of IL-17. After 1 year of treatment the patient's severe chronic periodontitis and a deep sacral ulcer had resolved without serious infections or adverse reaction. This approach to therapy offers additional insight into the complex pathophysiology of LAD-1 inflammation and suggests the possibility of novel therapeutic approaches to symptomatic management.[19]

KEY CONCEPTS

Leukocyte Adhesion Defect

- Three types of adhesion defect are known: leukocyte adhesion defect (LAD)-1, -2, and -3. There are two phenotypes for LAD-1: moderate and severe.
- LAD-1 results from mutations in CD18; LAD-2 is caused by mutations in sialyl Lewis^x (CD15s); LAD-3 results from mutations in *FERMT3*.
- High white blood cell count, delayed umbilical cord separation, recurrent bacterial infections, skin ulcers, defective wound healing, gingivitis, and periodontitis are the hallmarks of LAD-1.

Leukocyte Adhesion Defect-2

A distinct defect of leukocyte adhesion with susceptibility to infection was described by Etzioni et al. in 1992 and named LAD-2.[15,20] It is characterized by growth retardation and cognitive impairment, hypotonia, seizures, dysmorphic features, strabismus, and persistent periodontitis. In contrast to LAD-1, wound healing is not impaired, nor is the susceptibility to bacterial infections as severe. Hypofucosylation of the protein (SLe^X) on neutrophils impairs the rolling step of neutrophil adhesion. The underlying defect is in guanosine diphosphate (GDP)–fucose biosynthesis, resulting from mutations in the GDP–fucose transporter-1 (*FUCT1* or *SLC35C1*), hence the designation of this disease as a congenital disorder of glycosylation IIc (CDGIIc). In addition to severe impairment in neutrophil migration as in LAD-1, lymphocyte homing to skin is also defective. Patients with LAD-2 have had relatively mild courses of infections with several pneumonias and superinfection of varicella lesions,[20] and some have reportedly improved with fucose supplementation. In addition to SLe^X, fucosylated blood group antigens are also affected, leading to the Bombay blood group phenotype (lack of the H antigen) and Lewis a^-b^- in these patients. Absence of CD15 on patient neutrophils can be detected by flow cytometry. Effective and prompt treatment of infections is central to the management of LAD-2.

Leukocyte Adhesion Defect-3

A third leukocyte adhesion deficiency was recognized, LAD-3, initially named LAD-1/variant (LAD-1v), has a distinct infantile bleeding diathesis similar to Glanzmann-type thrombasthenia along with defective leukocyte adhesion.[21] Although CD18/CD11a (lymphocyte function–associated antigen-1 [LFA-1] or $\alpha_1\beta_2$) is the main integrin on leukocytes, $\alpha_{IIb}\beta_3$ (also called *GPIb-IIIa*) allows platelets to bind fibrinogen to promote clotting. This was initially described in Turkish patients and ascribed to the fermitin (kindlin) family member 3 gene *FERMT3*, which encodes KINDLIN3, an adaptor protein expressed in hematopoietic cells that regulates integrin activation.[22,23] KINDLIN3 activates integrins through binding to distinct motifs on the short tails of the integrin β subunits. Phenotypically, leukocytes and platelets in LAD-3 have defective β_3, β_2, and β_1 integrin activation as a result of loss of "inside out" or chemokine-mediated LFA-1 activation and intrinsic LFA-1/$\alpha_1\beta_2$ adhesiveness. In addition, these cells have decreased adherence to endothelial cells and reduced expression of the Rap-1 guanine nucleotide exchange factor, CalDAG-GEFI (CDGI). Based on the location and severity of mutation, LAD-3 leukocytes may also display loss of adhesion to vascular cell adhesion molecule-1 (VCAM-1). LAD-3 platelets have decreased binding to soluble fibrinogen, respond poorly to thrombin via thrombin receptors (PARs), and therefore have poor platelet granule secretion through integrin activation. Bone marrow transplantation is necessary and can be curative.

CHRONIC GRANULOMATOUS DISEASE

CGD, first described in 1954, is characterized by recurrent infections and hypergammaglobulinemia. It results from defective phagocyte superoxide production leading to impaired microbial killing. CGD comprises six genotypes with a relatively consistent phenotype (Table 39.3) of recurrent severe bacterial and fungal infections and tissue granuloma formation.[24] CGD occurs at around 0.5 to 1/100,000 births. It is inherited in X-linked and autosomal recessive patterns, with the relative frequencies of recessive disease depending on the rates of local consanguinity. In the United States, the X-linked form accounts for about

TABLE 39.3 Genotype–Phenotype Correlations in X-Linked Chronic Granulomatous Disease

	X91⁰	X91⁻	X91⁺
gp91^{phox} protein levels	Undetectable	Normal to low	Normal
Residual superoxide production	Undetectable	Undetectable	Low
Cytochrome b_{558} spectrum	Absent	Low	Low or normal
Type of mutations in *CYBB*	Deletions, insertions, splice site mutations, missense mutations, nonsense mutations	Missense mutation, especially involving amino acids 310–587	Missense mutations, especially involving amino acids 1–309

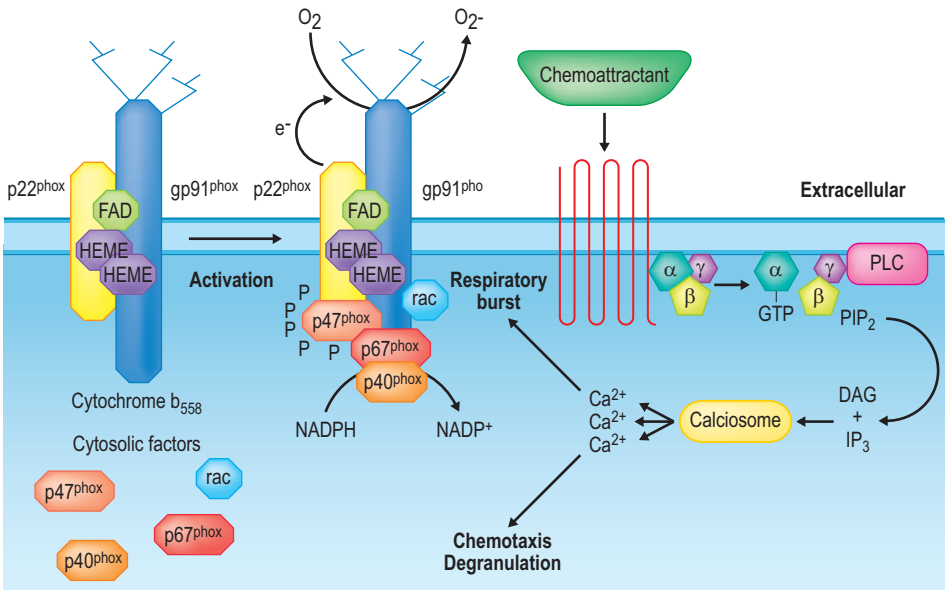


FIG. 39.4 Schematic Representation of the Nicotinamide Adenine Dinucleotide Phosphate (*NADPH*) Oxidase System. Chemoattractants interact with their receptors on the neutrophil surface, leading to an increase in intracellular calcium concentration. This activation results in the assembly of the NADPH oxidase complex following phosphorylation of cytosolic factors. This, in turn, leads to superoxide production. *DAG*, diacylglycerol; *FAD*, flavin adenine dinucleotide; *PIP_2*, phosphatidylinositol bisphosphate; *IP_3*, inositol triphosphate; *α, β, γ*, subunits of the guanosine triphosphate (*GTP*)–coupled receptors.

65% of cases and the autosomal recessive p47^{phox} (*ph*agocyte *ox*idase) deficiency for about 25%.

Patients with CGD often present with pneumonia, liver abscess, skin infections, lymphadenitis, or osteomyelitis; bacteremia is relatively uncommon. Initial presentation with immune dysregulation, particularly IBD, is not unusual. Exuberant tissue granuloma formation at the sites of infection, at surgical wounds, and in hollow viscera is a frequent problem seen primarily in patients with X-linked CGD.

The NADPH Oxidase and Its Activity

The nicotinamide adenine dinucleotide phosphate (NADPH) oxidase is a multicomponent system that transfers an electron to molecular oxygen by way of flavin adenine dinucleotide (FAD) and heme to form superoxide (O_2^-) (Fig. 39.4). Cytochrome b_{558} is a membrane-bound heterodimer lodged in the wall of the secondary granules; the large glycosylated β subunit is gp91^{phox} and the small nonglycosylated α subunit is p22^{phox}. The cytoplasmic tail of gp91^{phox} binds FAD, heme, and NADPH, which are required for electron transfer to oxygen (O_2). Neutrophil stimulation leads to aggregation and phosphorylation of p47^{phox}, p67^{phox}, p40^{phox}, and the small guanosine triphosphate (GTP)–binding proteins RAC1/RAC2, which dock with the cytochrome at the membrane through binding of p47^{phox} and p22^{phox}.

Much of the killing effect of neutrophils is, in fact, carried out by proteases enhanced by NADPH oxidase activity.[25] Charge created by electron flux across the membrane is compensated mostly by K^+ flux, which enhances microbial killing. Papayannopoulos and Zychlinksy identified neutrophil extracellular traps (NETs; extruded DNA with attached antimicrobial peptides), which depend on superoxide generation and are deficient in CGD.[26]

Mutations Leading to CGD

X-Linked CGD

The most common form of CGD is caused by mutations in *CYBB*, which encodes gp91^{phox} (located at Xp21.1) (see Table 39.3). Mutation types include deletions (22.2%), insertions (7%), deletion/insertion (1.5%), nonsense (29.8%), missense (19.4%), splice sites (19.5%), and promoters (0.6%).[27] The sporadic mutation rate is approximately 11%. Large interstitial deletions may include adjacent telomeric genes as well, leading to complex phenotypes, such as CGD along with McLeod syndrome (*KX*, or Kell antigen deletion), Duchenne muscular dystrophy (*DMD*), and X-linked retinitis pigmentosa (*RPGR*).[27] McLeod syndrome includes absent erythrocyte Kx protein and diminished levels of Kell blood group antigens. In patients with

McLeod syndrome, anti-Kx antibodies are formed when transfusions are given, making future transfusions extremely difficult. Patients with these large deletions may eventually develop progressive neurodegenerative symptoms, such as areflexia, dystonia, and choreiform movements. Deletions centromeric to *CYBB* may cause ornithine decarboxylase deficiency along with CGD.[28]

Autosomal Recessive CGD

Mutations in p47^{phox} (*NCF1*, located at 7q11.23) cause the majority of the recessive cases of CGD, around 25%, usually caused by homozygous deletions of the canonical GT splice site at the start of exon 2.[29] p22^{phox} (*CYBA*, located at 16q24) and p67^{phox} (*NCF2*, located at 1q25) are responsible for less than 5% of CGD cases each. p40^{phox} (*NCF4*, located at 22q13.1) deficiency has been reported in a boy with early-onset severe granulomatous fistulizing colitis without a significant infectious phenotype.[30] Essential for reactive oxygen species (EROS) is encoded by *CYBC1* and is required for transport of the gp91^{phox}/p22^{phox} complex to the cell surface; recessive defects in CYCB1 also result in CGD[30a]. No autosomal dominant cases of CGD have been identified.

Clinical Manifestations of CGD

The first severe infection usually occurs in infancy or childhood but can also occur in adulthood. Later diagnoses usually are seen in patients with residual superoxide production, either hypomorphic gp91^{phox} or p47^{phox} deficiency.[27,31] The constellation of signs and symptoms that suggest CGD range from failure to thrive, to IBD, visceral abscesses, recurrent sinopulmonary infections, and characteristic infections—most commonly pneumonia, lymphadenitis, liver abscess, skin abscess, perianal abscess, and osteomyelitis.[32] As in other neutrophil defects, the most common pathogen is *S. aureus*. Characteristic infections are caused by catalase-positive organisms such as *S. aureus, Burkholderia cepacia* complex*, Serratia marcescens, Nocardia* spp., and *Aspergillus* spp. (Fig. 39.5).

Staphylococcal liver abscesses in CGD are dense and necrotic and cause significant morbidity. Their fibrocaseous consistency makes percutaneous drainage difficult, and open surgery was formerly often performed. However, combined steroid and antibiotic therapy of CGD liver abscess has now become the preferred approach in many centers.[33] Liver involvement leading to portal hypertension is the likely cause of the splenomegaly commonly seen in CGD and is also closely tied to mortality.[34]

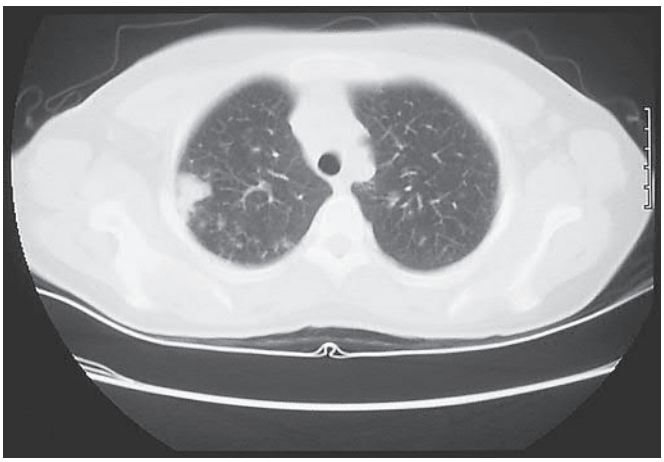


FIG. 39.5 Computed Tomography (CT) Scan of the Lungs of a Patient with Chronic Granulomatous Disease (CGD) and *Aspergillus* Pneumonia. *Aspergillus* pneumonia is often a peripheral consolidation in the lung parenchyma.

Pulmonary aspergillosis remains a major cause of death in CGD.[35] *Aspergillus fumigatus* is the most commonly isolated fungus, but it is successfully treated with azole antifungals. In contrast, severe disease in CGD is caused by *Aspergillus nidulans*, *Aspergillus viridinutans*, and other *Aspergillus* non-*fumigatus* species with low pathogenicity in the normal host.[35,36] Surgical resection of these infections is often required. *Aspergillus* infections in CGD are often unaccompanied by fever and leukocytosis.[35] Acute diffuse pulmonary fungal infection in CGD is referred to as "mulch pneumonitis," characterized by fever, hypoxia, and diffuse pulmonary infiltrates caused by inhalation of fungi, typically during mulching, leaf raking, or gardening.[37] This syndrome can be the initial presentation of CGD in older children and adults and is important to recognize, since it best responds to a combination of antifungals and steroids.

Septicemia is relatively uncommon but may occur with *B. cepacia* complex and *Chromobacterium violaceum. Granulibacter bethesdensis* is a gram-negative rod that causes chronic necrotic lymph node and spleen involvement pathognomonic for CGD.[38]

Inflammatory granulomata are a hallmark of CGD. Pyloric outlet obstruction, bladder outlet obstruction, and ureteral obstruction are common. Crohn-like IBD affects between 30% and 50% of patients, predominantly those with the X-linked form, and may involve the esophagus (Fig. 39.6), jejunum, ileum, cecum, rectum, and perirectal area.[39] GI manifestations can include diarrhea, malabsorption, abdominal pain, growth delay, and hypoalbuminemia. The median age of initial GI manifestations is 5 years, and abdominal pain is common. Interestingly, GI involvement has no effect on mortality, is not associated with liver disease, and is unaffected by the use of interferon-γ (IFN-γ).[32,39]

Granulomata respond very well to steroids and often require a slow taper over several weeks to months. Exuberant formation of granulation tissue and dysregulated cutaneous inflammatory responses lead to wound dehiscence and impaired wound healing (Fig. 39.7). Autoimmune and rheumatological problems have been reported at high rates in patients with CGD as compared to the general population.[40]

A comprehensive study of 287 patients with CGD from 244 kindred correlated the production of reactive oxygen intermediates with survival.[32] Patients with residual superoxide production had better long-term survival compared than those without residual superoxide production. Confirming the importance of this association, there was a direct correlation between the degree of superoxide and survival. Consistent with their previously recognized milder disease and better survival, patients with mutations in p47^{phox} had significant residual superoxide production. For those with gp91^{phox} mutations, the findings were more surprising. Patients with X-linked CGD with residual superoxide production were those with missense or splice mutations in the first 309 amino acids of gp91^{phox}. Those with missense mutations involving amino acids 310 to 587 had no residual superoxide production, regardless of protein levels. Therefore, identification of the specific molecular subtype of CGD and specific mutation has important implications for morbidity and survival. Interestingly, mortality curves

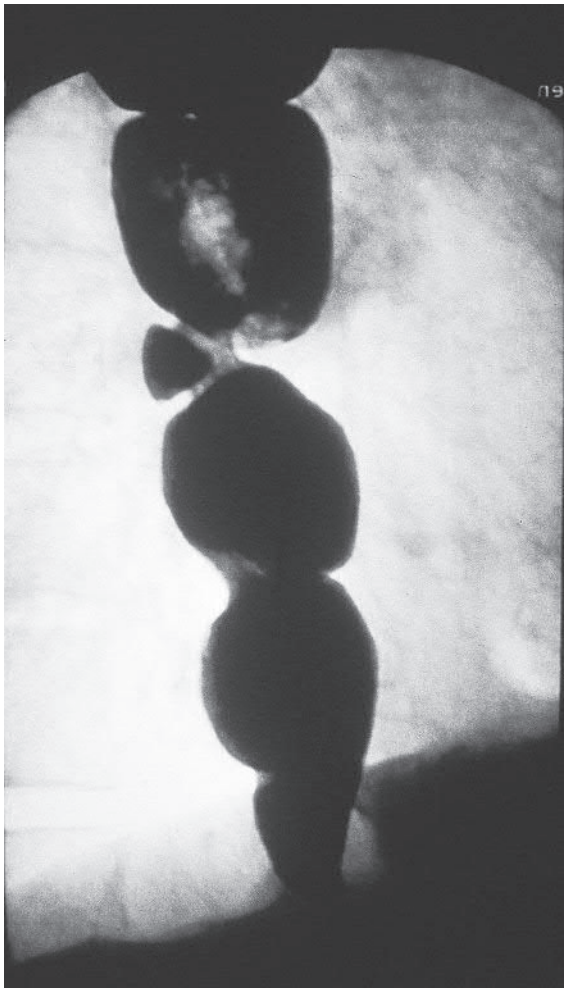


FIG. 39.6 Esophageal Involvement in Chronic Granulomatous Disease (CGD). Esophageal strictures caused by granuloma formation, as shown by barium swallow.

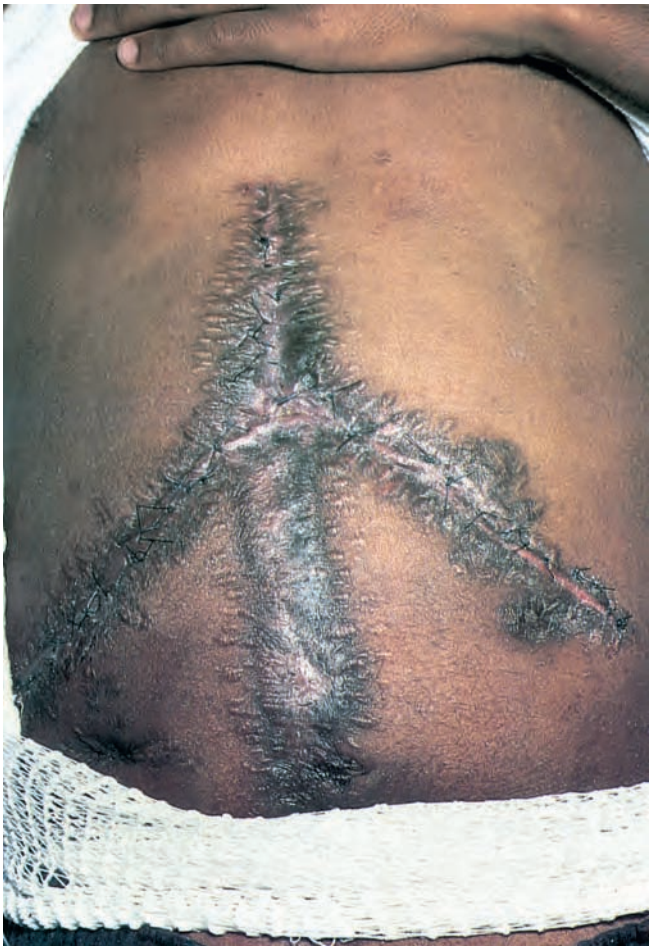


FIG. 39.7 Exuberant Granuloma Formation in Chronic Granulomatous Disease (CGD). Wound dehiscence and impaired wound healing at surgical incision sites as a result of dysregulated inflammatory responses in a patient with X-linked CGD.

did not diverge until after age 20 years, suggesting that residual superoxide production determines later toxicities, such as liver dysfunction, rather than early childhood mortality from infection. It is critical to keep in mind that this comprehensive study included data from patients followed for up to 30 years—that is, a significant number of patients were born before the advent of modern antimicrobials. Therefore, survival of a child born today who receives ideal management will probably exceed that in the study population above.

CLINICAL PEARLS

Chronic Granulomatous Disease

- Chronic granulomatous disease (CGD) comprises six inherited disorders with a relatively consistent phenotype.
- Major problems in CGD are infections with catalase-positive bacteria and fungi and formation of granulomata in the gastrointestinal and urinary tract.
- Oral prophylactic antibiotics and subcutaneous interferon (IFN)-γ injections three times a week are currently recommended for CGD.
- Diagnosis can be made via nitroblue tetrazolium (NBT) test or dihydrorhodamine (DHR) assay, the latter being a more sensitive diagnostic tool.
- Bone marrow transplantation is highly effective and can be curative.

Diagnosis of CGD

The diagnosis of CGD is most easily established by the dihydrorhodamine (DHR) assay, which measures the hydrogen peroxide–dependent conversion of DHR 123 to rhodamine 123, which is accompanied by fluorescence. This test is relatively reproducible, and a quantitative DHR index, obtained using flow cytometry that with residual superoxide production capacity. Other assays include nitroblue tetrazolium (NBT) reduction and dichlorofluorescein (DCF), but these older methods are more complicated and subjective (Fig. 39.8). One important false positive to keep in mind in DHR testing is myeloperoxidase (MPO) deficiency, which gives a DHR result consistent with CGD; however, superoxide production measured by NBT or the more specific ferricytochrome c reduction is normal to increased.

Treatment of CGD

Prophylactic trimethoprim–sulfamethoxazole (TMP-SMX) significantly reduces the frequency of bacterial infections in CGD, especially those caused by *S. aureus*. TMP-SMX prophylaxis is ineffective against fungal infections but does not encourage them. Prophylactic itraconazole prevents fungal infections. IFN-γ is beneficial as a prophylactic treatment in

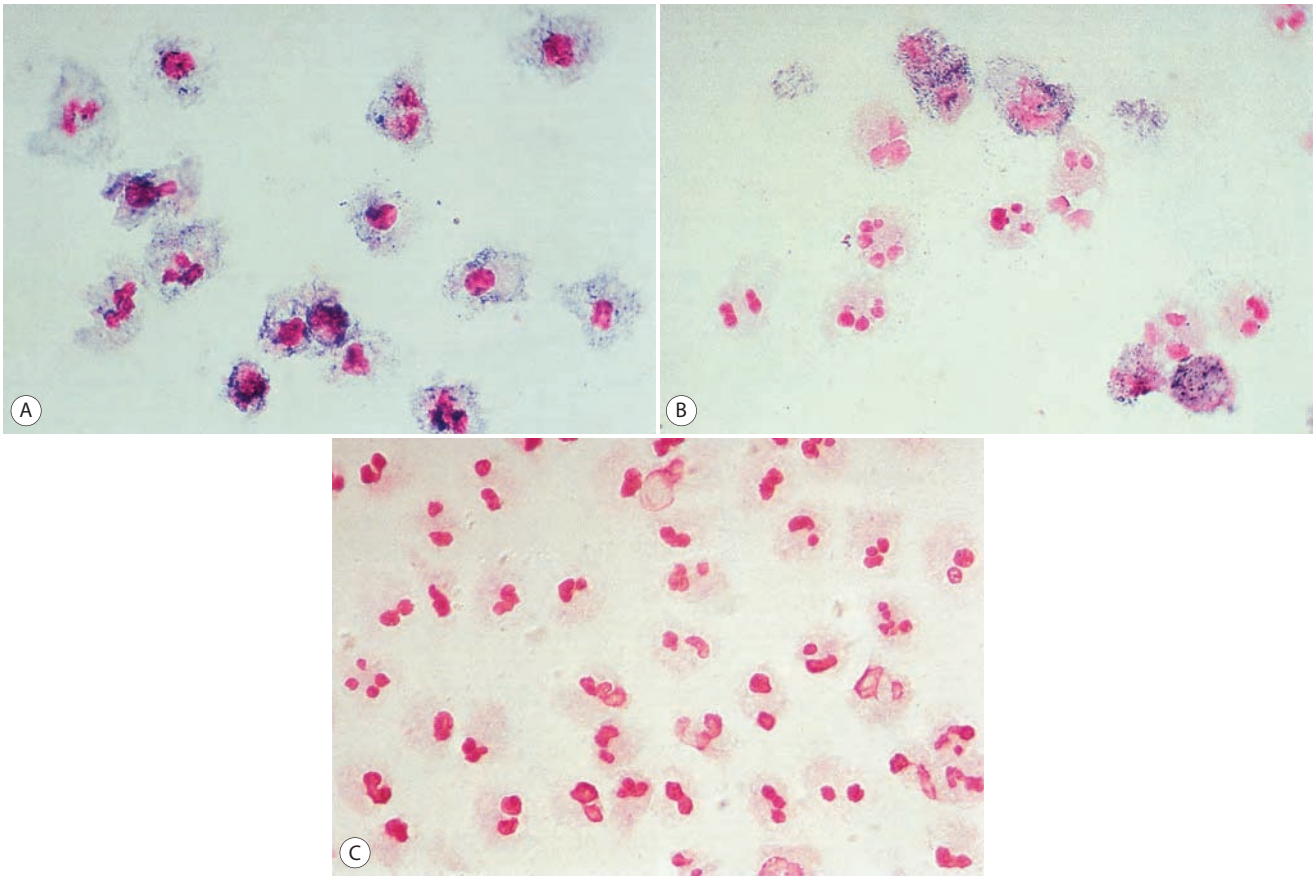


FIG. 39.8 Laboratory Diagnosis of Chronic Granulomatous Disease (CGD) with the Nitroblue Tetrazolium (NBT) Test. (A) Nitroblue tetrazolium reduction (NBTR) by purified normal neutrophils following stimulation with phorbol esters and calcium ionophore. NBT is reduced by all neutrophils, showing a *blue/purple deposit*. (B) NBTR by purified neutrophils from an X-linked CGD carrier; two different populations of cells are seen. Normal (unaffected cells) reduce the NBT dye and stain *blue/purple*, whereas affected cells fail to reduce the NBT dye and appear *clear*. (C) Neutrophils from a patient with CGD fail to reduce the NBT dye and appear *clear*. (Courtesy of Dr. Douglas B. Kuhns, Leidos, Frederick, MD.)

CGD. In a multicenter, placebo-controlled trial of IFN-γ, the number and severity of infections were significantly reduced by IFN-γ. The exact mechanism of action of IFN-γ is not known, but it has multiple effects, including stimulation of components of NADPH oxidase in partial deficiencies, increased bactericidal activity through neutrophil granule components, and Fc receptor expression. Subcutaneous administration of recombinant IFN-γ three times a week at a dose of 50 μg/m^2 (for those with body surface area >0.5 m^2) is recommended. Adverse effects of recombinant IFN-γ in patients with CGD include fever, chills, headache, flu-like symptoms, and diarrhea. During severe infections, leukocyte transfusions are sometimes used in addition to antibiotics, but this approach may lead to alloimmunization, compromising future bone marrow transplantation opportunities.

Because CGD is predominantly a hematopoietic disorder, bone marrow transplantation can cure CGD and has been successfully performed even in the setting of active infection.[41] The type of transplant that is used in patients with CGD varies among centers, but transplants using either fully myeloablative or partially myeloablative (reduced-intensity) conditioning have been effective. Although active infection is a relative contraindication for bone marrow transplantation overall, there are certain infections in CGD, especially those caused by atypical *Aspergillus* spp. infections, that are not curable with standard antifungal therapy. Bone marrow transplantation prevents not only recurrent life-threatening infections but also GI disease and growth retardation and is currently successful in about 90% of cases. Gene therapy for CGD appears to be effective but is still limited by the need for a preparative regimen to make space in the bone marrow for gene-corrected stem cells. Durability of the transduced cells remains to be demonstrated, and this treatment is not yet widely available.

MYELOPEROXIDASE DEFICIENCY

MPO is a heme-containing enzyme necessary for the conversion of hydrogen peroxide (H_2O_2) to hypochlorous acid (HOCl). MPO is expressed early in myeloid differentiation and resides in the azurophilic granules of neutrophils and the lysosomes of monocytes.[42] Mature MPO is a symmetrical molecule of four peptides, with each half consisting of a heavy–light chain heterodimer. Neutrophils of individuals with MPO deficiency fail to produce HOCl upon stimulation, whereas the NADPH oxidase system remains unaffected. Prolonged supranormal levels of superoxide and H_2O_2 production follow stimulation in MPO-deficient neutrophils. This may result from lack of negative feedback regulation of HOCl on the NADPH oxidase, although

the exact mechanism is unknown. MPO deficiency can be primary (congenital) or secondary (acquired).

Primary MPO Deficiency

Primary MPO deficiency is the most common phagocyte defect with a frequency of 1/4000 births. Both total and partial MPO deficiencies have been described. Patients with primary MPO deficiency do not usually have increased infections, probably because MPO-independent mechanisms compensate for the lack of MPO-dependent microbicidal activity. Visceral candidiasis occurring with concurrent diabetes has been reported in some patients. However, the frequency of such cases is very low. Affected individuals may develop nonfungal infections, malignancies, and certain skin disorders. In several cohorts of patients with complete MPO deficiency, an increased incidence of solid or hematological tumors has been observed.[42] MPO-deficient neutrophils have no apparent defect in the phagocytosis of bacteria or fungi, but microbicidal activity is slower than normal. MPO-deficient neutrophils are severely impaired in killing *Candida* spp. or *Aspergillus* spp. in vitro despite the fact that most patients with MPO deficiency do not develop significant fungal infections. This suggests that the mucosal barrier to fungal infection is independent of MPO activity and is able to prevent invasive infection.

The most common mutation is a missense replacement of arginine 569 with tryptophan (R569W), causing maturational arrest of the MPO precursor and preventing heme incorporation. Most patients are compound heterozygotes. The diagnosis of MPO deficiency is made by using anti-MPO monoclonal antibodies (mAbs) in flow cytometric analysis of neutrophils. No MPO expression is seen in congenital deficiency, whereas near-normal antigenic reactivity may be seen with the acquired form. Maintenance antibiotic or antifungal therapy is not routinely recommended. Prompt and prolonged therapy is advised in patients with diabetes mellitus and congenital MPO deficiency, as they may develop localized or systemic infections.

Secondary or Acquired MPO Deficiency

In the majority of patients, MPO deficiency is partial and transient. Secondary MPO deficiency can be seen in some hematological malignancies or disseminated cancers, exposure to cytotoxic agents or antiinflammatory medications, iron deficiency, lead intoxication, thrombotic diseases, renal transplantation, and pregnancy. MPO activity in bone marrow myeloid precursors as well as peripheral blood cells may vary from cell to cell. Successful treatment of the underlying condition typically corrects the defect. Secondary MPO deficiency is most likely linked to somatic mutations in the case of malignancy or toxic–metabolic effects on MPO activity.

Specific Granule Deficiency

Neutrophil-specific granule deficiency (SGD) is a rare disorder of leukocyte maturation in which neutrophil secondary granules and some primary granule proteins are absent as a result of mutations in CCAAT/enhancer binding protein epsilon C/EBPε (encoded by *CEBPE* located at 14q11.2), a member of the leucine zipper family of transcription factors. SGD is characterized by frequent, severe pyogenic infections, a paucity or absence of neutrophil-specific granule proteins and defensins, and atypical neutrophil nuclear structure with mostly bilobed nuclei. In vitro, these patients' cells show diminished neutrophil migration, reduced staphylococcal killing, reduced phagocytosis, and increased cell surface-to-volume ratio. Eosinophils and platelets are also affected. Platelets lack high-molecular-weight (HMW) von Willebrand factor multimers and have reduced platelet fibrinogen and fibronectin due to diminished platelet α granules. Bleeding diatheses and neutrophil phagocytosis of platelets are seen in SGD. In addition, SGD eosinophils are deficient in the eosinophil-specific granule proteins eosinophil cationic protein (ECP), eosinophil-derived neurotoxin (EDN), and major basic protein (MBP) despite the presence of messenger RNA (mRNA) transcripts for these proteins. An autosomal dominant form of SGD is caused by heterozygous mutations in *CEPPE*. Few patients have been reported to have survived beyond adolescence except for those with milder disease. SMARCE2 controls the expression of *CEBPE* mRNA, and mutations in the *SMARCD2* gene also cause SGD.[43] Few patients have been reported to have survived beyond adolescence except for those with a milder dominant form. Bone marrow transplantation should be considered early in the course of the disease.

KEY CONCEPTS

Specific Granule Deficiency

Specific granule deficiency (SGD) is caused by promyelocyte–myelocyte transition block as a result of a mutation in the *C/EBPε* gene.

- Pathological findings in SGD granulocytes include absence of secondary granule proteins and selective loss of the primary granule defensins.
- The prognosis is very poor in recessive forms of SGD.

CHEDIAK-HIGASHI SYNDROME

Chediak-Higashi syndrome (CHS) is a rare autosomal recessive disorder characterized by partial oculocutaneous albinism, increased susceptibility to infections, deficient NK-cell activity, and abnormal giant primary granules in neutrophils.[43] This immunodeficiency was first reported by Beguez-Cesar in 1943 and then further elaborated by Chediak and Higashi a decade later. The hallmark of CHS is giant abnormal granules in all granule-containing cells, including melanocytes (melanosomes are members of the lysosomal lineage of organelles), neutrophils, central and peripheral neural tissues, fibroblasts, and hair. The underlying defect is the inability to form appropriate lysosomes and cytoplasmic granules. CHS granulocytes lack cathepsin G and elastase, but the defensin content is normal. The giant granules of CHS are derived predominantly from azurophilic granules. CHS is classically a biphasic immunodeficiency in which the first phase is susceptibility to infection and the second is an accelerated lymphoproliferative syndrome with histiocytic infiltration of various tissues. Rarely, the accelerated phase may be the initial presentation. The giant organelles are derived from the late compartments of the endocytic pathway, affecting specifically late endosomes and lysosomes with minimal or no effect on early endosomes. *CHS1* encodes a 3801 amino acid lysosomal transporter (LYST), which has a vital role in lysosomal trafficking. Lysosomal exocytosis triggered by membrane wounding is impaired in Chediak-Higashi fibroblasts. The reduced survival of fibroblasts after wounding indicates that impaired lysosomal exocytosis inhibits membrane resealing. Inability of cells to repair plasma membrane lesions may contribute to the pathology of CHS. The degree of albinism can

vary from a slightly diluted skin pigment to hypopigmented skin and hair, photophobia, nystagmus, strabismus, macular hypoplasia, and reduced visual acuity. Skin biopsy shows large irregular melanin granules in melanocytes. Microscopic analysis of hair also shows poor distribution of melanin. Pancytopenia, neutropenia, and lack of NK-cell cytotoxicity result in frequent pyogenic infections, usually caused by staphylococci or streptococci. Hepatosplenomegaly and lymphadenopathy are common. A mild bleeding diathesis results from platelet storage pool deficiency. Neurological findings, including intellectual disability, seizures, cranial nerve palsies, and progressive peripheral neuropathy, have been noted in CHS.

The lymphoma-like lymphohistiocytic accelerated phase is characterized by increased hepatosplenomegaly, lymphadenopathy, and worsened pancytopenia, which may resemble the virus-associated hemophagocytic syndromes or familial hemophagocytic lymphohistiocytosis. Although chemotherapy can induce transient remissions, relapses are common. Bone marrow transplantation prevents the accelerated phase and restores NK-cell function, but it does not resolve the central or peripheral nervous system abnormalities. Demonstration of giant azurophilic cytoplasmic inclusions on peripheral blood smear is very suggestive of the diagnosis of CHS; mutation analysis confirms the diagnosis.

HYPER-IgE RECURRENT INFECTION, OR JOB SYNDROME

This syndrome was first published as hypoinflammatory recurrent infections with severe eczema and called "Job syndrome" by Ralph Wedgewood and colleagues in 1966. In 1972, Rebecca Buckley and coworkers recognized the IgE elevation that is characteristic of this disease. This multisystem autosomal dominant disorder is caused by heterozygous mutations in the gene encoding signal transducer and activator of transcription 3 (*STAT3*, located at 17q21).[44] Mutations in *STAT3* are mostly missense and are clustered in either the DNA-binding domain or Src homology 2 (SH2) domains of STAT3 Hyper-IgE recurrent infection syndrome (HIES, or Job syndrome) is characterized by recurrent infections of the lower respiratory tract and skin, chronic eczema, arterial anomalies, including coronary arterial tortuosity and aneurysms, extremely elevated IgE levels, and eosinophilia (Table 39.4). HIES occurs in all racial and ethnic groups.

TABLE 39.4 Clinical and Laboratory Findings in Patients With the Hyper-IgE Syndrome

Findings	Incidence (%)
Eczema	100
High IgE levels (>2000 IU/mL)	97
Eosinophilia (>2 SD above the mean for normals)	93
Boils	87
Pneumonia	87
Mucocutaneous candidiasis	83
Characteristic facies (in those ≥16 years)	83
Lung cysts	77
Scoliosis (for those ≥16 years)	76
Hyperextensible joints	68
Delayed shedding of primary teeth	72
Bone fractures	57

Adapted from Grimbacher B, Holland SM, Gallin JI, et al. Hyper-IgE syndrome disorder. *N Engl J Med.* 1999;340:692.

Facial, Skeletal, and Dental Abnormalities

Facial differences seen in the majority of the patients are a protruding, prominent mandible and forehead, apparent ocular hypertelorism, a broad nasal bridge, and a wide, fleshy nasal tip with increased interalar distance (Fig. 39.9). A high-arched palate is also common, as are skeletal abnormalities. Grimbacher et al. noted pathological bone fractures in 57% and scoliosis in 76% (Fig. 39.10). Low bone density and cortical bone loss are also seen, but not clearly correlated with the rate of bone fracture. Other infrequent skeletal abnormalities in HIES are craniosynostosis, spina bifida, bifid rib, wedge-shaped lumbar vertebra, hemivertebra, and pseudoarthritis of the hip. Joints are hyperextensible. A unique dental abnormality in HIES is retention of primary teeth, causing delayed eruption of permanent teeth.

KEY CONCEPTS

Hyper-IgE Recurrent Infection Syndrome or Job Syndrome

- Recurrent pneumonias and skin infections, chronic eczema, extremely elevated immunoglobulin E (IgE) levels, and eosinophilia are the hallmarks of the syndrome.
- Facial, skeletal, and dental abnormalities are very common.
- Lung abscesses and pneumatoceles following pneumonias caused by *Staphylococcus aureus* and *Haemophilus influenzae* cause major morbidity.

Infections and Immunological Characteristics

Moderate to severe eczema presenting within the first hours to weeks of life is almost universal in HIES. Mucocutaneous candidiasis involving fingernails and toenails, mouth, vagina, and intertriginous areas is seen in most patients. Primary pulmonary infections are caused by *S. aureus, Haemophilus influenzae*, and *Streptococcus pneumoniae*. These pneumonias are

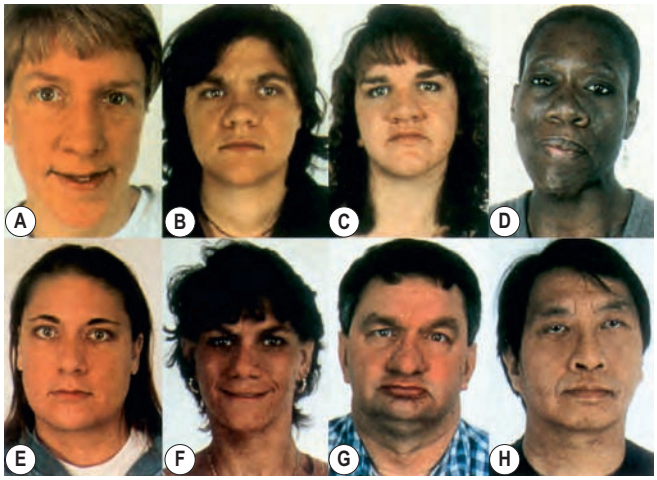


FIG. 39.9 Facial Abnormalities Seen in Patients with Hyper-IgE Recurrent Infection Syndrome (HIES). Prominent mandible and forehead, apparent hypertelorism, broad nasal bridge with a wide nasal tip, and increased interalar distance are commonly seen facial features of HIES. (With permission from Grimbacher B, Holland SM, Gallin JI, et al. Hyper-IgE syndrome with recurrent infections—an autosomal dominant multisystem disorder. *N Engl J Med.* 1999;340:692.)

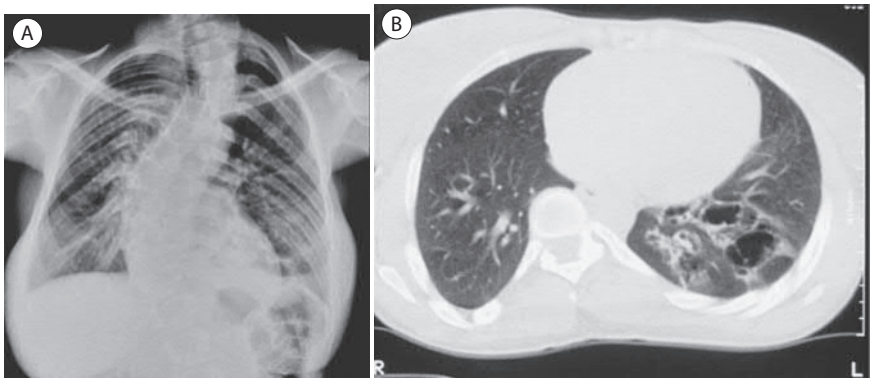


FIG. 39.10 Thoracic Pathology in Hyper-IgE Recurrent Infection Syndrome (HIES). (A) Chest X-ray of a patient with scoliosis. (B) Computed tomography (CT) scan of the lungs in the same patient demonstrates multiple pneumatoceles caused by prior infections.

often associated with abscess formation and usually lead to the development of pneumatoceles (see Fig. 39.10). Once lung cavities are formed, they provide an attractive environment for superinfection with *Pseudomonas* or *Aspergillus* spp. The clinical morphotype suggests abnormal tissue remodeling. *Pneumocystis jiroveci* pneumonia, cryptococcosis, histoplasmosis, and coccidioidomycosis have been reported. IgE levels are usually above 2000 IU/mL, but substantial fluctuations in IgE levels have been recorded over time, and the IgE levels do not correlate with disease activity or eosinophilia. Total serum IgG levels are usually within the normal range. Eosinophilia is common; the white blood cell (WBC) count is usually normal to low.

Mutations in *STAT3* lead to disruption of cytokine signaling, including IL-6, -10, -11, and -17 and -23. STAT3 deficiency leads to elevated tumor necrosis factor (TNF) and IFN-γ, but reduced IL-17 producing T cells (Th17 cells). This latter defect may explain the predisposition to mucocutaneous candidiasis.[45] Numbers of memory B and T cells are low. Shingles vesicles are increased despite normal recovery from primary varicella-zoster virus (VZV) infections. Frequency of lymphomas, but not epithelial malignancies, are increased.

In contrast to the dominant negative STAT3 mutations causing HIES gain-of-function mutations in STAT3 associated with high IL-6 response have recently been identified in children with failure to thrive, arthritis, livedo, and lung disease.

DOCK8 Deficiency (Autosomal Recessive Hyper-IgE Syndrome)

Several distinct diseases have elevated IgE, eczema, and eosinophilia. Deficiency of dedicator cytokinesis 8 (encoded by the gene *DOCK8*, located at 9p24) includes food allergies, asthma, herpesvirus infections, human papilloma virus (HPV), and molluscum contagiosum infections, which are not part of dominant negative STAT3 deficiency. In addition to the infection susceptibility, patients with DOCK8 deficiency are predisposed to cutaneous and lymphoid malignancies.[46] Although Th17 cells are markedly diminished in patients with STAT3 deficiency, they are less severely reduced in those with DOCK8 deficiency (patients with HIES-like disease[45]). Transplantation in highly effective in DOCK8 deficiency and should be considered early in life.

GATA2 Deficiency (Monomac Syndrome)

The *GATA2* gene (located at 3q21.3) encodes an early hematopoietic transcription factor most active in myeloid development.[47] There are five disease names associated with defects in *GSTS2: mpnp<AC (*monocytopenia and mycobacterial disease). DCML (dendritic cell, monocyte, B and NK deficiency), familial AML/MDS, Emberger syndrome, and NK-cell deficiency. In general, affected individuals develop late-childhood or adult-onset disseminated nontuberculous mycobacterial disease or disseminated fungal disease. Patients have absolute circulating monocytopenia, NK-cell cytopenia, and B-cell lymphopenia. Despite these circulating cytopenias, there are tissue macrophages and plasma cells, and Ig levels are normal to elevated. Neutrophils are variably affected, but persistent neutropenia is not uncommon. Other infections include HPV, molluscum contagiosum, histoplasmosis, and aspergillosis. Progressive pulmonary alveolar proteinosis is common, as are cytogenetic abnormalities of bone marrow, such as trisomy 8 and monosomy 7. Lymphedema develops in a minority and may have onset in late childhood. Pediatric myelodysplasia is another common manifestation. Diagnosis is suspected on the basis of the infections and hematological abnormalities. Since most routine hematological laboratory studies allow quite low ranges of normal monocyte percentages, it is necessary to look at absolute no the abnormal results on hemography. Since most routine hematological laboratory studies allow quite low ranges of normal monocyte percentages, it is necessary to look at absolute monocyte numbers, which are often ≤10% of normal. Other presentations of this syndrome include aplastic anemia, acute myelogenous leukemia, lymphedema, and chronic myelomonocytic leukemia. Sequencing of *GATA2* is required for diagnosis. Transplantation is highly effective if done before leukemia has developed.[47]

ASSESSMENT OF NEUTROPHIL FUNCTION

Discrete abnormalities in neutrophil function lead to recurrent bacterial or fungal infections. Assays have been developed to interrogate those functions.[48] However, since neutrophils cannot be viably stored or frozen, samples are usually examined fresh with simultaneous normal volunteer controls. The techniques discussed here are reviewed in Chapter 39.

Isolation of Neutrophils

Most assays require neutrophils to be purified from other blood components. Blood is usually anticoagulated using either citrate or heparin (10 units/mL) tubes and maintained at 20°C to 25°C in polypropylene containers. Typically, 1 to 2 × 10^6 neutrophils can be isolated per milliliter of whole blood.

Neutrophil Adherence

The adhesive function of phagocytes is commonly assessed by passage of 1 mL of whole blood through a column filled with nylon wool. Adherence is measured as the difference in the ANC of the precolumn sample and of the sample after passage through nylon wool. Alternatively, isolated neutrophils can be induced to bind to plastic using a 96-well plate either uncoated or coated with fetal bovine serum, a ligand like ICAM-1, or a specific ECM protein, such as fibrinogen or fibronectin. Endothelial cell monolayers harvested from human umbilical veins can serve as a more physiological substrate for the measurement of cell adhesion. Isolated neutrophils are preloaded with the cell permeant, acetoxymethyl ester derivative of the fluorescent dye, calcein (calcein-AM). Nonspecific esterases in the cytosol cleave the ester linkage, trapping the fluorescent probe in the cytosol. The labeled neutrophils are added to each well and incubated in the absence or presence of phorbol myristate acetate (PMA) to promote adherence through activated integrins. At the end of the incubation, the wells are washed to remove nonadherent cells. The fluorescence of each well is determined with a fluorescent microplate reader and compared with the fluorescence of a control well with a fixed number of fluorescent cells. Under control conditions, fewer than 10% of the neutrophils adhere to plastic or to plastic coated with fetal bovine serum. Slightly more neutrophil adherence is observed on wells coated with fibrinogen. Treatment of normal neutrophils with PMA for 30 minutes results in the adherence of 100% of the neutrophils under all conditions. Adherence is abnormal in patients with LAD. Neutrophils isolated from patients with typical LAD-1 generally exhibit markedly reduced adherence under both unstimulated and PMA conditions.

Neutrophil Chemotaxis

Neutrophil chemotaxis in vivo can be evaluated by using skin windows. Skin blisters are gently raised on the volar surface of the forearm using a vacuum pump and a blister device, with little hemorrhage or vascular damage. The roof of the blister is removed, and the exposed dermis is bathed with autologous serum with the use of a skin window chamber. In 24 hours, exudative neutrophils accumulate in the autologous serum bathing the skin lesion. The skin chamber provides a mechanism for characterizing the immune cells as well as the soluble immune mediators that accumulate during the evolution of the inflammatory response. Chemotaxis *in vitro* is generally measured by using a Boyden chamber. The Boyden chamber includes three components: a lower (chemoattractant) chamber, a nitrocellulose or polycarbonate filter layer, and an upper cell chamber. The lower compartments are filled with a chemoattractant, such as fMLF; 10^{-8} M) or IL-8 (10 ng/mL). Rapid fluorescence-based assays of neutrophil chemotaxis uses a 96-well chemotaxis chamber and a fluorescence microplate reader.

Expression of Surface Antigens

The expression of cell surface antigens on neutrophils relies on labeled mAbs analyzed by flow cytometry. The panel may include the β_2 integrins (CD11a, CD11b, CD11c, and CD18), selectins (CD62L), Fcγ receptors I, II, and III (CD64, CD32, and CD16), leukosialin (CD43), the common leukocyte antigen (CD45), and markers for the specific granules (CD67), and azurophilic granules (CD63). The expression of surface antigens can be used to assess the responsiveness of neutrophils to particular ligands, such as fMLF and lipopolysaccharide (LPS).

Neutrophil Degranulation

The proteases, acid hydrolases, and inflammatory mediators released from storage granules in neutrophils can mediate bacterial killing, tissue damage, healing, and immune regulation. Lactoferrin from specific granules can chelate iron, resulting in a bactericidal or bacteriostatic effect. Stimulation of neutrophils with various secretagogues can release granular enzymes into the extracellular fluid. Treatment of the neutrophils with cytochalasin b (5 μg/mL) disrupts microfilament assembly and facilitates the release of both specific and azurophilic enzymes. To differentiate degranulation from cell lysis, release of the cytosolic enzyme lactate dehydrogenase should be monitored simultaneously. The release of azurophilic granules can be assessed by determination of β-glucuronidase activity. Supernatant fluids or cell extracts obtained from stimulated neutrophils are incubated with 4-methylumbelliferyl-β-D-glucuronide. Alternatively, MPO can be determined by using commercially available enzyme-linked immunoassays. CD63 is also found in the membrane of azurophilic granules and migrates to the neutrophil surface after stimulation with fMLF in the presence of cytochalasin b. The release of specific granules can be assessed by determination of lactoferrin levels using an enzyme-linked immunoassay. The carcinoembryonic antigen CD66b (formerly CD67) is found on the neutrophil surface and in the specific granules, and its expression on the surface of the neutrophils is increased after stimulation with fMLF or LPS. Detection of the constituents of secretory granules can be assessed by flow cytometric analysis of the change in expression of surface proteins, such as adhesion molecules, and cytochrome b_{558} of the NADPH oxidase. Neutrophil NET formation can also be measured in vivo and in vitro.

Generation of Reactive Oxygen Species

The production of O_2^- can be detected by using the reduction of cytochrome c. Because O_2^- causes a one-to-one stoichiometric reduction of ferricytochrome c to ferrocytochrome c, the resultant increase in the absorption spectrum at 550 nM can be used to quantitate the production of O_2^-. Superoxide dismutase is added to an identical tube to control for the nonspecific reduction of cytochrome c. However, since cytochrome is not permeable to the cells, the detection of O_2^- is limited to that released into the extracellular milieu. Neutrophils isolated from patients with CGD produce little O_2^- in response to PMA in 10 minutes. However, some patients with forms of CGD associated with residual superoxide production have low but detectable O_2^- production in 60 minutes. Neutrophils from X-linked heterozygous carriers of X-linked CGD can yield a full spectrum of O_2^- production, whereas neutrophils from autosomal recessive carriers of CGD generally yield a normal response. Although the detection of O_2^- by reduction of cytochrome c is useful in the diagnosis of patients with CGD, it cannot be used in the diagnosis of carriers because of the wide spectrum of responses that result from the random process of degree of X-chromosome inactivation or lyonization.

The extracellular release of H_2O_2 can be measured by using horseradish peroxidase–induced oxidation of either phenol red or Amplex red. Neutrophil suspensions in the presence of horseradish peroxidase and a chromophore are exposed to either PMA or buffer alone. Changes in optical density of phenol red at 600 nm can be determined with a standard microplate reader. With Amplex red—a much more sensitive fluorescent chromophore—H_2O_2-dependent changes in fluorescence can be determined with a fluorescence microplate reader.

The NBT test is a qualitative assay of O_2^- production and yields a visual record of the reduction of the NBT dye to insoluble, blue-black deposits of formazan. Whole blood or isolated neutrophils are mixed with NBT in a chamber slide and stimulated with PMA for 15 to 30 minutes at 37°C. The slide is counterstained with 0.1% safranin and examined under a microscope. Normal neutrophils, but not neutrophils from patients with CGD, reduce the yellow dye to black-brown-blue aggregates in the cells (see Fig. 39.8, A–C). The NBT test can be used to diagnose X-linked carriers of CGD but cannot differentiate autosomal carriers from normal subjects.

An alternative to the NBT test is a flow cytometric assay, using dihydrorhodamine-1,2,3 (DHR). Neutrophils are loaded with the nonfluorescent dye and then stimulated with PMA for 15 minutes at 37°C. H_2O_2 produced upon PMA stimulation oxidizes the DHR and results in increased fluorescence, detectable with a flow cytometer. Catalase is added to prevent cell-to-cell diffusion of H_2O_2. Since DHR is localized to the cytoplasm, and catalase is present in the extracellular fluid, the assay detects the intracellular production of reactive oxygen metabolites. Stimulation of normal neutrophils with PMA results in a two-log increase in the fluorescence intensity. The major advantages of the DHR assay are the sensitivity, the high signal-to-noise ratio, and ease of recording data from a large number of cells.

Western Blot for Measuring NADPH Oxidase

Determination of the protein expression in CGD by Western blot analysis provides direction for the genetic defect. A validated healthy control and a typical gp91^{phox} CGD sample are included on each blot to ensure adequate visualization of p22^{phox}. Patients with p47^{phox} CGD are Western blot negative. Patients with p67^{phox} CGD are generally Western blot negative. Because p22^{phox} and gp91^{phox} exist as a membrane complex, patients with a defect are p22^{phox}, in general, Western blot negative for both p22^{phox} and gp91^{phox}. In contrast, defects in gp91^{phox} yield more variable results. Neutrophils from patients with nonsense defects in gp91^{phox} exhibit low but detectable levels of p22^{phox}. Patients with missense mutations in gp91^{phox} that yield detectable gp91^{phox} protein exhibit proportionately higher levels of p22^{phox} and EROS proteins that can also be detected by Western blotting.

TRANSLATIONAL RESEARCH

The progress made over the past 60 years in understanding and managing phagocyte defects has been remarkable, spanning initial identification, phenotyping, and molecular characterization. The last 20 years have seen the advent of oral antifungals, potent oral antibiotics, oral antivirals, reduced intensity conditioning for bone marrow transplantation, and gene therapy. These advances have transformed the quality of life and longevity of all patients with immune deficiencies. We now need further in-depth study of the mechanisms that drive disease pathophysiology to gain novel insights. The advent of therapeutic cytokines, small molecule inhibitors and agonists, RNA inhibitors, and effective means of gene transfer should make directed approaches to disease modification a reality. However, we must be sure we know exactly what functions need addressing. The downstream effects of genetic defects are surprisingly complex and not always as straightforward as has been anticipated. More detailed clinical and functional phenotyping, careful examination of developmental and gene expression effects, and continued longitudinal studies of large cohorts will convert these severe diseases punctuated by acute, life-threatening infections into chronic conditions that are successfully treated medically or successfully managed until successful bone marrow transplantation becomes available.

ON THE HORIZON

- Early recognition and molecular diagnosis of all phagocyte defects, leading to prophylactic antimicrobial treatment, where indicated
- Improvement in bone marrow transplantation technology to allow for early, safe, successful, fertility-preserving transplantation in all cases
- Understanding the mechanisms of the hepatic complications of chronic granulomatous disease (CGD) that correlate with mortality
- Characterizing the complex somatic and immune pathways that utilize the signal transducer and activator of transcription 3 (STAT3) to understand the basis of hyper-IgE recurrent infection syndrome (HIES) and disorders arising from STAT3 gain-of-function mutations.

ACKNOWLEDGMENT

This work was supported by the intramural program of the National Institute of Allergy and Infectious Diseases.

REFERENCES

1. Kaushansky K. Lineage-specific hematopoietic growth factors. *N Engl J Med*. 2006;354:2034–2045.
2. Berliner N. Lessons from congenital neutropenia: 50 years of progress in understanding myelopoiesis. *Blood*. 2008;111:5427–5432.
3. Perdiguero EG, Geissmann F. The development and maintenance of resident macrophages. *Nat Immunol*. 2016;17:2–8.
4. Boztug K, Klein C. Genetics and pathophysiology of severe congenital neutropenia syndromes unrelated to neutrophil elastase. *Hematol Oncol Clin North Am*. 2013;27:43–60.
5. Kollner I, Sodeik B, Schreek S, et al. Mutations in neutrophil elastase causing congenital neutropenia lead to cytoplasmic protein accumulation and induction of the unfolded protein response. *Blood*. 2006;108:493–500.
6. Dale DC, Welte K. Cyclic and chronic neutropenia. *Cancer Treat Res*. 2011;157:97–108.
7. Klein C, Grudzien M, Appaswamy G, et al. HAX1 deficiency causes autosomal recessive severe congenital neutropenia (Kostmann disease). *Nat Genet*. 2007;39:86–92.
8. Germeshausen M, Grudzien M, Zeidler C, et al. Novel HAX1 mutations in patients with severe congenital neutropenia reveal isoform-dependent genotype-phenotype associations. *Blood*. 2008;111:4954–4957.
9. Boztug K, Appaswamy G, Ashikov A, et al. A syndrome with congenital neutropenia and mutations in G6PC3. *N Engl J Med*. 2009;360:32–43.
10. Boocock GR, Marit MR, Rommens JM. Phylogeny, sequence conservation, and functional complementation of the SBDS protein family. *Genomics*. 2006;87:758–771.
11. Huang JN, Shimamura A. Clinical spectrum and molecular pathophysiology of Shwachman-Diamond syndrome. *Curr Opin Hematol*. 2011;18:30–35.
12. Capsoni F, Sarzi-Puttini P, Zanella A. Primary and secondary autoimmune neutropenia. *Arthritis Res Ther*. 2005;7:208–214.
13. Maheshwari A, Christensen RD, Calhoun DA. Immune neutropenia in the neonate. *Adv Pediatr*. 2002;49:317–339.
14. Nauseef WM, Borregaard N. Neutrophils at work. *Nat Immunol*. 2014;15:602–611.
15. van de Vijver E, van den Berg TK, Kuijpers TW. Leukocyte adhesion deficiencies. *Hematol Oncol Clin North Am*. 2013;27:101–116.
16. Kishimoto TK, Hollander N, Roberts TM, et al. Heterogeneous mutations in the beta subunit common to the LFA-1, Mac-1, and p150,95 glycoproteins cause leukocyte adhesion deficiency. *Cell*. 1987;50:193–202.

17. Moutsopoulos NM, Konkel J, Sarmadi M, et al. Defective neutrophil recruitment in leukocyte adhesion deficiency type I disease causes local IL-17-driven inflammatory bone loss. *Sci Transl Med*. 2014;6: 229–240.
18. Roos D, Meischl C, de Boer M, et al. Genetic analysis of patients with leukocyte adhesion deficiency: genomic sequencing reveals otherwise undetectable mutations. *Exp Hematol*. 2002;30:252–261.
19. Moutsopoulos NM, Zerbe CS, Wild T, et al. Interleukin-12 and interleukin-23 blockade in leukocyte adhesion deficiency type 1. *New Engl J Med*. 2017;376:1141–1146.
20. Etzioni A, Gershoni-Baruch R, Pollack S, et al. Leukocyte adhesion deficiency type II: long-term follow-up. *J Allergy Clin Immunol*. 1998;102:323–324.
21. Alon R, Etzioni A. LAD-III, a novel group of leukocyte integrin activation deficiencies. *Trends Immunol*. 2003;24:561–566.
22. Malinin NL, Zhang L, Choi J, et al. A point mutation in KINDLIN3 ablates activation of three integrin subfamilies in humans. *Nat Med*. 2009;15:313–318.
23. Svensson L, Howarth K, McDowall A, et al. Leukocyte adhesion deficiency-III is caused by mutations in KINDLIN3 affecting integrin activation. *Nat Med*. 2009;15:306–312.
24. Marciano BE, Spalding C, Fitzgerald A, et al. Common severe infections in chronic granulomatous disease. *Clin Infect Dis*. 2015;60:1176–1183.
25. Segal AW. How neutrophils kill microbes. *Annu Rev Immunol*. 2005;23:197–223.
26. Papayannopoulos V, Zychlinsky A. NETs: a new strategy for using old weapons. *Trends Immunol*. 2009;30:513–521.
27. Roos D, Kuhns DB, Maddalena A, et al. Hematologically important mutations: X-linked chronic granulomatous disease (third update). *Blood Cells Mol Dis*. 2010;45:246–265.
28. Royer-Pokora B, Kunkel LM, Monaco AP, et al. Cloning the gene for an inherited human disorder—chronic granulomatous disease—on the basis of its chromosomal location. *Nature*. 1986;322:32–38.
29. Deardorff MA, Gaddipati H, Kaplan P, et al. Complex management of a patient with a contiguous Xp11.4 gene deletion involving ornithine transcarbamylase: a role for detailed molecular analysis in complex presentations of classical diseases. *Mol Genet Metab*. 2008;94:498–502.
30. Roos D, Kuhns DB, Maddalena A, et al. Hematologically important mutations: the autosomal recessive forms of chronic granulomatous disease (second update). *Blood Cells Mol Dis*. 2010;44:291–299.
31. Matute JD, Arias AA, Wright NA, et al. A new genetic subgroup of chronic granulomatous disease with autosomal recessive mutations in p40 phox and selective defects in neutrophil NADPH oxidase activity. *Blood*. 2009;114:3309–3315.
32. Kuhns DB, Alvord WG, Heller T, et al. Residual NADPH oxidase and survival in chronic granulomatous disease. *N Engl J Med*. 2010;363:2600–2610.
33. Leiding JW, Freeman AF, Marciano BE, et al. Corticosteroid therapy for liver abscess in chronic granulomatous disease. *Clin Infect Dis*. 2012;54:694–700.
34. Feld JJ, Hussain N, Wright EC, et al. Hepatic involvement and portal hypertension predict mortality in chronic granulomatous disease. *Gastroenterology*. 2008;134:1917–1926.
35. Segal BH, DeCarlo ES, Kwon-Chung KJ, et al. *Aspergillus nidulans* infection in chronic granulomatous disease. *Medicine (Baltimore)*. 1998;77:345–354.
36. Vinh DC, Shea YR, Jones PA, et al. Chronic invasive aspergillosis caused by *Aspergillus viridinutans*. *Emerg Infect Dis*. 2009;15:1292–1294.
37. Siddiqui S, Anderson VL, Hilligoss DM, et al. Fulminant mulch pneumonitis: an emergency presentation of chronic granulomatous disease. *Clin Infect Dis*. 2007;45:673–681.
38. Greenberg DE, Shoffner AR, Zelazny AM, et al. Recurrent *Granulibacter bethesdensis* infections and chronic granulomatous disease. *Emerg Infect Dis*. 2010;16:1341–1348.
39. Marciano BE, Rosenzweig SD, Kleiner DE, et al. Gastrointestinal involvement in chronic granulomatous disease. *Pediatrics*. 2004;114:462–468.
40. De Ravin SS, Naumann N, Cowen EW, et al. Chronic granulomatous disease as a risk factor for autoimmune disease. *J Allergy Clin Immunol*. 2008;122:1097–1103.
41. Güngör T, Teira P, Slatter M, et al. Reduced-intensity conditioning and HLA-matched haemopoietic stem-cell transplantation in patients with chronic granulomatous disease: a prospective multicentre study. *Lancet*. 2014;383:436–448.
42. Lanza F. Clinical manifestation of myeloperoxidase deficiency. *J Mol Med*. 1998;76:676–681.
43. Kaplan J, De Domenico I, Ward DM. Chediak–Higashi syndrome. *Curr Opin Hematol*. 2008;15:22–29.
44. Freeman AF, Holland SM. Clinical manifestations, etiology, and pathogenesis of the hyper-IgE syndromes. *Pediatr Res*. 2009;65:32R–37R.
45. Milner JD, Sandler NG, Douek DC. Th17 cells, Job's syndrome and HIV: opportunities for bacterial and fungal infections. *Curr Opin HIV AIDS*. 2010;5:179–183.
46. Zhang Q, Davis JC, Lamborn IT, et al. Combined immunodeficiency associated with DOCK8 mutations. *N Engl J Med*. 2009;361:2046–2055.
47. Spinner MA, Sanchez LA, Hsu AP, et al. GATA2 deficiency: a protean disorder of hematopoiesis, lymphatics, and immunity. *Blood*. 2014;123: 809–821.
48. Elloumi HZ, Holland SM. Diagnostic assays for chronic granulomatous disease and other neutrophil disorders. *Methods Mol Biol*. 2014;1124:517–535.

40

Complement and Complement Disorders

John P. Atkinson, M. Kathryn Liszewski, Anuja Java, and Hrishikesh S. Kulkarni

OVERVIEW: AN EVOLUTIONARY AND HISTORICAL PERSPECTIVE

The complement system arose early in evolution, possibly even in single-cell organisms. In vertebrates, it features a proteolytic cascade to coat microorganisms with cleavage fragments that are recognized by receptors on phagocytic cells.[1,2] This phenomenon, known as *opsonization*, leads to immune adherence followed by internalization. A related strategy employed by primates is to express a complement receptor on erythrocytes. In blood, the opsonized pathogen becomes adherent to a cluster of abundant erythrocytes, which, like a taxi, delivers them to the liver and/or the spleen for transfer to monocytes and macrophages to be neutralized. In this manner, bacteria are both immobilized and prevented from traveling free in the circulation to such sites as the brain.

To mediate opsonization, the complement system needed to develop a system to transfer a plasma component onto the pathogen. Lectins and, subsequently in evolution, antibodies (Abs) faced this same challenge. Of note, both of the latter eventually utilized the complement system (lectin pathway and classical pathway [CP]) to "complement" their attachment strategies. In contrast to lectins and antibodies, the alternative pathway (AP) earlier solved this problem through development of a thioester bond in C3. C3's homologous cousin, α_2-macroglobulin, also utilizes cleavage of a thioester bond to covalently attach to and thereby inactivate proteases. The parallels are apparent—upon its generation following C3 activation, C3b can transiently (within microseconds) attach to nearby hydroxyl or amino groups to form an ester or amide linkage, respectively. This mechanism creates essentially an almost unbreakable bond and places the complement fragment irreversibly on the pathogen's surface. The components of the complement system are listed by category in Table 40.1.

The major function of the complement system is to modulate the membrane of a microbial target, leading to immune adherence and internalization of the opsonized antigen. The target is selected by lectins in the lectin pathway and predominantly by antibodies in the CP. However, the ancient AP does not feature selective or specific recognition. Instead, 1% to 2% of C3 "ticks over" each hour, serving as a surveillance system. The activated C3 has a few microseconds to bind to a target, such as a bacterium, or it will be inactivated by water [now termed "C3(H_2O)"]. If it lands on a pathogen, the AP feedback loop engages and C3b is quickly generated. For example, this rapid amplification can deposit several million copies of C3b on a single *Escherichia coli* bacterium in 2 to 3 minutes. If C3b remains in the fluid phase, it is promptly inactivated by plasma regulators. If it binds to healthy self-tissues, it is inactivated by ubiquitously expressed membrane complement inhibitors. Thus, the AP initially comprised the original complement system and likely consisted of three elemental proteins (C3 and two proteases known as *factor B* and *factor D*). These three were sufficient to generate a C3 convertase (splitting) enzyme and also to form a feedback loop. The positive regulator of this remarkable enzymatic feedback/amplification loop, known as *properdin*, probably came later in evolution, as did lectins and antibodies.

Still later to develop was the membrane attack complex (MAC, often called the *terminal pathway*) that is common to all three cascades. The goal of its five sequentially interacting, nonenzymatic proteins (C5b, C6, C7, C8, and C9) is also to attach to and then alter the surface of a pathogen by membrane perturbation, often ending in lysis. In humans, a deficiency of any one of these five proteins leads to meningococcal infections. Interestingly, properdin deficiency also predisposes to recurrent meningococcemia. Thus, such infections likely drove specialization of this terminal wing of the complement system.

In contrast to the rapidly acting AP, an initial limitation with the lectin and Ab systems is that their triggers are present in limited quantities. It takes several days to ramp up synthesis of a particular lectin (acute phase type response) and at least a week to develop a specific immunoglobulin M (IgM) and even longer for an IgG immune response. For a host with an opportunistic pathogen invading the bloodstream, this time delay for an adaptive humoral immune response is far from adequate. Consequently, the complement system is often called the "guardian of the intravascular space." Once a "pumped" circulation of blood developed in evolution, a rapidly acting, abundant, and highly efficient system to prevent pathogens from entering, traveling, and reproducing in plasma was mandatory.

The second function of the complement system is to promote the inflammatory response. This is primarily accomplished by the anaphylatoxins C3a and C5a. Upon cleavage of C3 to C3b (the major complement opsonin) and C5 to C5b (the trigger of the MAC), the ~10 kilodalton (kDa) C3a and C5a fragments are released, and these can engage their respective receptors to initiate vascular and cellular changes, rapidly leading to a proinflammatory state. These receptors are expressed on many cell types, including endothelial, epithelial, and immune cells. Upon receptor engagement, defensive strategies are initiated and result in increased blood flow and stimulated phagocytes that are now more efficient at binding and ingesting C3 fragment-coated antigens.

Through these same interactions, the complement system instructs the adaptive immune response. Antigens decorated by complement proteins are taken up by monocytes, follicular-dendritic cells (FDCs), B lymphocytes, and other antigen-presenting cells (APCs), resulting in an adaptive immune response (the complement system is often called "nature's

TABLE 40.1 Proteins of the Complement System

Component	Function
Classical Pathway (CP)	
C1q	Part of C1. Binds to immunoglobulin M (IgM), IgG, pentraxins, and ligands on apoptotic cells to initiate CP activation.
C1r	Part of C1. After auto-activation, cleaves C1s.
C1s	Part of C1. After activation by C1r, cleaves C4 and C2.
C4	Cleaved by C1s to form C4a and C4b. C4b is an opsonin and part of the CP and LP C3 and C5 convertases. C4a is an untethered agonist for protease-activated receptor (PAR1) and PAR4.
C2	Binds to C4b and then cleaved by C1s to form C2a and C2b; C2b becomes part of the enzymatic component of the CP and LP C3 and C5 convertases. C2a is released.[a]
Lectin Pathway (LP)	
MBL	Recognition component for LP activation. Binds to mannose-rich glycans through C-type lectin domains.
MASP-1 and MASP-3	Associated with MBL and ficolins. Cleaves C2, but not C4. Cleaves profactor D.
MASP-2	Associated with MBL and ficolins. Cleaves C2 and C4.
Ficolins 1–3	Recognition components for LP activation. Bind to glycans through fibrinogen-like recognition domains.
Alternative Pathway (AP)	
C3	Cleaved by C3 convertases to form C3b and C3a. C3b is opsonic. A small fraction becomes part of the AP C3 convertase and part of all C5 convertases. C3b is further cleaved to opsonic iC3b and the CR2 ligands C3dg and C3d. C3a is an anaphylatoxin.
Factor B	Binds to C3b and then cleaved by factor D to form Bb, the enzymatic component of the AP C3 and C5 convertases. Ba is released.[b]
Factor D	Cleaves factor B bound to C3b to form AP convertases.
Properdin	Stabilizes AP convertases. Binds to microbial ligands to initiate AP activation.
Membrane Attack Complex (MAC)	
C5	Cleaved by C5 convertases to form C5b and C5a. C5b initiates MAC formation. C5a is an anaphylatoxin.
C6	Part of the MAC. Binds membranes.
C7	Part of the MAC. Binds membranes.
C8	Part of the MAC. Initiates pore formation.
C9	Part of the MAC. Polymerizes to form lytic pores.
Soluble Regulatory Proteins	
C1-INH	Serine protease inhibitor of C1r, C1s, MASP-1, MASP-2, kallikrein, factor XII.
C4BP	Binds C4b and prevents interaction with C2. Decay accelerating and cofactor activities for C4b-containing convertases.
FH	Binds C3b and polyanions. Prevents factor B binding. Has decay-accelerating and cofactor activities for C3b-containing convertases.
FI	Cleaves C3b and C4b bound to a cofactor protein.
Vitronectin	Binds C5b-7, prevents membrane insertion and lysis.
Clusterin	Binds C8 and C9, prevents MAC assembly and lysis.
Membrane Regulatory Proteins	
CD55 (DAF)	Accelerates decay of C3 and C5 convertases.
CD46 (MCP)	Cofactor for FI cleavage of C3b and C4b.
CD59	Binds to C8 and C9, prevents MAC assembly and lysis.
Receptors	
CD35 (CR1)	Opsonic receptor for C3b and C4b. Has decay-accelerating and cofactor activity for C4b and C3b and convertases containing these fragments.
CD21 (CR2)	Receptor for C3dg and C3d. Enhances B-cell activation.
CD11b/CD18 (CR3)	Opsonic receptor for iC3b. Leukocyte adhesion integrin.
CD11c/CD18 (CR4)	Opsonic receptor for iC3b. Leukocyte adhesion integrin.
CRIg	Opsonic receptor for iC3b and C3c. Inhibits C5 convertases.
C5aR (CD88)	Proinflammatory and chemotactic receptor for C5a.
C5L2	Receptor for C5a. Function not fully defined.
C3aR	Proinflammatory and chemotactic receptor for C3a.
C4aR	Cell activation; endothelial cell permeability.

[a]No defined function upon release.
[b]Several putative functions reported but requires further study.

adjuvant.") Thus, complement activation is required for optimal Ab responses to most foreign antigens. Individuals lacking C3 are predisposed at an early age to bacterial infections, predominantly by encapsulated bacteria.

Complement deficiencies are instructive anomalies of nature. Surprisingly, a complete deficiency in an early component of the classical complement pathway predisposes to autoimmune diseases, particularly systemic lupus erythematosus (SLE) (Chapter 52). Greater than 80% of patients with C1q or C4 deficiency present with SLE. This association indicates that the complement system is required not only for host defense against foreign agents but also to identify and safely

clear self-materials (debris or garbage removal), particularly RNA and DNA species.

Much complement-mediated pathology revolves around disturbances of the AP and thus its potent amplification loop. It must be rigorously regulated to prevent activation on normal self and excessive activation on injured self. Even haploinsufficiency of its major inhibitors predisposes to endothelial damage in atypical hemolytic–uremic syndrome (aHUS) and retinal damage in age-related macular degeneration (AMD).

Genetically based deficiencies have taught us that individuals lacking in a complement activating component present with recurrent infections and/or autoimmunity. While those lacking proper regulation of the system have an excessive response to tissue injury.

Knowledge of how complement is activated and how it can be controlled points to opportunities for the development of therapeutic agents. One such example is anti-C5 monoclonal antibody (mAb) therapy, which has been recently approved to treat several complement-dependent pathogenic disorders. Other new complement therapeutics and diagnostics are on the horizon as genetic evaluations are increasingly utilized to define diseases in which the complement system is involved, and biotechnology companies are pursuing development of novel complement inhibitors.

COMPLEMENT PATHWAYS

The three pathways of complement activation are the CP, lectin pathway (LP), and AP (Fig. 40.1).[1,2] The CP is initiated by IgM or IgG Ab binding to an antigen. The CP can also be activated by innate pattern recognition molecules, such as the pentraxins and C-reactive protein (CRP). They participate with natural antibodies in early host defense and provide a mechanism for clearance of cells via immune complexes and apoptosis. The LP uses most of the CP components but is activated by mannan-binding lectin (MBL) and the ficolins, which are lectins that recognize repeating carbohydrate patterns on microorganisms.[3] The AP is the most ancient pathway and also has the broadest recognition ability. The AP is engaged by surface components of all types of microorganisms, including bacteria, fungi, parasites, and viruses.[4] It is continuously turning over and autoactivates if its inhibitors are lacking. Activation of the AP can also be initiated by properdin (P), a molecule that binds to pathogens and apoptotic cells. This mechanism further promotes its function as an innate and rapid responder to infection. The AP is an important amplification mechanism for CP or LP activation, resulting in greater opsonization and generation of the terminal lytic pathway. For example, the initial trigger on a pathogen surface may be IgM or a lectin, but the majority of the C3b deposited is via AP's amplification or feedback loop that is engaged by just having C3b on a target.

The cleavage of C3 to C3a and C3b is central to all three pathways of complement activation. This enzymatic step exposes a highly reactive thioester bond through which C3b covalently attaches to nearby molecules (Fig. 40.2). Activation of C3 to C3b also exposes sites for interactions with other complement proteins, inhibitors, and receptors. Recent results have shed new light on the structural basis for C3 activation. In 2006, several studies reported the first x-ray structure of C3b, the activated product C3.[5,6] The results revealed a major conformational change in C3 upon cleavage to C3b that exposes the reactive thioester group as well as cryptic binding sites for complement receptors and regulatory proteins. Moreover, the crystal structures of five binding proteins in complex with C3b have now been solved.[7]

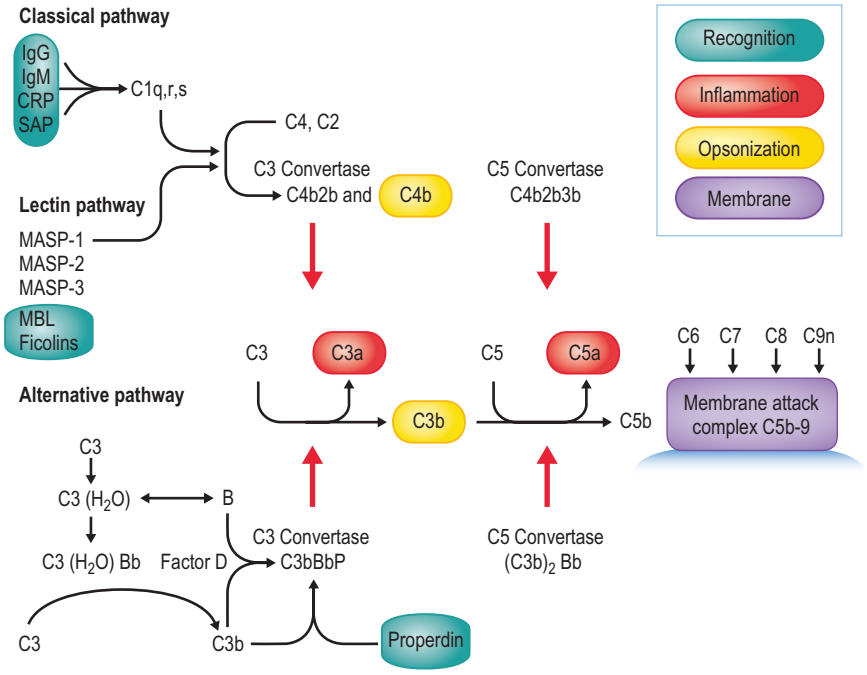


FIG. 40.1 Overview of the Complement Pathways Indicating Components Required for Recognition, the Enzymatically Active Fragments and Complexes and the Major Opsonic, Inflammatory, and Membranolytic Products. Note: not shown are the released fragments of C4 (C4a), C2 (C2b), and factor B (Ba). *MASP*, mannan-binding lectin–associated serum proteases.

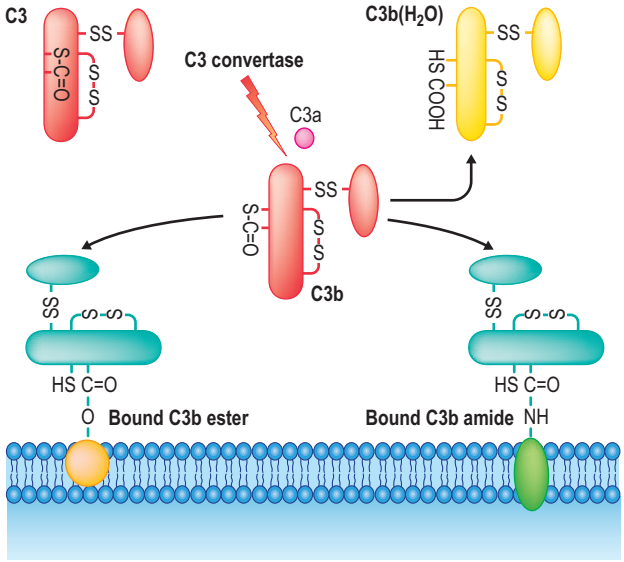


FIG. 40.2 Exposure and Reactivity of the C3 Thioester Bond. C3 cleavage by a C3 convertase generates metastable C3b with a reactive thioester. Metastable C3b may be hydrolyzed to form C3b(H_2O) or may react with hydroxyl (ester linkage) or amino (amide linkage) groups to become covalently bound to a surface. Note: C3 is composed of two disulfide-linked chains, the α-chain (110 kDa) and the β-chain (75 kDa).

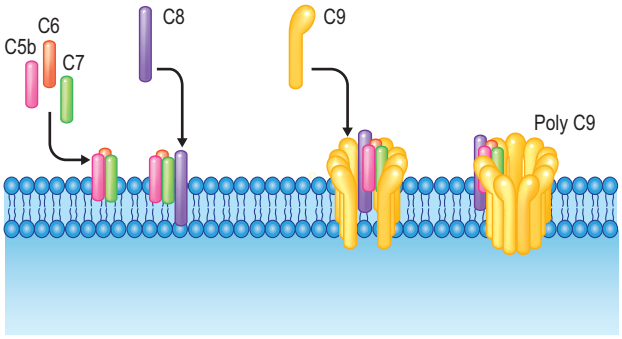


FIG. 40.3 Sequence of Protein Interactions in the Assembly of the Membrane Attack Complex (MAC). C5b, generated by a convertase cleaving C5, combines with C6 and C7 to form a hydrophobic complex capable of a membrane interaction. Binding of C8 allows the complex to insert further into the membrane and forms a site for C9 polymerization. C9 (~18 monomers) polymerizes to form a transmembrane pore to mediate cell lysis.

Classical Pathway

KEY CONCEPTS

Structural and Functional Homologies in Complement Pathways

Recognition: C1q, MBL, ficolins, CRP
Initiating enzymes: C1r, C1s, MASP-1, MASP-2, FD
C3 convertases: C4b2b, C3bBb
C5 convertases: C4b2b3b, (C3b)2Bb
Enzyme subunits of convertases: C2a, Bb
Assembly subunits: C3b, C4b (both covalently bound to target)
Anaphylatoxins: C3a, C5a
MAC subunits: C5b, C6, C7, C8, C9
Regulatory proteins: C4BP, FH, CR1, MCP, DAF, CSMD1
Receptor proteins: CR1 and CR2; CR3 and CR4, CRIg
Major opsonins: C3b and C4b

The CP is initiated primarily by Ab binding to a target antigen. In general, the ability of Ab to activate complement is ranked as follows: IgM > IgG3 > IgG1 > IgG2 > IgG4. Binding of these Abs exposes sites in the Fc region for attachment of the first subcomponent of complement, C1q.[8,9] C1 is a large calcium-dependent complex composed of C1q and two molecules each of the proenzymes, C1r and C1s. C1q is a 410-kDa protein with six globular heads connected by a collagen-like tail. IgM, IgG, and CRP bind C1q through its globular head groups. For IgM, which is pentameric, binding to antigen creates a conformational change that exposes the C1q binding site. For IgG, at least two closely bound molecules are required to provide multiple attachment points for C1q binding.

Once C1q binds to an activator, C1r is cleaved by an autocatalytic process. Activated C1r then cleaves and activates C1s, which, in turn, cleaves circulating C4. Note that there are two isotypes of C4: C4A and C4B. These differ in their preference for acceptor nucleophiles, amino groups for C4A and hydroxyl groups for C4B.

C4 and C3 are highly homologous proteins that share an unusual posttranslational modification known as an *internal thioester bond* (see Fig. 40.2).[5] Cleavage of C4 releases the C4a fragment and exposes the reactive thioester bond in the larger C4b fragment. This allows C4b to attach covalently to nearby target structures through hydroxyl or amino groups to form an ester or amide bond, respectively. The exposed thioester bond is highly but transiently reactive as it is susceptible to rapid hydrolysis. For example, within microseconds, about 5% of the generated C4b typically becomes attached to the target. Bound C4b provides an anchor site for C2 attachment, which is then also cleaved by C1s, releasing the smaller fragment C2a and leading to the formation of C4b2b.

The complex C4b2b is termed the *CP C3 convertase* because it cleaves C3 to C3b, releasing C3a. The C2b component of the complex contains the active enzymatic site. C3 cleavage is similar to C4 cleavage in that the larger fragment, C3b, contains a thioester site (see Fig. 40.2) that mediates covalent attachment to nearby surface structures. C3 is found at a three- to fourfold higher concentration in serum compared with C4, and its cleavage is amplified by the AP. Thus, efficient complement activation will result in clusters of multiple bound C3b molecules that can be recognized by cellular receptors. C3b that attaches to C4b within the C3 convertase produces the trimolecular complex C4b2b3b, which is a C5 convertase. Cleavage of C5 produces C5a, which has potent inflammatory activity, and C5b, which initiates the formation of the MAC or, as it is also known, the *terminal complement complex* (TCC) (Fig. 40.3).

Lectin Pathway

The LP is similar to the CP, except that it uses pattern recognition molecules, MBL, and ficolins-1, -2, and -3, instead of Ab to target activation.[3,10] MBL is structurally similar to C1q, with a collagen-like region and globular heads. The globular heads of MBL are C-type lectin domains specific for repeating carbohydrate structures found on microorganisms. Like C1q, MBL and ficolins are in complex with serine proteases, MBL-associated

serum proteases (MASPs), which are structurally and functionally similar to C1r and C1s. MASP-1 and -2 are active proteases, but only MASP-2 cleaves both C4 and C2 to generate C4b2a, the same C3 convertase as the CP. MASP-1 can supplement activation by cleaving C2 but not C4. Two nonproteolytic splice products of the *MASP2* and *MASP1/3* genes, sMAP and MAP-1, compete with MASP-1 and -2 for binding to MBL to regulate the LP. Subsequent steps in the LP are identical to those in the CP. Interestingly, MASP-3 bridges the LP and AP since it is the main enzyme that cleaves profactor D to mature factor D.[11,12]

Alternative Pathway

The AP uses proteins that are structurally and functionally homologous to those of the CP, but this pathway has unique features that play three important roles in the complement cascade. First, the surveillance role of the AP is mediated by a continuous low level of spontaneous activation that results from the hydrolysis of the C3 thioester bond.[4] Hydrolyzed C3, C3(H_2O), assumes a conformation similar to that of C3b and can bind factor B (homologous to C2), which is cleaved by factor D (homologous to C1s) to form a fluid-phase C3 convertase. This convertase cleaves C3 to generate C3b, which can covalently bind to nearby structures and provide the basis for a bound C3 convertase (C3bBb). Because C3b is both a part of this enzyme and a product of the reaction, a positive feedback loop that rapidly deposits more C3b is formed. This "idling-like" activation process is tightly regulated on host cells and tissues by plasma and membrane-bound complement regulatory proteins. The plasma protein factor H (FH) is particularly important in controlling AP activation, both in the fluid phase and on "nonactivating" surfaces. The latter recruits FH through its binding sites for polyanions, including sialic acid and glycosaminoglycans. "Activating" surfaces, such as microbial polysaccharides, lipopolysaccharides, and foreign glycoproteins, provide C3b attachment sites that are protected from regulatory proteins. Similar to the CP, the AP C5 convertase (C3bBb3b) is formed when a second C3b attaches to the C3 convertase. The AP C3 and C5 convertases are stabilized by P (factor P or P), for which this pathway was originally named.

Second, an additional role for properdin (P) in initiating AP activation was rediscovered.[4,13,14] P is a pattern-recognition molecule with specificity for microbes and damaged cells. Once bound, P can recruit C3b and thereby provide a platform for the assembly of the AP convertase. Thus, P binding can direct AP activation, similar to MBL in the LP. P binding to certain *Neisseria* species potently activates the AP, and this may account for the susceptibility of P-deficient individuals to infection with *N. meningitidis*.

The third important role of the AP is the amplification of C3b deposition and C5 convertase generation that is initiated by the CP or the LP.[4] This function of the AP is critical in complement-mediated pathology, as it increases the generation of C5a and the MAC, the most inflammatory components of the system. It is this amplification role of the AP that makes it an attractive therapeutic target.

MEMBRANE ATTACK COMPLEX

All three complement pathways merge with the cleavage of C5 into C5a and C5b. Although C5 is structurally homologous to C3 and C4, it lacks an internal thioester bond that allows covalent attachment to surfaces. C3a and C5a are also structurally homologous and, as described below, are the most potent proinflammatory mediators of the complement system. C5b initiates the formation of the MAC (see Fig. 40.3). The membrane-bound MAC precursor complex that is composed of C5b, C6, C7, and C8 recruits C9 from plasma. Unfolding of the initial C9 exposes binding sites for the next C9 molecule sequentially until a ring is formed with 18 copies of C9.[15] This complex, as indicated by its name, penetrates membrane bilayers to form pores that disrupt the osmotic barrier, leading to swelling and lysis of susceptible cells. Lysis of Ab-sensitized erythrocytes by the MAC is the basis of the total hemolytic complement (THC) assay or CH_{50}. C5b initiates the formation of the MAC. Without further proteolytic steps, C5b binds to C6 and this complex binds to C7. The C5b67 complex is lipophilic and associates with cell membranes, if available, or with serum lipoproteins. Once bound to a membrane, C5b67 recruits C8, and the complex penetrates more deeply into the membrane. However, efficient lysis requires C9, a pore-forming molecule with homology to perforin, a protein used by cytotoxic T cells and natural killer (NK) cells for killing virus-infected targets. The complex of C5b678 forms a nidus for C9 binding and polymerization. Although complement-dependent lysis of bacteria can be observed in vitro, many pathogens have evolved mechanisms to circumvent this activity of complement.[16] Opsonization by C3b is therefore the most potent mechanism for destruction (adherence followed by ingestion) of bacteria by the complement system. The sublytic MAC is proinflammatory because of its membrane-perturbing capabilities for host cells, and it contributes to the deleterious effects of complement activation in inflammatory diseases.

REGULATION OF COMPLEMENT ACTIVATION

The complement cascade is rapidly activated and highly amplified by the generation of C3 and C5 convertases. Control mechanisms exist at three main levels that limit the potential harm that uncontrolled complement activation might cause: (1) the initiation step in the CP and the LP; (2) the C3 and C5 convertases of all three pathways; and (3) the assembly of the MAC. Both soluble and membrane-bound regulatory proteins serve these functions, which direct it to the appropriate targets and then terminate complement activation.[17]

C1 Esterase Inhibitor

C1 esterase inhibitor (C1-INH) is a plasma serine proteinase inhibitor (serpin) that covalently binds to activated C1r and C1s, irreversibly inhibiting their activity and thereby limiting CP activation. C1-INH inactivation of C1r and C1s also removes them from the C1 complex, exposing sites on the collagen-like region of C1q. Likewise, C1-INH inhibits MASP-1 and MASP-2, kallikrein, factor XIa, factor XIIa, and plasmin of the LP and the contact, coagulation, and fibrinolytic systems. Inherited deficiency of C1-INH is the basis of hereditary angioedema, a disease characterized by recurrent attacks of subcutaneous or submucosal edema (Chapter 46).[18]

Regulators of the C3 and C5 Convertases

The C3 and C5 convertases are central to the generation of the inflammatory and opsonic products of complement activation and are highly regulated by fluid-phase and membrane-bound regulatory proteins. First, the membrane-deposited C4b and C3b may be bound by the regulator to prevent convertase formation with C2 or FB. Second, the convertases themselves are complexes

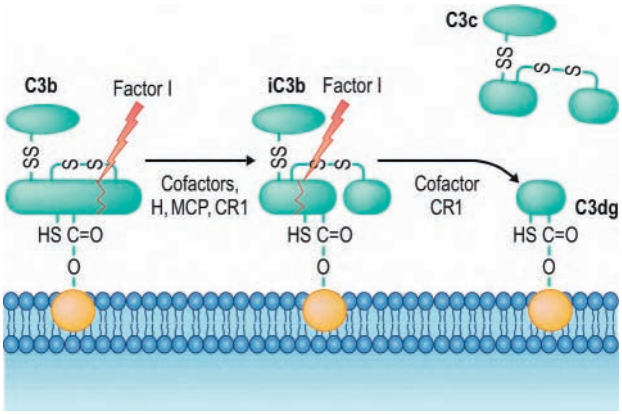


FIG. 40.4 Factor I *(FI)*–Dependent Cleavage of C3 Showing the Structures of the Products and the Required Cofactors. The cofactor protein binds first and then the serine protease factor I cleaves the C3b at two sites in two locations on its α-chain. *MCP,* Membrane cofactor protein.

of two or three components, and one mechanism of regulation is the dissociation of these complexes. This type of regulation is termed *decay acceleration*. A third mechanism of regulation is the enzymatic inactivation of the C4b and C3b components of the convertases (Fig. 40.4). This is accomplished by the plasma serine protease factor I (FI), which, however, only acts on C4b or C3b in association with one of several regulatory proteins. The binding of regulatory proteins to C4b or C3b to enable FI cleavage is termed *cofactor activity*.

Factor I

Factor I (C3b inactivator, C3bINA) cleaves C4b and C3b into products that are recognized by specific cellular receptors (as discussed below). The sequential cleavages of C3b by FI to iC3b and C3dg are depicted in Fig. 40.4. C4b is cleaved in an analogous manner to C4d. (The iC4b intermediate is found only transiently.) The regulatory proteins that facilitate this cleavage by cofactor activity and those that inactivate C3 and C5 convertases by decay-accelerating activity are members of a family of structurally related proteins encoded within the regulators of complement activation (RCA) genetic locus.[19–21] This family is characterized by a repeating structure that consists of subunits, termed *complement control protein repeats* (CCP) or sushi domains, of about 60 amino acids with a conserved pattern of two disulfide bonds per repeat, and are usually encoded by a single exon.

Soluble Regulatory Proteins, C4b-Binding Protein, and FH

C4b-binding protein (C4BP) and FH are fluid-phase regulatory proteins with both decay-accelerating and cofactor activities. C4BP is multimeric, being composed of seven identical subunits, each containing eight CCPs. FH is a single-chain protein composed entirely of 20 CCPs. C4BP is specific for C4b and the C4b-containing convertases of the CP (C4b2b, C4b2b3b), whereas FH regulates C3b and C3b-containing convertases (C3bBb, C3bBb3b, C4b2b3b). FH is essential for regulation of C3 "tickover," and FH deficiency results in an acquired deficiency of C3. Additional binding sites on FH that recognize polyanions, such as sialic acid and glycosaminoglycans, provide targeted regulation of AP activation on surfaces.[17,22,23]

Membrane Regulatory Proteins

The RCA family includes the membrane regulatory proteins decay-accelerating factor (CD55, DAF), membrane cofactor protein (CD46, MCP), and complement receptors CR1 (CD35) and CR2 (CD21).[19,21,24,25] Additionally, other recently identified membrane regulators are the CUB (for C1r/C1s Bmp1, Uegf, Bmpf) and sushi multiple domains protein 1 (CSMD1) and complement receptor of the immunoglobulin superfamily (CRIg).

CD55 (DAF) and CD46 (MCP), as their names imply, have decay-accelerating and/or cofactor activity, respectively, that inhibits complement activation on cell membranes.[19] Each has an extracellular domain composed exclusively of four CCPs. CD55, a glycosylphosphatidylinositol (GPI)–anchored protein, and CD46, a transmembrane protein, are widely distributed on cells in contact with blood, with the notable exception of erythrocytes that lack CD46. Soluble CD55 is also found in most biological fluids. Both protect cells from complement-mediated lysis. CD35 (CR1) has decay-accelerating and cofactor activity and is a receptor for bound C4b and C3b. The function of CD35 as a complement receptor is discussed later in the chapter.

Two more recently described regulators are CSDM1 and CRIg. CSDM1 is a type 1 transmembrane protein that is widely expressed and contains 28 CCPs and 14 CUB domains. It inhibits the classical pathway and serves as a cofactor for cleavage of C3b.[26,27] CRIg binds C3b and iC3b and functions to inhibit the alternative pathway, clear systemic pathogens, and regulate the adaptive immune response. CRIg is widely expressed on macrophages in many tissues, being highly prevalent in the Kupffer cells in the liver.[6]

Properdin

In contrast to the regulatory proteins discussed above, the plasma protein P (factor P) stabilizes C3 and C5 convertases of the AP, increasing their activity.[4,13,14] This enhancer of AP activation is found as noncovalently linked dimers, trimers, tetramers, and larger species composed of identical 56-kDa chains. The majority of this plasma protein consists of a series of six thrombospondin type 1 modules. P binds to C3b and to Bb, preventing the spontaneous or induced decay of the AP C3 and C5 convertases. Its multimeric structure promotes interaction with clustered C3b. As discussed above, bound P can also recruit C3b to provide a site of assembly for the AP C3 convertase.

Regulators of the Membrane Attack Complex

The MAC is also regulated by both fluid-phase and membrane regulatory proteins.[1,17,25,28]

Soluble Membrane Attack Complex Inhibitors: Vitronectin and Clusterin

Soluble hydrophobic proteins block the incorporation of the MAC into membranes. Two well-characterized proteins with this activity are vitronectin (S protein) and clusterin (SP-40,40, apolipoprotein J).[17,25] Vitronectin is in plasma and the extracellular matrix and binds to C5b-7. C8 and C9 can still bind to the complex, but membrane insertion and C9 polymerization are prevented. Soluble complexes of vitronectin and C5b-9 are in plasma during complement activation, and an enzyme-linked immunosorbent assay (ELISA) specific for this complex has been used to monitor activation of the MAC. Clusterin forms

a complex with C5b-9, preventing membrane insertion. It is found in plasma, in the male reproductive tract, and on endothelial cells of normal arteries.

Membrane Attack Complex Inhibitor CD59

The primary membrane-bound inhibitor of the MAC is CD59.[17,25] CD59 is a GPI-anchored protein expressed by most cells. CD59 binds to C8 and C9, preventing their incorporation and thus blocking polymerization of C9.

COMPLEMENT RECEPTORS

Many of the biological effects of complement activation are mediated by cellular receptors for fragments of complement proteins. These include receptors for the small soluble complement fragments C5a and C3a and receptors for bound complement fragments C1q, C4b, and C3b as well as their cleaved fragments. Receptors are specific for C3b and for its further breakdown products generated by the enzymatic processing by FI in conjunction with the cofactor proteins mentioned above. The breakdown of C3b and its intermediate products are shown in Fig. 40.4 and the receptors for these components in Fig. 40.5.

C1q Receptors

C1q is one of a family of proteins termed *soluble defense collagens* that includes the "collectins" (MBL, surfactant proteins A and D, conglutinins, and the ficolins). Each of these proteins is composed of a collagen-like linear stem region terminated by multiple globular recognition domains or head groups. The collectins recognize carbohydrates with their C-type lectin head groups, and the ficolins recognize acetyl groups on carbohydrates and other molecules with fibrinogen-like recognition domains. In contrast, the globular head groups of C1q do not recognize carbohydrates but, rather, bind to amino acid motifs on IgG, IgM, and pentraxins. In general, the soluble defense collagens broadly recognize pathogen-associated carbohydrate patterns and damaged or apoptotic cells. Reported direct effects of this group on leukocytes include the enhancement of phagocytosis, triggering of the respiratory burst, and regulation of cytokine responses. Several cell-surface proteins have been proposed to facilitate these activities, including CD93 (C1qRp), CD35 (CR1), $\alpha_2\beta_1$ integrin, calreticulin in complex with CD91, gC1q-binding protein, and SIPRα. However, none has been definitively established as a receptor in the classic sense.[1,10,29,30]

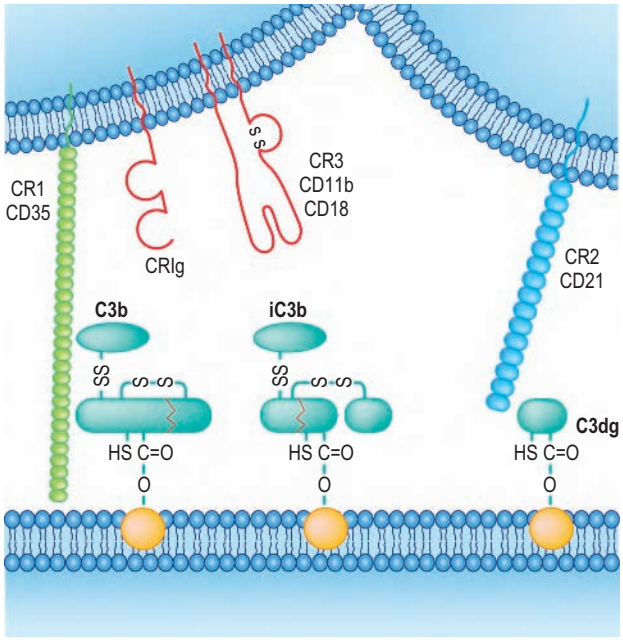


FIG. 40.5 Receptors for Bound C3b and Its Cleavage Products. Receptors shown are CD35 and CD21 composed of CCP (SCR) subunits; CD11b/CD18 (CR3), a β_2 integrin; and CRIg with one or two immunoglobulin domains. The specificities of the receptors are CD35 for C4b and C3b (C4b > C3b), CRIg for iC3b > C3b; CD11b/CD18 for iC3b; CD21 for C3dg and C3d. CD11c/CD18 (CR4) is similar to CD11b/CD18 and is not shown. Receptors are not drawn to scale. Their molecular weights are listed in Table 40.1.

Complement Receptor 1 (CR1, CD35)

There are five identified receptors for bound fragments of C3 and/or C4. CD35 is a large protein composed of a linear string of CCPs, a transmembrane region, and a short intracytoplasmic domain. Different allelic forms of CD35 are found, the most common being composed of 30 CCPs with a molecular weight of 190 kDa. These CCPs are organized into groups of seven, creating structures termed *long homologous repeats* (LHRs), each of which contains a single binding site. The predominant allele of CD35 contains two binding sites for C3b, three for C4b, and one for C1q. CR1 is expressed on human erythrocytes, monocytes and macrophages, neutrophils, B lymphocytes, a small percentage of T lymphocytes, eosinophils, FDCs, and glomerular podocytes.

CD35 on primate erythrocytes provides a mechanism for clearing immune complexes from the circulation. Although the number of receptors on each erythrocyte is low, the large number of erythrocytes provides the major pool of CR1 in the circulation. Soluble immune complexes that fix complement attach quickly to erythrocytes in the circulation, bypassing monocytes and neutrophils. These erythrocyte-bound complexes are taken to the liver and spleen, where they are transferred to tissue-specific macrophages such as hepatic Kupffer cells expressing Fc and complement receptors. The erythrocytes exit into the circulation to pick up more immune complexes. This clearance pathway is impaired in patients with SLE because of decreased complement in the circulation, decreased CD35 on erythrocytes, and saturated Fc receptors in the liver and spleen.

CD35 on monocytes and neutrophils promotes binding of microbes carrying C3b and C4b on their surface (immune adherence reaction), facilitating their phagocytosis through Fc receptors. CD35 can directly mediate the uptake of microbes when phagocytic cells have been activated by chemokines or integrin interactions with matrix proteins. CD35 is a member of the RCA family and has decay-accelerating and cofactor activity in addition to its function as a receptor. It differs from the membrane regulatory proteins DAF (CD55) and MCP (CD46) in its ability to also bind to C3b and C4b extrinsically (on targets other than the cell expressing it) and in its cofactor activity for iC3b processing. CD35 is the most effective cofactor for FI cleavage of C3b and iC3b to the smallest covalently bound fragment C3dg. C3dg is the major ligand for CR2 on B lymphocytes (described below). The cofactor activity of CD35 on B lymphocytes processes bound C3b to C3dg, facilitating binding to CR2 and lowering the threshold for B-cell activation.[21,25,31]

Complement Receptor 2 (CR2, CD21)

CD21 is also an RCA family protein composed of 15 to 16 CCPs. CD21 has a limited range of expression that includes B lymphocytes, FDCs, and some epithelial cells. CD21 is specific for the smallest covalently bound C3 fragments, C3dg and C3d, and has weaker binding to iC3b. CD21 is also the Epstein-Barr virus (EBV) receptor on B cells and nasopharyngeal epithelial cells and binds to CD23, a low-affinity IgE receptor.[21,31]

CD21 on B lymphocytes serves a costimulatory role. It is expressed on mature B cells as a complex with CD19 and CD81 (TAPA-1). Coligation of CD21 and the B-cell antigen receptor induces the phosphorylation of CD19, activating several signaling pathways and strongly amplifying B-cell responses to antigen. This role of CD21 is believed to contribute to the strong adjuvant effect produced by attaching C3d to antigen.[21,31]

Complement Receptors 3 and 4

CR3 and CR4 are the β_2 integrins commonly known as CD11b/CD18 (Mac-1) and CD11c/CD18.[21] β_2 integrins are large heterodimers found on neutrophils and monocytes with multiple roles in adhesion to endothelium and matrix molecules as well as direct recognition of microbial pathogens. The binding activities of β_2 integrins are regulated by cellular activation often through chemokine receptors. Both CD11b/CD18 and CD11c/CD18 are expressed primarily on neutrophils, monocytes, and NK cells and bind to iC3b and, to a lesser extent, C3b. CD11b/CD18 has been studied more extensively than CD11c/CD18. CD11b/CD18 expression, clustering, and conformation are all rapidly upregulated by chemokine activation of neutrophils, leading to increased responses to ligand. CR3 plays an essential role in neutrophil attachment to and migration through activated endothelium to sites in inflammation and in the regulation of neutrophil apoptosis. Deficiency of the β_2 chain (CD18) results in leukocyte adhesion deficiency, characterized by recurrent pyogenic infections, and defects in inflammatory and phagocytic responses. Complement receptors CD11b/CD18 and CD11c/CD18 provide an essential function for the removal of microbial pathogens following complement activation, since C3b processing to iC3b often occurs rapidly after deposition.

Complement Receptor of the Immunoglobulin Superfamily (CRIg)

CRIg is a receptor for iC3b and C3b present on Kupffer cells in the liver as well as other tissue macrophages but is absent from splenic macrophages, peripheral blood cells, bone marrow–derived macrophages, and monocyte/macrophage cell lines.[32] Two alternatively spliced forms of human CRIg were identified with one and two Ig domains. The mouse receptor has a single Ig domain. CRIg removes C3b or iC3b-opsonized particles from the circulation by the liver.

C5a and C3a Receptors

During complement activation, the homologous proteins C3 and C5 are each cleaved near the amino-terminus of the α chains to release a soluble peptide fragment of approximately 8 kDa. These fragments are designated C3a and C5a. C5a may also be generated locally by direct cleavage of C5 by thrombin or leukocyte proteases.[33] C3a and C5a are termed *anaphylatoxins* because of their ability to increase vascular permeability, contract smooth muscle, and trigger the release of vasoactive amines from mast cells and basophils.[34,35] C5a is 10- to 100-fold more active than C3a. These peptides are also chemotactic: C5a is specific for neutrophils, monocytes, and macrophages, whereas C3a is specific for mast cells and eosinophils. Biological activities of complement anaphylatoxins are summarized in Table 40.2.

TABLE 40.2 Cellular Targets and Effects of Complement Anaphylatoxins

	Targets Bearing Receptors	Effects
C3a, C5a	Mast cells, basophils	Degranulation, release of vasoactive amines: contraction of smooth muscle, increased vascular permeability
C3a	Eosinophils	Chemotaxis, degranulation
C5a	Endothelium	Increased adhesion of leukocytes; augmented chemokinesis and cytokine synthesis
C5a	Neutrophils, monocytes/ macrophages, eosinophils, basophils, astrocytes	Chemotaxis
C5a	Neutrophils, monocytes/macrophages	Priming: activation of receptors, assembly of nicotinamide adenine dinucleotide phosphate (NADPH) oxidase; activation: degranulation, respiratory burst
C5a	Resident macrophages	Regulation of FcγR expression (↑activating, ↓inhibitory)
C5a	Hepatocytes	Acute phase protein synthesis
C3a, C5a	Lymphocytes (antigen-presenting cells)	Regulation of T-cell responses to antigen

Structurally, anaphylatoxins are compact structures consisting of multiple helices cross-linked by disulfide bonds with more flexible carboxy-terminal regions. The C-terminal peptide of C3a interacts with the C3aR and can reproduce C3a agonist activity. In contrast, C5a interacts with the C5aR at multiple sites. Plasma carboxypeptidases cleave the C-terminal arginine from C3a and C5a, producing the des-Arg forms. This inactivates C3a; however, C5a-desArg retains much of its biological activity. The C5aR (CD88 and C5L2) and the C3aR are rhodopsin-type receptors with seven transmembrane-spanning domains coupled to G-protein signaling pathways. C5aR is expressed at high levels on neutrophils and is also expressed by macrophages, mast cells, basophils, smooth muscle cells, and endothelial cells. If C5a is generated locally, for example, at an extravascular site of infection, it helps induce an acute local inflammatory response, including vasodilation, edema, neutrophil chemotaxis, and activation of neutrophils and macrophages for enhanced phagocytosis and killing. The inflammatory activities of C5a can also contribute to complement-mediated pathology in some conditions, such as sepsis, acute respiratory distress syndrome, and ischemia/reperfusion (I/R) injury, making the C5a–C5aR interaction an attractive therapeutic target.

The C5L2 receptor binds to both C5a and C5a-desArg. C5L2 was initially believed to be a default or decoy receptor for C5a because it is uncoupled from G proteins. Genetic deletion of C5L2 (*Gpr77*$^{-/-}$) in mice resulted in enhanced neutrophil infiltration and cytokine production in the pulmonary Arthus

reaction, supporting an antiinflammatory role for C5L2 in immune complex disease, where genetic deletion of C5aR is fully protective. However, studies in a cecal ligation and puncture (CLP) model of sepsis found increased survival in mice lacking either C5aR or C5L2. The results suggest an active proinflammatory role for C5L2 that requires C5a and results in the release of the inflammatory signal, high-mobility group box-1 protein (HMGB1) from phagocytic cells. Thus, both C5aR and C5L2 may contribute synergistically to harmful inflammatory events during sepsis.

COMPLEMENT IN HOST DEFENSE AND IMMUNITY

Complement in Host Defense

Complement activation provides a coordinated response to infection that results in the opsonization of microbial pathogens and the attraction and activation of phagocytic cells to kill them. Complement-dependent opsonization is of greatest importance in infections with encapsulated extracellular bacteria, and individuals with deficiencies in Ab production, neutrophil function, or C3 share increased susceptibility to these organisms, including *Streptococcus pneumoniae* and *Haemophilus influenzae*. MBL deficiency is also associated with recurrent pyogenic infections in young children. In general, activation of complement by natural Ab or MBL results in C3b and iC3b deposition on these pathogens, overcoming the antiphagocytic effects of the capsule. Phagocytic cells ingest and kill the organisms using CD35, CD11b/CD18, and CD11c/CD18 receptors in conjunction with other innate and Fc receptors. C5aR signaling activates these receptors, leading to increased phagocytosis. Gram-negative bacteria are also susceptible to complement-dependent lysis. This is evident in the increased incidence of disseminated neisserial infection in individuals deficient in C3, any of the MAC components, or P, as discussed below.

Complement in Inflammation

An essential function of complement in host defense is the coordination of the local inflammatory response. C5a is the most potent complement product in this activity.[35] Sublytic deposition of the MAC on endothelial cells and platelets and C3a interaction with the C3aR also contribute to the proinflammatory effects of complement activation. As discussed below, these potent inflammatory fragments of complement, when generated in high amounts or targeted inappropriately, result in many of the disease-related deleterious effects of complement. Local production of C5a at a site of infection occurs either through local complement activation or through direct cleavage of C5 by tissue macrophages or thrombin.[33,34] This C5a is released and sets up a chemotactic gradient for neutrophils and macrophages. In addition, C5a activates endothelial cells to express P-selectin and synthesize chemokines, including interleukin-8 (IL-8). Interaction of C5a with mast cells releases vasoactive amines, increasing endothelial permeability. Neutrophils and macrophages are "primed" by interaction of C5a with its receptor. Priming includes enhancement of chemotaxis, activation of complement receptors for phagocytosis increased expression of activating FcγR, and assembly of the nicotinamide adenine dinucleotide phosphate (NADPH) oxidase that is required for effective killing of microbes after phagocytosis. C5a also prevents neutrophil apoptosis, prolonging survival and contributing to local accumulation. Together, these actions result in the attraction and activation of potent antimicrobial cells and resolution of infection.

Pathogen Evasion of Complement

Further evidence of the host defense function of complement is the association of complement evasion strategies with virulence. Pathogenic gram-negative bacteria, such as *Salmonella*, have lipopolysaccharides with long O-polysaccharide side chains that promote rapid shedding of the MAC and prevent its insertion into the cell membrane. *Neisseria* species have several FH-binding components that help restrict AP activation and protect against lysis. Group A and B streptococci and *S. pneumoniae* have cell-surface components (M protein, Bac or beta, PspC, Hic) that bind to FH and/or C4BP, restricting complement activation. Other organisms, including type 3 group B streptococci, elaborate sialic acid–containing capsules or cell walls to limit AP activation.

Although complement deficiencies are not generally associated with viral infections, the importance of complement in host defense against viruses is suggested by the multiple strategies used by viruses to evade complement.[16,36] Several viruses

KEY CONCEPTS

How Pathogens Abuse and Evade the Complement System: Some Examples

1. Bacteria
Block C1, C3b deposition
Streptococcus pneumoniae
Block MAC access to plasma membrane
Salmonella
Limit access of C3b, iC3b to complement receptors by capsule
Streptococcus pneumoniae
Haemophilus influenzae
Block AP activation by sialylation
Streptococcus agalactiae (GBS) type III
Neisseria
Bind (highjack) FH, C4BP to inhibit complement activation
Streptococcus pneumoniae (Hic)
Streptococcus pyogenes (GAS) (M protein)
Neisseria
Use CD55 (DAF), CD46 (MCP) for attachment to cells
Streptococcus pyogenes (GAS) (M protein)
Neisseria
Escherichia coli
Use complement receptors for entry
Mycobacterium tuberculosis (CR3)
Bacillus anthracis spores (CR3)

2. Viruses
Express complement regulatory proteins homologous to those synthesized by the host
Herpes simplex virus (glycoprotein C)
Poxviruses (SPICE/VICE)
Express unique complement regulatory proteins
Flaviviruses (Dengue, West Nile)
Use CD55 (DAF), CD46 (MCP) for attachment to cells
Measles virus, adenovirus, herpes virus 6 (CD46)
Picornaviruses, hantavirus (CD55)
Use complement receptors for entry
Epstein-Barr virus (CD21)
Human immunodeficiency virus (CD35, CR3)

3. Parasites
Express complement regulatory proteins
Schistosoma (CRIT)
Acquire complement regulatory proteins from host
Schistosoma (CD55)
Use complement receptors for entry
Leishmania (CR1, CR3)

produce complement regulatory proteins, including vaccinia virus complement control protein and herpes virus glycoprotein C, which facilitate breakdown of C3b and C4b. Some viruses, such as human immunodeficiency virus (HIV), incorporate complement regulatory proteins into the viral envelope, a strategy that is also used by other pathogens, such as *Schistosoma*.[16,36,37]

There are also many examples of complement receptors and membrane regulatory proteins being exploited as receptors for pathogens to invade cells. Examples of these include strategies of direct pathogen binding to receptors as well as deposition of C3 fragments followed by invasion through host C3 receptors.[16]

Role of Complement in Adaptive Immunity

Over the past 10 years there has been renewed interest in the role of the innate immune system in adaptive immune responses.[4] The importance of complement in humoral immunity has been recognized since the observation that complement depletion of mice before immunization decreased Ab responses to thymus-dependent antigens. Further studies have shown that complement receptors CR1 (CD35) and CR2 (CD21) are also essential for immunomodulation.[21,31] In humans, these receptors are found together on B cells and FDCs. CD35 is also expressed on a number of other cell types (described above), including erythrocytes and phagocytic cells.

Effects of Complement on the Humoral Immune Response

Results obtained by experimental manipulation of C3, C4, and their receptors in mouse models indicate roles for these complement components at multiple levels in the humoral immune response.[21,31] One caveat regarding these studies is that in the mouse CD35 and CD21 are alternative splice products of the same gene, and genetically deficient animals lack both receptors.[21] In humans, CD35 and CD21 are encoded by separate genes. The first role of CD35/CD21 is in B-cell development, indicated by a pronounced defect in B-1-cell development in CD35/CD21–deficient mice. B-1 cells are generally found outside lymphoid follicles, have a restricted repertoire, and are essential in the production of natural Ab to pathogens, such as *S. pneumoniae*, and to self-antigens exposed on damaged cells, such as phosphatidylcholine and DNA. Although the mechanism of this defect in CD35/CD21–deficient mice is not fully understood, these mice have an altered repertoire of natural Ab and B-1 cells.[21,31] Decreased natural Ab may contribute to susceptibility to infection and autoimmune disease in hereditary complement deficiency (discussed below).

A second role for complement in the Ab response is the well-described function of CD21 as a coreceptor for the mature B-cell response to antigen.[21,31,38] As described above, CD21 is associated with the signaling complex of CD19 and CD81 (TAPA-1) in the B-cell membrane. Coligation of CD21 with the B-cell antigen receptor occurs naturally when the antigen activates complement and covalently binds C3dg. This coligation of the B-cell receptor with CD21 greatly decreases the threshold for B-cell activation and blocks Fas-initiated apoptosis of B cells. B cells activated by complement-opsonized antigen have increased ability to present antigen as well as survival and proliferation during encounters with T-dependent antigens.

The expression of CD35 and CD21 on FDCs is also important in the Ab response. FDCs trap antigen in the germinal centers and provide selection of somatically mutated high-affinity B-cell clones. Antigen trapped on FDCs also provides a source of long-term stimulation for maintenance of memory B cells. FDCs use complement receptors (CD35 and CD21) and FcγR to trap and retain antigen for these functions. Expression of CD21 on both FDCs and B cells is required for effective affinity maturation of the Ab response and for the development and maintenance of memory B cells.

Complement and T-Cell Biology

For many years, T cells were thought to express only a limited repertoire of complement proteins. However, we now know that T cells express many complement components, although at lower amounts relative to APCs.[39] Complement proteins can exert both a direct and indirect effect on T cells. For example, CR1 is expressed by ~12% of blood-circulating $CD4^+$ and $CD8^+$ T cells and is upregulated upon TCR engagement. However, CR1 is considered to be a negative controller of T-cell activity since CR1 stimulation during $CD4^+$ T-cell in vitro activation inhibits proliferation and IL-2 production and induces IL-10 secretion.[39] Further, CR1 may have additional signaling roles in T cells.

The first evidence that complement indirectly influences T-cell protective responses was shown by studies in primary pulmonary infection with influenza in which C3-deficient mice have a defect in influenza-specific $CD4^+$ and $CD8^+$ T-cell priming.[38] The mechanism may be more efficient uptake and presentation of C3-opsonized virus by APC through CR3 and CR4 or stimulation of T-cell responses through the C3aR.

Other complement components may directly impact T-cell function. For example, C1q-opsonized immune complexes (ICs) bind to T cells via C1qR and can induce cell activation. Additionally, the anaphylatoxin receptors can directly regulate T-cell responses as indicated by studies of *C3ar* and *C5ar1* double-knockout mice that demonstrated reduced ability to generate Th1 responses.

Further, costimulation of human T cells in vitro through CD3 and CD46 leads to the development of T cells with a regulatory phenotype characterized by synthesis of IL-10 in the absence of other Th2 cytokines (IL-2, IL-4) (Chapter 14). The induction of regulatory T cells (Tregs) was seen in response to both anti-CD46 cross-linking and natural ligands (C3b dimers, streptococcal M protein).

CD55-deficient mice showed enhanced T-cell responses to immunization and increased T cell–dependent autoimmune disease. These effects were complement dependent and apparently involved the loss of CD55 regulation of local complement synthesis by APCs during cognate interactions with T cells. One postulated mechanism is that CD55 inhibits the generation of C3a and C5a by APCs, preventing their interactions with C3aR and C5aR on T cells.[40]

The Intracellular Complement System (Complosome)

For many years, the complement system was considered a serum-centric system. However, a growing number of new studies are revealing that complement also has an intracellular arsenal of components that provide not only immune defense but also key interactions for host cell functioning (reviewed by Liszewski et al.[41] and Arbore et al.[42]). While early work has centered primarily on T cells, the intracellular complement system (the complosome) likely functions in most, if not all, cells. Some of these functions may trace their origin to the ancient complement system that likely began with a primitive form of C3 responsible for protection from intracellular pathogens. With continuing

evolution and creation of more components, this system expanded to the extracellular space as C3 became secreted to guard the vasculature from infectious agents. It is no surprise, then, that contemporary cells retain elements of this vestigial system.

Thus, current evolving understandings of this system recognize the following functions: (1) C3 serves as a damage-associated molecular pattern that, in particular, coats intracellular pathogens; (2) most cells contain intracellular stores of C3 and recycle C3(H2O); (3) intracellular C3 assists in cellular survival and metabolic reprogramming; (4) other components of the complosome include C5, factor B, properdin, and complement receptors and regulators (e.g., CD46, FH, C3aR, C5aR). As the complosome becomes better elucidated, new targets for the next generation of complement therapeutics may emerge.

Role of Complement in Clearance of Apoptotic Cells

Damaged tissue and dead and dying cells activate complement through several pathways. This can increase local inflammation and cellular damage, as in I/R injury, AMD, and hemolytic–uremic syndrome (HUS) (discussed below). Complement activation by apoptotic cells contributes to their opsonization and clearance and may prevent the development of autoimmunity. The deleterious consequences of complement activation following tissue damage are mainly attributable to AP-dependent generation of C5a and the MAC, whereas the beneficial effects are dependent on early CP components and innate recognition molecules.[43,44]

Necrosis, as occurs following ischemic tissue injury, exposes phospholipids and mitochondrial proteins that activate complement directly or indirectly. The pathways are different depending on the tissue involved.[43] For example, renal reperfusion injury appears to be initiated by the AP, possibly secondary to the loss of regulatory proteins on tubular epithelial cells. Intestinal (I/R) injury is initiated by natural IgM Ab and requires both the CP for initiation and the AP for injury. MBL and CRP-initiated complement activation have been proposed to contribute to myocardial reperfusion injury after coronary artery ligation.

Apoptotic cells are recognized by multiple receptors and opsonins.[44,45] The association between early CP deficiencies and SLE (see below and Chapter 52) has been attributed to a failure of complement-dependent opsonization, resulting in accumulation of apoptotic cells and released autoantigens. Support for this hypothesis is provided by studies of mice deficient in C1q, IgM, or serum amyloid P (SAP), all of which develop autoantibodies against phospholipid and nuclear antigens characteristic of SLE, and by the therapeutic effect of CRP in mouse models of SLE.[29] The role of complement in apoptotic cell recognition and uptake by macrophages is depicted in Fig. 40.6. MBL, C1q, and surfactant protein-D (SP-D) bind to apoptotic cells and facilitate clearance through direct binding to cellular receptors as well as complement activation.[45] Natural IgM Ab, CRP, and SAP bind to phospholipids exposed on late apoptotic cells. All three proteins can also activate the CP generating C1q, C4b, C3b, and iC3b ligands for complement receptors. CRP and SAP also directly opsonize apoptotic cells for uptake through Fcγ receptors.[46] Phagocytosis of apoptotic cells generally induces antiinflammatory cytokines transforming growth factor-β (TGF-β) and IL-10.

Targeted and Restricted Activation of the Complement System

Interestingly, CRP and SAP also bind complement regulatory proteins FH and C4BP, which helps limit complement activation to the deposition of opsonic components with little or no lysis or generation of C5a.[47] This type of targeted complement activation was also observed on acrosome-activated spermatozoa. In this case, the CP was activated by CRP from follicular fluid, resulting in bound C3b and iC3b, which are proposed to bind complement receptors on the egg and facilitate fertilization. Such engagement of the complement systems is likely a physiological (normal) mechanism for handling intracellular and extracellular debris that is continually produced in the body. For example, complement fragments are almost universally observed at sites of tissue injury in conditions such as atherosclerosis and Alzheimer disease. Whether such deposition is helpful or harmful (or both, depending on timing, duration, and magnitude) remains to be defined.

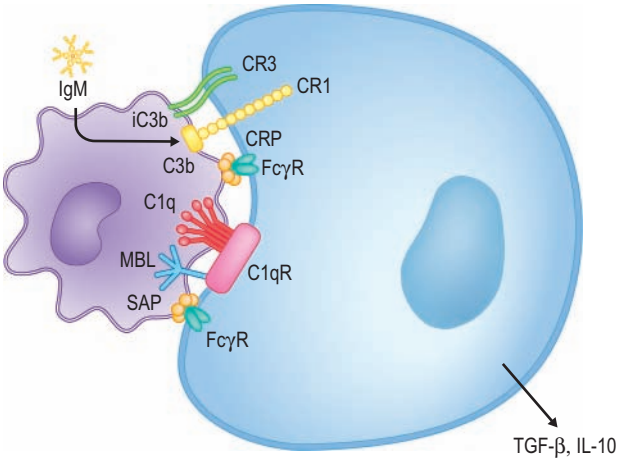


FIG. 40.6 Pathways of Opsonization of Apoptotic Cells by Complement. Innate recognition of apoptotic cells by natural immunoglobulin M *(IgM)*, cross-reactive protein *(CRP)*, serum amyloid P *(SAP)*, C1q, and mannose-binding lectin *(MBL)* is shown. Each reaction activates complement, leading to opsonization by C3b and iC3b. In addition, C1q and MBL bind to collectin receptors, and CRP and SAP bind to FcγR on macrophages. Cytokine responses to apoptotic cells opsonized by complement include the antiinflammatory cytokines transforming growth factor-β *(TGF-β)*, and interleukin-10 *(IL-10)*.

COMPLEMENT DEFICIENCIES

Genetics and Incidence

Complete genetic deficiencies of complement proteins are rare, with an estimated combined prevalence of 0.03% for any inherited complete deficiency (excluding MBL deficiency) in the general population.[2,48–50] For most components, inheritance is autosomal and expression is codominant, so complete deficiency is homozygous recessive and heterozygotes express half levels. There are two C4 genes (*C4A* and *C4B*), so a range of partial deficiencies can be observed. All cases of C1-INH deficiency have been heterozygous, and P deficiency is X-linked. MBL is found in multiple allelic forms with different levels of expression ranging from 5 nanograms per milliliter (ng/mL) to more than 5 micrograms per milliliter (μg/mL) in plasma. Deficiencies specific to the LP are not detected by the screening assays described below but can be determined by specific assays.[51]

The most common clinical presentations of patients with complement deficiencies are recurrent infections with encapsulated bacteria, neisserial infections, and systemic autoimmune disease

TABLE 40.3 Clinical Effects of Genetic Complement Deficiency

Deficient Component	Resulting Defect	Clinical Associations
C1q, C1r, C1s, C4, or C2	Inability to activate the CP	Systemic lupus erythematosus, bacterial infections
Factor D (FD), factor B (FB)	Inability to activate the AP	Infections (encapsulated bacteria)
MBL, MASP-2	Decreased or absent ability to activate the LP	Recurrent childhood infections, pyogenic bacteria
Properdin	Reduced efficiency of AP activation	Neisserial infections
C3	Decreased opsonization. No MAC. No activation of AP. Decreased inflammation (no C3a).	Recurrent childhood infections, *N. meningitidis, Streptococcus pneumoniae*, other encapsulated bacteria; autoimmune disease (uncommon)
FH, FI,	Lack of regulation of fluid-phase C3 convertases, severe acquired C3 deficiency (C3NeFs)	Infections, membranoproli ferative glomerulonephritis, C3G
C5, C6, C7, C8, C9	Inability to form the MAC	Infection—recurrent, disseminated neisserial
Serum carboxypeptidase-N	Failure to control C3a, C5a, bradykinin	Recurrent angioedema
C1-INH	Loss of regulation of C1 and bradykinin	Recurrent angioedema (HAE)
FH, FI, CD46 (haploinsufficiency)	Decreased regulation of C3 convertases	Atypical hemolytic–uremic syndrome, age-related macular degeneration[a]
DAF, CD59	Failure to regulate complement activation on autologous cells (especially red blood cells)	Paroxysmal nocturnal hemoglobinuria (PNH), early-onset protein-losing enteropathy and thrombosis, PLE (DAF)

[a]Heterozygous C3 and factor B variants that lead to a gain of function cause atypical hemolytic–uremic syndrome (aHUS), age-related macular degeneration (AMD), and C3G.
C1 INH, C1 esterase inhibitor; *C3G*, C3 glomerulopathy; *DAF*, decay-accelerating factor; *HAE*, hereditary angioedema; *MAC*, membrane attack complex; *MASP*, MBL-associated serine protease; *MBL*, mannan-binding lectin; *NeF*, nephritic factor (stabilizing autoAb to convertase).

(Table 40.3). Populations with these disease manifestations have a much higher incidence of complement deficiency. For example, in Caucasian patients with SLE, the incidence of C2 deficiency is nearly 1%, 100-fold higher than in the general population. Screening of patients with autoimmune disease for complement deficiencies is useful, as these individuals are at higher risk for certain disease manifestations and may be at greater risk for infectious complications. Complement deficiency is found in as many as 20% of patients with recurrent disseminated neisserial infections. Evaluation of complement function is highly recommended in patients with systemic neisserial infections so that appropriate immunization and antibiotic prophylaxis can be initiated.

Complement deficiencies are most readily detected by hemolytic screening assays (the CH_{50} and AH_{50}), which determine the dilution of patient's serum needed to lyse 50% of erythrocytes sensitive to the CP (CH_{50}) or the AP (AH_{50}).[52] Deficiency of any C1 subcomponent, or any of the other CP components (C2 to C8), will result in little or no lysis in the CH_{50} (CH_{50} values <5%). C9-deficient patients may have residual activity in this assay (CH_{50} values <30%). Little or no lysis is observed in the AH_{50} assay if factor D, P, or any of the components C3 to C9 are deficient. By comparing the results of the two assays, it is possible to narrow down the search for the deficient component.

Hemolytic and antigenic assays may be done for each individual component to confirm the deficiency.

CLINICAL PEARLS

Value of Screening for Complement Deficiencies

1. Patients with recurrent bacterial infections and normal white blood cells and immunoglobulins should be analyzed for a complement deficiency (obtain CH_{50} and AP_{50}).
2. Patients with recurrent or disseminated neisserial infection should be evaluated for deficiency of C3–C9 by CH_{50} and for properdin by AP_{50}.
3. Prophylactic antibiotics and immunization should be considered in complement-deficient individuals, especially for pneumococcus and neisserial species.
4. Patients with systemic lupus erythematosus (SLE) (especially young children and those with familial lupus, and recurrent bacterial infections) should be screened with CH_{50}.

Classical Pathway Deficiencies

Patients with deficiencies of early CP components (C1, C4, or C2) are most commonly identified as having systemic autoimmune disease but are also at increased risk of infection.[48–50] The primary infectious agents in these patients are encapsulated bacteria, *S. pneumoniae, H. influenzae, N. meningitidis*, and *S. agalactiae*, which are cleared by Ab and CP opsonization.

C1 Deficiency

C1-deficient patients most commonly lack C1q, but C1r or C1s deficiency also results in nonfunctional C1 and no CP activity. Absence of C1q is highly associated with the development of SLE, with an incidence of 90%.[48–50,52] It has been proposed that this association is related to defective clearance of apoptotic cells.[53] Apoptotic cells may be opsonized by IgM, or pentraxins, leading to activation of the CP, which may be initiated by IgM or pentraxin (CRP and SAP). Cells can also be cleared by direct C1q binding, leading to attachment and uptake through other phagocytic receptors (e.g., phosphatidylserine receptor). Other proposed mechanisms to account for the strong association between C1 and C4 deficiency (see below) and SLE include defective immune complex clearance and defective development and maintenance of B-cell tolerance.

C4 Deficiency

There are two C4 genes, *C4A* and *C4B*, located within the major histocompatibility complex (MHC) on chromosome 6.[50] The two forms of C4 protein have similar function, but different substrate preferences for the covalent binding reaction that occurs on activation to C4b. C4A is more efficient in attaching to amino groups on proteins, such as immune complexes, whereas C4B is more efficient in attaching to carbohydrates. Complete C4 deficiency requires four null alleles and is rarely found but is highly associated with SLE (75% incidence). Partial C4 deficiencies with one to three null alleles, however, are relatively common, found in up to 25% of individuals. Complete C4A deficiency is greatly overrepresented in the SLE population. C4A deficiencies are found in about 1% of the general population and 10% to 15% of patients with SLE. Complete C4B deficiencies are more commonly associated with bacterial infections, suggesting

that the functionally different C4 genes contribute differently to host defense and autoimmunity.

C2 Deficiency

The gene encoding C2 is also located within the MHC. C2 deficiency is the most common complete complement deficiency, with about a 0.01% incidence in the population.[48–50] About half of C2-deficient individuals are clinically healthy. The remaining individuals have recurrent pyogenic infections and/or rheumatological diseases. The most common infectious agents are *S. pneumoniae, H. influenzae, N. meningitidis*, and *S. agalactiae*. Infections are invasive and mainly occur in childhood, suggesting that the defect may be partially overcome by development of acquired immune defenses. Rheumatological diseases associated with C2 deficiency include SLE (15%), vasculitis, polymyositis, and Henoch-Schönlein purpura. SLE associated with C2 deficiency has some features that distinguish it from other types of SLE; these features include equal expression in males and females, early onset, increased photosensitivity, decreased incidence of renal disease, lower frequency of anti-dsDNA Ab, and higher frequency of anti-SSA/Ro and anti-C1q Ab.[48–50]

Lectin Pathway Deficiencies

MBL deficiency was originally found as a serum defect in the opsonization of yeast in pediatric patients with recurrent infections. There are multiple MBL polymorphisms in the population, in both the promoter and coding regions of the gene, and MBL deficiency is common (estimated to be 14% in the normal Swedish population).[48] In addition to the association of MBL deficiency in children with recurrent infections, there is a two- to threefold increased frequency of MBL deficiency in SLE, and these individuals have more frequent and more severe infections during the course of their disease. Serious infectious complications are also more frequent in the subgroups of patients with cystic fibrosis and rheumatoid arthritis (RA) with MBL deficiency.

Although rare, in one reported homozygous MASP-2 deficiency,[51] the patient was asymptomatic until the age of 13 years when he was diagnosed with ulcerative colitis. Additional autoimmune manifestations developed along with recurrent severe infections with *S. pneumoniae*.

Alternative Pathway Deficiencies

Individuals with complete deficiencies of factor D or P have been reported. Patients with factor D deficiency have presented with recurrent infections by *Neisseria* and other organisms. Properdin deficiency is X-linked, and patients most commonly have severe childhood infections with *N. meningitidis*.[13,48,52]

C3 Deficiency

C3 is central to all three complement activation pathways. Nineteen families with primary inherited deficiency of C3 have been reported. The most common presentation is recurrent life-threatening infections before the age of 2 years, sometimes followed by immune complex disease. The infections observed are primarily respiratory tract infections (48%) and meningitis (34%) with a variety of pathogens, especially encapsulated bacteria. The organisms most often involved are *N. meningitidis* and *S. pneumoniae*, but other encapsulated gram-negative and gram-positive bacteria have also been observed. Recurrent infections are seen in more than 50% of patients with C3 deficiency. This clinical presentation is similar to that seen in hypogammaglobulinemia.

Acquired C3 Deficiency: Genetic Deficiencies of FH and FI and C3 and C4 Nephritic Factors

Factors H and I are required to control C3 convertase in the fluid phase of the AP. Complete deficiency of either protein results in C3 cleavage and depletion to very low levels. C5, factor B, and P levels may also be reduced. The clinical presentation of patients with FH or FI deficiency resembles that of patients with primary C3 deficiency. The highest disease association is recurrent infection with *N. meningitidis* and *S. pneumoniae*, and there is also an increased incidence of SLE. FH deficiency is more commonly associated with renal disease compared with C3 or FI deficiency (73% of individuals with FH deficiency compared with 13% of individuals with FI deficiency and 26% of those with C3 deficiency).

Nephritic factors (NeFs) are autoantibodies specific to the CP or the AP C3 convertase (C4b2a or C3bBb) or the AP C5 convertase that stabilizes these enzyme complexes and prevents normal regulatory control. The AP C3Nef induces unregulated complement activation, resulting in acquired C3 deficiency. NeFs are often associated with C3 glomerulopathy and with partial lipodystrophy.

Deficiencies of Complement Receptors

Deficiencies of CR1 (CD35) and CR2 (CD21)

Complete genetic deficiencies of CR1 or CR2 have not been reported. However, partial deficiencies of CR1 on erythrocytes, B lymphocytes, and polymorphonuclear leukocytes and of CR2 on B lymphocytes have been reported in patients with SLE. Decreased CR1 on erythrocytes may be acquired as a result of immune complex clearance.[48,52]

Leukocyte Adhesion Deficiency: CR3 and CR4 Deficiency

Leukocyte adhesion deficiency (LAD; Chapter 39) is a syndrome caused by mutations of the common β_2-integrin chain, CD18, found in LFA-1, CR3, and CR4. Defects are related to adhesion and activation of phagocytic cells, and the clinical presentation includes childhood infections with pyogenic bacteria.

Deficiencies of Regulatory Proteins

Hereditary Angioedema: C1-INH Deficiency

Hereditary angioedema (HAE) is found in individuals with heterozygous (autosomal dominant pattern of inheritance) deficiency of C1-INH.[18] C1-INH is a serine protease inhibitor (serpin) with regulatory activity for C1r, C1s, MASP-1, and MASP-2 of the complement system; factor XII (Hageman factor) and kallikrein of the contact system; factor XI and thrombin of the coagulation system; and plasmin and tissue plasminogen activator (tPA) of the fibrinolytic system. Although previous studies implicated a C2 product (C2 kinin) as a mediator, more recent data, including studies in a C1-INH–deficient mouse model, indicate that bradykinin is the primary biological mediator of angioedema in HAE.[18] In the more common form of HAE (type I, 85% of patients), reduced synthesis of C1-INH is found (5% to 30% of normal), along with decreased serum C4 and C2. In type II HAE, an abnormal C1-INH is synthesized, making antigenic levels normal or elevated with reduced functional activity and decreased C4 and C2. Clinically, type I and type II HAE are indistinguishable.

HAE presents in childhood or adolescence as recurrent episodes of nonpainful, nonpruritic, and nonpitting swelling that

are subcutaneous and/or submucosal. Urticaria is not present. Episodes are self-limiting, usually peaking at 24 hours and resolving over 2 to 5 days. Attacks are variable in frequency, severity, duration, and location, and initiating factors are poorly understood. The most common areas involved are the extremities, face, genitals, and respiratory and gastrointestinal tracts. Intestinal attacks are often associated with vomiting and diarrhea and are extremely painful (partial obstruction from the bowel wall edema). Laryngeal attacks may result in life-threatening respiratory tract narrowing. Recurrent attacks continue throughout the life of the patient and may involve multiple sites or progress from one site to another. Diagnosis of HAE is suggested by family history and clinical findings. Confirmation is based on decreased C1-INH functional activity (<10% to 35% of normal). It is important to note that although C1-INH protein is decreased in type I HAE, it can be normal or even elevated in type II HAE. C4 levels are below normal in 95% of patients with HAE. Acquired forms of C1-INH deficiency have been described, usually in older patients with lymphoproliferative diseases. These are usually caused by autoantibodies to C1-INH and are distinguished from HAE by a lack of family history and decreased C1q as well as C4. The management and treatment of HAE are discussed in Chapter 46.

Paroxysmal Nocturnal Hemoglobinuria: Decay-Accelerating Factor and CD59 Deficiency

Paroxysmal nocturnal hemoglobinuria (PNH) is a rare acquired disorder in which a somatic mutation in the *PIGA* gene in a clone of bone marrow stem cell results in defective synthesis of GPI-anchored proteins. PNH is characterized clinically by intravascular hemolysis and venous thrombosis. DAF and CD59 are GPI-anchored complement regulatory proteins expressed on erythrocytes, and PNH erythrocytes are highly susceptible to lysis. Studies of individuals with isolated DAF and CD59 deficiencies indicate that hemolysis is more highly associated with CD59 deficiency. The basis for thrombosis in PNH is poorly understood. A mAb to C5 has been approved by the US Food and Drug Administration (FDA) to treat PNH.

Control of Localized Complement Activation: Atypical Hemolytic–Uremic Syndrome, Age-Related Macular Degeneration

Hemolytic–uremic syndrome (HUS) is a rare disease characterized by microangiopathic hemolytic anemia, thrombocytopenia, and acute renal failure. "Typical" HUS occurs in children and is caused by *E. coli*, mainly O157:H7, producing a shiga-like toxin. Atypical HUS (aHUS) affects primarily older children and adults and is associated with either loss-of-function mutations in complement regulators FH, FI, or CD46 or gain-of-function mutations in C3 or FB. Mutations in complement proteins are identified in approximately 50% of patients. FH mutations associated with aHUS are predominantly clustered in the C-terminal end of the molecule in CCP19–20, a region that is required for FH binding to polyanions and endothelial cells. The ability of FH to regulate fluid-phase AP activation is therefore not affected, and C3 levels are normal. These findings have led to the hypothesis that local complement regulation is essential for preventing renal disease following endothelial cell injury and that FH acts locally after binding to exposed matrix or damaged endothelium.

Genetic variants in *CFI* and *CFH* have also been associated with age-related macular degeneration (AMD), a major cause of blindness in older adults. These variants range from common polymorphisms, conferring relatively low to moderate risk, to rare variants with nearly complete penetrance and high risk. FH polymorphism (Tyr/His402) is commonly associated with AMD[54] and is located in CCP7 (in a region of FH that binds heparin and CRP).[23,55,56] As is the case for the mutations associated with aHUS , this region of FH is not required for regulation of the fluid-phase AP convertase. AMD develops when abnormal deposits of a protein, termed drusen, form in the retina. Recent findings support the view that the local inflammatory response, including complement activation with MAC deposition, damages the retina, leading to vision loss. Although complement factors are not the only genes linked to AMD, common variants are estimated to account for more than 50% of cases and rare variants commonly lead to haploinsufficiency in about 10% of cases. These findings are driving the development of new complement-based therapeutics that could provide protection from a very significant form of age-related visual loss.

COMPLEMENT IN DISEASE

Measurement of Complement in a Clinical Setting

Laboratory tests for complement include functional assays for the CP (CH_{50}), the AP (AH_{50}), and the LP (LP_{50}) as well as antigenic levels and functional assays for each of the individual components. The CH_{50} is a hemolytic assay in which sheep erythrocytes sensitized with rabbit anti-sheep RBC Ab are incubated with serial dilutions of the patient's serum. The titer is the reciprocal of the dilution of serum at which 50% of the sheep erythrocytes are lysed. The CH_{50} requires all of the CP and terminal components (C1 to C9). Diseases associated with CP activation result in decreased CH_{50}, C4, and C3 levels. These are primarily immune complex–associated diseases, both autoimmune and infectious, and are listed in Table 40.4.

Another cause of selective CP activation is essentially a laboratory artifact, in which clotting of the blood sample in the cold is associated with consumption of the early CP. Plasma CH_{50} is often normal, but the serum CH_{50} value is markedly decreased. C3 and C4 antigenic tests may be normal or moderately reduced, but their functional activity is lost. In these situations, the clinician should consider diseases associated with cold reactive Ab: namely, cryoglobulinemias and cold agglutinin syndrome.

A comparable assay for the AP (AH_{50}) uses a buffer that blocks CP activation and employs rabbit erythrocytes in place of sensitized sheep erythrocytes. Rabbit erythrocytes spontaneously activate the human AP and are lysed in the assay. The AH_{50} requires all of the AP and terminal components (factor B, D, and P and C3 to C9). The combined use of the CH_{50} and AH_{50} is the most effective screening method for genetic deficiencies of complement components. Complete deficiency will generally result in titers of < 5% in one or both assays. Because C3 to C9 are common to both pathways, the combined results of the two assays can rapidly determine whether the deficiency is one of these shared components, one of the CP components (C1, C2, C4), or one of the AP components (factors B, D, P).

Properdin deficiency results in low, but not absent, lysis in the AH_{50}, and patients with C9 deficiency may have values up to 30% of normal in the CH_{50}. The AP is activated in gram-negative sepsis, aHUS, AMD, C3 glomerulopathy, IgA nephropathy, and PNH. As described above, the majority of these conditions are associated with genetic variants in complement proteins. Screening for complement genetic mutations that includes

a multigene panel is available in many laboratories. The gene panel offered may vary slightly among the different laboratories conducting these tests. All variants are usually reported according to Human Genome Variation Society (HGVS) nomenclature and classified based on the guidelines established by the joint consensus of the American College of Medical Genetics (ACMG) and the Association of Molecular Pathology. Laboratories that specialize in functional analysis of genetic variants can further assist in defining the significance of the variants. Laboratory values may show decreased C3 with decreased or normal CH_{50}, and normal C4 levels (see Table 40.4). However, normal C3 levels do not exclude the presence of mutations in complement regulatory proteins, because C3 is an acute-phase reactant with elevated synthesis during disease states. Also, rarely is the predisease value known and, furthermore, there is a rather wide normal range.

Evaluation of AP function also includes measurement of serum concentrations of FH, FI, anti-FH Ab, and CD46/MCP expression on leukocytes by flow cytometry. Since most patients carry a heterozygous mutation in a complement protein, patients in whom the mutant protein is not expressed or is dysfunctional would be expected to have a 50% functional level. Measurement of serum antigenic levels will detect ~25% of patients with mutations in *CFH* and ~40% of mutations in *CFI*. Flow cytometry to measure CD46/MCP expression will detect ~75% of the mutations.

LP function (and MBL deficiency) is determined by using a specific ELISA, in which the patient's serum is placed into wells coated with mannan. Binding MBL and activation of the LP results in the deposition of C4b and C4d, which are detected with mAbs. MBL levels may also be determined antigenically.

Heterozygous C1-INH deficiency, as described above, is associated with the clinical syndrome of HAE.[18] The diagnosis can be made on the basis of clinical findings and family history. C1-INH activity is reduced in these patients, and C4 protein is also low in 95% of patients, especially during attacks of edema. In type I HAE (85% of cases), C1-INH protein levels are low, but in type II HAE (15% of cases), an abnormal C1-INH protein is made, and antigenic levels are normal or elevated. There is an acquired form of C1-INH deficiency associated with autoantibodies to the inhibitor commonly seen in patients with lymphoma. In this case, low C1-INH is usually accompanied by decreased C1q as well as C4 and C2.

In clinical practice, evaluation of complement levels and genetic variants may be useful in a variety of circumstances. Initial consideration of complement deficiency may be appropriate in patients presenting with autoimmune conditions or repetitive pyogenic infections in the setting of a normal white blood cell (WBC) count and normal quantitative immunoglobulin levels (see Table 40.3). The complement profile can also be helpful in differential diagnoses of SLE and its look-alikes (see Table 40.4). Monitoring complement levels is frequently used to follow disease activity and response to therapy in a variety of conditions. In addition, identification and characterization of genetic variants helps to predict the risk of recurrence of a disease.

Role of Complement in Specific Immunological Diseases

Complement activation is involved in the pathogenesis of many immunological diseases. A general concept emerging from current research in this area is that most of the pathogenic effects of complement depend on the generation of C5a and the MAC. Further, it is becoming recognized that regardless of the initial activation mechanism, AP amplification is often needed to produce sufficient quantities of these mediators to cause disease. Finally, systems biology approaches increasingly reveal that members of the complement cascade interact with other inflammatory mediators, resulting in diseases that are a product of complex gene–environment interactions, such as asthma and Alzheimer disease.

Systemic Lupus Erythematosus (Chapter 52)

CLINICAL PEARLS

Complement Tests for Diagnosis and Monitoring of Systemic Lupus Erythematosus

- A low C4 and C3 assist in the diagnosis of systemic lupus erythematosus (SLE).
- Decreased C3 and C4 are associated with increased severity of disease, and especially with lupus nephritis.
- On serial observations, decreasing C3 and C4 levels predict and help to establish an SLE flare-up.
- Note: Decreases in C4 may precede decreases in C3.
- Remission after treatment of lupus often shows return toward normal levels of C4, followed by increases in C3.
- Note: Patients with SLE who have partial C4 deficiency may have persistently low C4 levels.
- Complete absence of CH_{50} implies the existence of a hereditary deficiency of one of the classical complement pathway components, usually C1q, C4, or C2.

Complement plays a dual role in SLE.[48,49,57,58] There is a strong association of genetic deficiencies of C1q, C1r, C1s, C4, C2, and, to a lesser degree, C3 with SLE, indicating a protective role. Three main complement-dependent mechanisms have been proposed: (1) complement-dependent clearance of immune complexes; (2) modulating the adaptive immune system, particularly through the development and maintenance of self-tolerance in B lymphocytes; and (3) a requirement for complement in the clearance of apoptotic cells and potential autoantigens released from damaged cells. The pathogenesis of SLE results, in large part, from an inflammatory response to immune complexes formed by autoantibodies (e.g., Ab to double-stranded DNA [anti-dsDNA]) binding to antigens from dead and dying cells. However, complement activation is believed to play a pathogenic role in tissue damage induced by autoantibodies in SLE. There is evidence for complement activation in skin and renal lesions of patients with SLE, as well as in autoantibody-mediated hemolytic anemia and thrombocytopenia.

Antiphospholipid Syndrome (Chapter 61)

Antiphospholipid syndrome is characterized by antiphospholipid Ab, recurrent fetal loss, vascular thrombosis, and thrombocytopenia. Antiphospholipid Abs are found in 50% of patients with SLE, and thrombotic events occur in about 50% of them. Antiphospholipid Abs identified in patients without SLE have similar clinical consequences. Disease pathogenesis has been attributed to the procoagulant effects of antiphospholipid Abs. A mouse model of antiphospholipid Ab syndrome has been used to demonstrate that injection of pregnant mice with human IgG antiphospholipid Ab results in fetal loss and wasting. In this model, complement is required for pathogenesis, and treatment with complement inhibitors is protective. Studies in the mouse model are consistent with initial complement activation by antiphospholipid Ab bound to the decidua, followed by C5a

generation and recruitment of neutrophils. The AP as well as the CP was required for pathology. Interestingly, C3 deposition in the decidua was decreased if neutrophils were depleted, suggesting an amplification pathway mediated either by tissue damage or by neutrophil release of complement components.

Rheumatoid Arthritis (Chapter 53)

Patients with RA generally have normal or elevated complement values systemically.[58,59] There is, however, evidence for local complement activation in joint fluid, in synovia, and in rheumatoid nodules. In addition to being elevated in the joints of patients with RA, complement activation products are also found in patients with osteoarthritis, SLE, Reiter syndrome, and gout. Concentrations of C3a and C5a in joint fluid are higher in RA than in other types of arthritis. An important role for complement activation in the pathogenesis of RA is suggested by studies in two animal models—collagen-induced arthritis and the K/BxN-derived Ab transfer model. In the first model, inflammatory joint disease was ameliorated by treatment with an Ab to C5 that blocks its cleavage, preventing generation of C5a and the MAC. In the second model, disease was prevented by genetic deficiency of factor B, but not C4, indicating an essential involvement of the AP.

Vasculitis (Chapters 59 and 60)

Human vasculitides encompass a spectrum of disease mechanisms and clinical manifestations. Some, such as giant-cell arteritis and the antineutrophil cytoplasmic Ab (ANCA)–associated vasculitides, granulomatosis with microscopic polyangiitis, and eosinophilic granulomatosis with polyangiitis, are not usually associated with local complement deposition or evidence of systemic complement depletion. Despite that, a "self-fueling inflammatory amplification loop," as a result of the generation of C5a by activated neutrophils and neutrophil priming by C5a, appears to be able to drive necrotizing vascular injury. Additionally, in vasculitides associated with circulating immune complexes, C3b, MAC, and/or AP components are deposited in lesions, and complement profiles consistent with CP and/or AP activation are found (see Table 40.4).

Immunological Renal Diseases (Chapter 68)

Complement activation is evident in most types of glomerulonephritis, with the site and pathway of activation dependent on the location of immune complex or autoantibody deposition. AP activation has been identified in IgA nephropathy, poststreptococcal glomerulonephritis, and C3 glomerulopathy (C3G). More recent results have implicated activation of the LP in IgA nephropathy. Glomerular deposition of MBL has been associated with greater histological damage and higher proteinuria.

C3 glomerulopathy is a chronic progressive form of glomerulonephritis characterized by complement dysregulation, which results in prominent C3 deposition in kidney biopsies on immunofluorescence. C3 glomerulopathy has been historically called MPGN and was divided into three histological groups, designated type I, II, and III based on electron microscopy findings. Given that the historical classification of MPGN did not help delineate disease pathogenesis and as MPGN was increasingly being viewed as immunoglobulin-mediated (associated with CP activation) and non–immunoglobulin-mediated (associated with AP activation), the term C3 glomerulopathy was introduced in 2010. *C3 glomerulopathy* is used to describe those cases of glomerular involvement secondary to the AP activation, with an emphasis on immunofluorescence microscopy findings. The two major subgroups of C3 glomerulopathy are dense deposit disease (DDD) and C3 glomerulonephritis (C3GN) based on the location of the deposits on electron microscopy. DDD was previously classified as MPGN type II.

In glomerulonephritis secondary to immune complex disease (immunoglobulin-mediated, such as SLE and forms of MPGN), complement activation is primarily by the CP, and C4 is detected along with C3 and IgG in glomerular deposits. Complement activation contributes to renal disease by attracting and activating inflammatory cells through the anaphylatoxin C5a and by direct damage to cells through the MAC. Pathology caused by inflammatory cell infiltration is predominant when subendothelial immune complex deposition and complement activation occur.

In contrast, in non–immunoglobulin-mediated glomerulonephritis (current C3G), defects in the tightly regulated AP are believed to result in excessive activity of the C3 convertase. This can occur either in the presence of the C3 nephritic factor (C3Nef), which is a stabilizing autoantibody, or in the setting of deficient functional FH activity, either through mutations or acquired defects. C3NeF is a pathogenic autoantibody that binds to the AP C3 convertase (C3bBb), preventing its decay and regulation by FH and FI. Absence of functional FH results in unregulated C3 convertase activity, resulting in uncontrolled glomerular inflammation and renal disease. Understandably, such reclassification has helped target treatment, for example, by using plasma infusion or exchange and even the anti-C5 mAb eculizumab in certain cases.

Asthma (Chapters 43)

Asthma is a chronic inflammatory disease of the lung, in which Th2 responses to environmental allergens frequently play a critical role. The development of mice deficient in receptors for C3a and C5a has led to a new understanding of the roles of

TABLE 40.4 Complement Test Interpretation

Pathway	CH_{50}	C4	C3	Related Diseases
CP	↓	↓	↓	SLE, serum sickness, vasculitis, subacute bacterial endocarditis, MPGN (type I)
AP	↓	N	↓	Poststreptococcal glomerulonephritis, MPGN (type II)
Fluid-phase activation of the CP	↓	↓	N	C4NeF, HAE, cryoglobulinemia
Fluid-phase activation of the AP	↓	N	↓	FH or FI deficiency, C3NeF, MPGN (type II)
Acute-phase response	↑	↑	↑	Acute and chronic inflammation
Decreased CH_{50} (sample collection problems)	↓	N	N	Cryoglobulins, cold activation, sample mishandling; coagulation-associated activation
Decreased CH_{50} (biosynthetic)	↓	N	↓	Severe liver disease; decreased C3, C6, C9

AP, Alternative pathway; *CP*, classical pathway; *HAE*, hereditary angioedema; *MPGN*, membranoproliferative glomerulonephritis; *N*, normal; *NeF*, nephritic factor, *SLE*, systemic lupus erythematosus.

the complement anaphylatoxins in asthma. Several studies have demonstrated a correlation between C3a and C5a release in asthmatic lungs and the influx of eosinophils and neutrophils. C3-deficient and C3aR-deficient mice were protected from development of acute bronchoconstriction, airway inflammation, and airway hyper-responsiveness. C5a inhibition had similar effects on the response to challenge in an established allergic environment. However, in contradiction to these findings, C5 deficiency was genetically linked to susceptibility to experimental allergic asthma. Further studies found that C5a signaling (most likely through the C5aR on pulmonary DCs) during initial pulmonary exposure to allergen decreased Th2 cytokine and IgE production, thereby preventing the initiation of the allergic response. Thus, it appears that both the C3a–C3aR and C5a-C5aR axes contribute to asthma pathogenesis. However, how disruption of their homeostatic roles on different immune cells versus the bronchial epithelium contributes to asthma pathogenesis remains to be understood.

Neurological Disease

Proteins from the complement system are normally found in the central nervous system (CNS) and the peripheral nervous systems. Low levels of hemolytic complement (0.25% of serum levels) can be measured in the cerebrospinal fluid if care is taken to stabilize it with gelatin during storage. Levels of anaphylatoxins are increased in the CNS when the blood–brain barrier is impaired. Complement proteins and regulatory proteins are synthesized by glial cells and astrocytes, and their synthesis is enhanced by inflammatory cytokines, such as IL-6. There is evidence both from human multiple sclerosis (MS) (Chapter 65) and the animal model, experimental allergic encephalitis (EAE), that complement activation with the generation of the MAC contributes to the demyelination in these diseases. Generation of the MAC can lead to oligodendrocyte death, generation of inflammatory mediators, or a repair process in which myelin synthesis is decreased. Complement activation on myelin and oligodendrocytes is initiated by antimyelin Ab or directly by myelin through the CP. There is evidence of MAC formation in the cerebrospinal fluid of patients with MS, and complement depletion, inhibition, and genetic deficiency are protective in rat and mouse models of EAE.

In any disease in which the host synthesizes an autoantibody, it may bind to its target antigen and fix complement. In turn, the complement system could then contribute to tissue injury and thereby disease pathogenesis and thus be a therapeutic target. Neuromyelitis optica (NMO) and myasthenia gravis (MG) are two such examples in which the target antigen is known, autoantibodies are made, and complement activation occurs at the disease site. Moreover, in both diseases a mAb to C5 has been approved for treatment by the FDA.

There is also evidence of complement activation in degenerative neurological conditions such as Alzheimer disease. In Alzheimer disease, neurofibrillary tangles and senile plaques composed of β-amyloid and other proteins develop, resulting in neuronal loss and dementia with progressive loss of cognitive function. Complement activation products C1q, C4, C3, and MAC components, as well as clusterin (ApoJ) and vitronectin (S40), are found deposited in areas of β-amyloid, suggesting CP activation. Peptides derived from β-amyloid were shown to activate C1 directly by binding to the collagen-like domain. SAP, a component of all types of amyloid, including β-amyloid, activates the CP as well. There are limited data on the role of complement in the pathogenesis of Alzheimer disease, with some studies reporting enhanced disease following complement inhibition and another finding decreased inflammatory changes and neuronal degeneration in C1q deficiency. Finally, excessive complement activity, notably C4, has been implicated in the development of schizophrenia and has been associated with reduced numbers of synapses. This suggests that the role of complement proteins in neuropsychiatric illness extends beyond inflammation-mediated tissue damage.[60,61]

Ischemia/Reperfusion Injury

I/R injury refers to injury induced by inflammatory mediators, such as reactive oxygen intermediates produced by activated neutrophils, following the reperfusion of hypoxic tissue. Different pathways of complement activation may be important in different sites of injury, probably because of differences in expression of complement regulatory proteins and the nature of the tissue damage and the antigens exposed to the innate immune system. The primary complement mediators of tissue injury are C5a and the MAC acting locally and, in some cases, C5a acting systemically. In experimental renal I/R injury and in human tubular necrosis, the AP appears to be directly activated and neither Ab nor the CP is required. However, in intestinal I/R injury, the CP as well as the AP are required, and a natural IgM Ab to a newly exposed antigen on damaged endothelium initiates complement activation. In coronary artery ligation/reperfusion models, innate recognition of epitopes of ischemic tissue by MBL and CRP leads to lectin and CP activation, respectively.

ON THE HORIZON

Future Directions in Complement Research

- Functional analysis of polymorphisms and rare variants in complement proteins identified in genome-wide association studies (GWAS) and by next-generation sequencing of inflammatory and autoimmune diseases will be functionally characterized to provide insight into pathogenesis and treatment.
- Genetic sequencing of entire complement activation pathways and their regulators and receptors in patients with inflammatory and autoimmune diseases will reveal novel pathogenic mechanisms and approaches to diagnosis and therapy.
- RNA sequence analysis will identify "up and down" regulation of complement proteins in human disease.
- Structural analysis of complement protein complexes will lead to targeted small molecules to inhibit or enhance complement activation.
- Therapeutic trials of existing agents and those in development will dramatically refine therapy of complement-mediated diseases.
- Proteome studies in patients with infectious, inflammatory, and autoimmune diseases will reveal patterns of complement activation and biomarkers for diagnosis, disease activity, and monitoring responses to therapy.

COMPLEMENT-BASED THERAPEUTICS

The multiple roles of complement in inflammatory and autoimmune diseases make it an attractive target for therapeutic intervention. Recombinant complement inhibitors, inhibitory mAb, and peptide-based receptor inhibitors have been developed to block the detrimental effects of the complement activation fragments.[61] As described above, products of the complement cascade have many beneficial effects in host defense and the adaptive immune response. The detrimental effects of complement

activation are commonly associated with C5a and the MAC. Thus, targeting either the generation of C5a or its association with C5aR might be expected to control inflammation while maintaining other important functions, such as opsonization. An anti-C5 mAb that prevents C5 cleavage (eculizumab) has been approved for human use in the treatment of PNH,[61] atypical HUS, NMO, and MG. Other drugs targeting the C5 pathway, as well as mAb directed to components of the AP (e.g., factor B, factor D), are under investigation for C3G and AMD. Peptides and mAb directed at the C5aR, as well as upstream of C3 convertase production, have shown promise in a number of inflammatory models in animals and are being evaluated for the treatment of sepsis, reperfusion injury, asthma, and IgA nephropathy. Other approaches that are being developed will target complement regulatory proteins to specific cell or tissue targets. As the importance of this system is clarified in a variety of inflammatory diseases, it is likely that further research will establish new complement-based therapeutic agents for additional applications.

REFERENCES

With a few exceptions, reviews are cited rather than original reference due to space limitations. For a more in-depth understanding of each component of the complement pathways, readers are referred to *The Complement FactsBook*, second edition, Barnum, S and Schein, T, editors, Academic Press, 2018. For more clinically focused insights, the reader is referred to chapters on Complement by Liszewski, MK and Atkinson, JP in the online and CD-ROM clinical resource "UpToDate."

1. Ricklin D, Hajishengallis G, Yang K, et al. Complement: a key system for immune surveillance and homeostasis. *Nat Immunol*. 2010;11(9):785–797.
2. Atkinson JP. Complement system in disease. In: *Goldman-Cecil Medicine*. Vol. 1, 25th ed. 2015:240–246.
3. Dobo J, Pal G, Cervenak L, et al. The emerging roles of mannose-binding lectin-associated serine proteases (MASPs) in the lectin pathway of complement and beyond. *Immunol Rev*. 2016;274(1):98–111.
4. Lachmann PJ. The amplification loop of the complement pathways. *Adv Immunol*. 2009;104:115–149.
5. Janssen BJ, Christodoulidou A, McCarthy A, et al. Structure of C3b reveals conformational changes that underlie complement activity. *Nature*. 2006;444(7116):213–216.
6. Wiesmann C, Katschke KJ, Yin J, et al. Structure of C3b in complex with CRIg gives insights into regulation of complement activation. *Nature*. 2006;444(7116):217–220.
7. Forneris F, Wu J, Xue X, et al. Regulators of complement activity mediate inhibitory mechanisms through a common C3b-binding mode. *EMBO J*. 2016;35(10):1133–1149.
8. Gaboriaud C, Juanhuix J, Gruez A, et al. The crystal structure of the globular head of complement protein C1q provides a basis for its versatile recognition properties. *J Biol Chem*. 2003;278(47):46974–46982.
9. Gaboriaud C, Thielens NM, Gregory LA, et al. Structure and activation of the C1 complex of complement: unraveling the puzzle. *Trends Immunol*. 2004;25(7):368–373.
10. Matsushita M. Ficolins: complement-activating lectins involved in innate immunity. *J Innate Immun*. 2010;2(1):24–32.
11. Dobo J, Szakacs D, Oroszlan G, et al. MASP-3 is the exclusive pro-factor D activator in resting blood: the lectin and the alternative complement pathways are fundamentally linked. *Sci Rep*. 2016;6:31877.
12. Pihl R, Jensen L, Hansen AG, et al. Analysis of factor D isoforms in Malpuech-Michels-Mingarelli-Carnevale patients highlights the role of MASP-3 as a maturase in the alternative pathway of complement. *J Immunol*. 2017. https://doi.org/10.4049/jimmunol.1700518.
13. Kemper C, Atkinson JP, Hourcade DE. Properdin: emerging roles of a pattern-recognition molecule. *Annu Rev Immunol*. 2010;28:131–155.
14. Agarwal S, Ferreira VP, Cortes C, et al. An evaluation of the role of properdin in alternative pathway activation on *Neisseria meningitidis* and *Neisseria gonorrhoeae*. *J Immunol*. 2010;185(1):507–516.
15. Serna M, Giles JL, Morgan BP, et al. Structural basis of complement membrane attack complex formation. *Nat Commun*. 2016;7:10587.
16. Lambris JD, Ricklin D, Geisbrecht BV. Complement evasion by human pathogens. *Nat Rev Microbiol*. 2008;6(2):132–142.
17. Zipfel PF, Skerka C. Complement regulators and inhibitory proteins. *Nat Rev: Immunol*. 2009;9(10):729–740.
18. Zuraw BL. Clinical practice. Hereditary angioedema. *N Engl J Med*. 2008;359(10):1027–1036.
19. Liszewski MK, Farries TC, Lublin DM, et al. Control of the complement system. *Adv Immunol*. 1996;61:201–283.
20. Liszewski MK, Atkinson JP. Complement regulators in human disease: lessons from modern genetics. *J Inter Med*. 2015;277(3):294–305.
21. Holers VM. Complement and its receptors: new insights into human disease. *Ann Rev Immunol*. 2014;32:433–459.
22. Manuelian T, Hellwage J, Meri S, et al. Mutations in factor H reduce binding affinity to C3b and heparin and surface attachment to endothelial cells in hemolytic uremic syndrome. *J Clin Invest*. 2003;111(8):1181–1190.
23. Morgan HP, Schmidt CQ, Guariento M, et al. Structural basis for engagement by complement factor H of C3b on a self surface. *Nat Struct Mol Biol*. 2011;18(4):463–470.
24. Liszewski MK, Java A, Schramm EC, et al. Complement dysregulation and disease: insights from contemporary genetics. *Annu Rev Pathol*. 2017;12:25–52.
25. Kim DD, Song WC. Membrane complement regulatory proteins. *Clin Immunol*. 2006;118(2–3):127–136.
26. Holmquist E, Okroj M, Nodin B, et al. Sushi domain-containing protein 4 (SUSD4) inhibits complement by disrupting the formation of the classical C3 convertase. *FASEB J*. 2013;27(6):2355–2366.
27. Escudero-Esparza A, Kalchishkova N, Kurbasic E, et al. The novel complement inhibitor human CUB and Sushi multiple domains 1 (CSMD1) protein promotes factor I-mediated degradation of C4b and C3b and inhibits the membrane attack complex assembly. *FASEB J*. 2013;27(12):5083–5093.
28. Bubeck D. The making of a macromolecular machine: assembly of the membrane attack complex. *Biochemistry*. 2014;53(12):1908–1915.
29. Marnell L, Mold C, Du Clos TW. C-reactive protein: ligands, receptors and role in inflammation. *Clin Immunol*. 2005;117:104–111.
30. Bohlson SS, Fraser DA, Tenner AJ. Complement proteins C1q and MBL are pattern recognition molecules that signal immediate and long-term protective immune functions. *Mol Immunol*. 2007;44(1–3):33–43.
31. Roozendaal R, Carroll MC. Complement receptors CD21 and CD35 in humoral immunity. *Immunol Rev*. 2007;219:157–166.
32. van Lookeren Campagne M, Wiesmann C, Brown EJ. Macrophage complement receptors and pathogen clearance. *Cell Microbiol*. 2007;9(9):2095–2102.
33. Huber-Lang M, Sarma JV, Zetoune FS, et al. Generation of C5a in the absence of C3: a new complement activation pathway. *Nat Med*. 2006;12(6):682–687.
34. Klos A, Tenner AJ, Johswich KO, et al. The role of the anaphylatoxins in health and disease. *Mol Immunol*. 2009;46(14):2753–2766.
35. Ward PA. Functions of C5a receptors. *J Mol Med*. 2009;87:375–378.
36. Ojha H, Panwar HS, Gorham RD Jr, et al. Viral regulators of complement activation: structure, function and evolution. *Mol Immunol*. 2014;61(2):89–99.
37. Yu Q, Yu R, Qin X. The good and evil of complement activation in HIV-1 infection. *Cell Mol Immunol*. 2010;7(5):334–340.
38. Kopf M, Abel B, Gallimore A, et al. Complement component C3 promotes T-cell priming and lung migration to control acute influenza virus infection. *Nat Med*. 2002;8(4):373–378.
39. West EE, Kolev M, Kemper C. Complement and the regulation of T cell responses. *Annu Rev Immunol*. 2018;36:309–338.
40. Hawlisch H, Kohl J. Complement and Toll-like receptors: key regulators of adaptive immune responses. *Mol Immunol*. 2006;43(1–2):13–21.
41. Liszewski MK, Elvington M, Kulkarni HS, et al. Complement's hidden arsenal: new insights and novel functions inside the cell. *Mol Immunol*. 2017;84:2–9.
42. Arbore G, Kemper C, Kolev M. Intracellular complement—the complosome—in immune cell regulation. *Mol Immunol*. 2017;89:2–9.

43. Thurman JM, Holers VM. The central role of the alternative complement pathway in human disease. *J Immunol*. 2006;176(3):1305–1310.
44. Taylor PR, Carugati A, Fadok VA, et al. A hierarchical role for classical pathway complement proteins in the clearance of apoptotic cells in vivo. *J Exp Med*. 2000;192(3):359–366.
45. Poon IK, Hulett MD, Parish CR. Molecular mechanisms of late apoptotic/necrotic cell clearance. *Cell Death Differ*. 2010;17(3):381–397.
46. Du Clos TW, Mold C. Pentraxins (CRP. SAP) in the process of complement activation and clearance of apoptotic bodies through Fcgamma receptors. *Curr Opin Organ Transplant*. 2011;16(1):15–20.
47. Mold C, Gewurz H, Du Clos TW. Regulation of complement activation by C-reactive protein. *Immunopharmacology*. 1999;42(1–3):23–30.
48. Sjoholm AG, Jonsson G, Braconier JH, et al. Complement deficiency and disease: an update. *Mol Immunol*. 2006;43(1–2):78–85.
49. Bryan AR, Wu EY. Complement deficiencies in systemic lupus erythematosus. *Curr Allergy Asthma Rep*. 2014;14(7):448.
50. Cambridge MA. Elsevier Academic Press, 2016. Atkinson JP, Yu CY. The complement system in systemic lupus erythematosus. In: Tsokos GC, ed. *Systemic Lupus Erythematosus, Basic, Applied and Clinical Aspects*. 2016:81–112.
51. Sorensen R, Thiel S, Jensenius JC. Mannan-binding-lectin-associated serine proteases, characteristics and disease associations. *Springer Semin Immunopathol*. 2005;27(3):299–319.
52. Wen L, Atkinson JP, Giclas PC. Clinical and laboratory evaluation of complement deficiency. *J Allergy Clin Immunol*. 2004;113(4):585–593; quiz 94.
53. Seelen MA, Roos A, Daha MR. Role of complement in innate and autoimmunity. *J Nephrol*. 2005;18(6):642–653.
54. Avery RL. The plague and macular degeneration. *Ophthalmology*. 2010;117(12):2442.
55. Ferreira VP, Pangburn MK, Cortes C. Complement control protein factor H: the good, the bad, and the inadequate. *Mol Immunol*. 2010;47(13):2187–2197.
56. Makou E, Herbert AP, Barlow PN. Functional anatomy of complement factor H. *Biochemistry*. 2013;52(23):3949–3962.
57. Mayilyan KR. Complement genetics, deficiencies, and disease associations. *Protein Cell*. 2012;3(7):487–496.
58. Sturfelt G, Truedsson L. Complement in the immunopathogenesis of rheumatic disease. *Nat Rev Rheumatol*. 2012;8(8):458–468.
59. Ballanti E, Perricone C, Greco E, et al. Complement and autoimmunity. *Immunol Res*. 2013;56(2–3):477–491.
60. Sekar A, Bialas AR, de Rivera H, et al. Schizophrenia risk from complex variation of complement component 4. *Nature*. 2016;530(7589):177–183.
61. Morgan BP, Harris CL. Complement, a target for therapy in inflammatory and degenerative diseases. *Nat Rev Drug Discov*. 2015;14(12):857–877.

41

Human Immunodeficiency Virus Infection and Acquired Immunodeficiency Syndrome

Susan L. Gillespie, Javier Chinen and Mary E. Paul

There is much to celebrate when considering the quest to end the human immunodeficiency virus (HIV)/acquired immunodeficiency syndrome (AIDS) epidemic worldwide. World Health Organization (WHO) initiatives, including the scale-up of HIV testing, early initiation of treatment, more efficacious antiretroviral (ARV) medications, and improvements in disease monitoring, have resulted in people infected with HIV now living longer and healthier lives with fewer AIDS-related deaths. The number of new infections has also declined due in part to multipronged HIV prevention strategies including pre-exposure prophylaxis (PrEP), improvements in interventions to prevent mother-to-child transmission (PMTCT), and early infant diagnosis and treatment.

According to the Joint United Nations Programme on HIV/AIDS (UNAIDS), at the end of 2018, an estimated 37.9 million people worldwide were living with HIV or AIDS (Fig. 41.1).[1] The annual number of deaths from AIDS-related illness globally has fallen from a peak of 1.7 million in 2004 to 770,000 deaths in 2018. Also encouraging is the decline in new infections. Since peaking at 2.9 million infections in 1997, year by year the number of new HIV infections has continued to decline, down to 1.7 million new infections in 2018.

KEY CONCEPTS

Trends in HIV Infection

- Global rates of HIV infection have shown a slow but steady decline, but there is great heterogeneity in disease incidence among countries and regions.
- Despite progress made in adoption and implementation of HIV policies, certain populations are lagging behind in the global response, including infants, girls, and women, and members of key populations such as men who have sex with men (MSM), transgender individuals, those who inject drugs, and sex workers.
- The age demographic most affected in the developing world—that is, those aged 25–44 years—includes men and women who are economically productive and women of childbearing potential.
- Worldwide, most infections are acquired through heterosexual contact.
- In the United States, African Americans and MSM are disproportionately infected.
- The numbers of infections transmitted from mother to child are declining as access to prophylactic medications to prevent infection improves.

Similar encouraging statistics exist for children and adolescents with HIV infection. Of the 37.9 million people living with HIV worldwide, 2.8 million are children and adolescents age 0 to 19. There were 360,000 new infections in 2018 with 160,000 infections occurring in children age 0 to 9 and 190,000 in adolescents age 10 to 19. AIDS-related deaths occurred in 120,000 children and adolescents.[1]

Despite these encouraging improvements in global statistics, however, there is great heterogeneity in the progress made among individual countries and regions and among specific demographic groups. For example, although new HIV infections are declining in most countries, particularly those in sub-Saharan Africa that had been most affected by HIV, there has been an increase in the annual number of new infections in other regions of the world such as Eastern Europe, central Asia, the Middle East, and North Africa.

Certain populations are also lagging behind in the HIV response. For example, early infant diagnosis and access to antiretroviral therapy (ART) for children has fallen short of international goals, with only 58.8% of infants born to women living with HIV being tested in a timely manner and only 54.2% of children age 0 to 14 receiving ARV treatment. Girls account for 75% of new infections among adolescents. Additionally, members of specific demographic groups referred to as key populations remain at extremely high risk for acquiring HIV. Key populations and their sexual partners represent more than half of new infections. For example, gay men and other men who have sex with men (MSM), people who inject drugs, and sex workers are each more than 20 times more likely to acquire HIV compared to the general population. Transgender people are 12 times more likely.[1]

US PERSPECTIVE

The United States Centers for Disease Control and Prevention (CDC) estimate that approximately 1.2 million adults and adolescents older than age 13 were living with HIV in the US at the end of 2018, including 161,800 (13.8%) whose infection was undiagnosed. Despite ongoing prevention efforts designed to reduce the number of new cases of HIV infection, the number of new infections annually in the United States has remained stable, with an estimated 36,400 new infections in 2018.

Within the United States, there is great variability in both the geographical and demographic distribution of the disease, with more people living with HIV residing in states in the south and northeast. Certain segments of the population are disproportionately affected, specifically MSM and ethnic and racial minorities, including African Americans and Hispanic Americans (Fig. 41.2).[2]

In the United States and other developed countries, the number of children newly infected with HIV has decreased dramatically as a consequence of successful interventions against perinatal mother-to-child transmission. Among children born during 2017, 39 children were diagnosed with perinatally acquired HIV. At the same time, new pediatric AIDS cases and AIDS deaths also have plummeted, in large part as a result of powerful combinations of ARV drugs.[2]

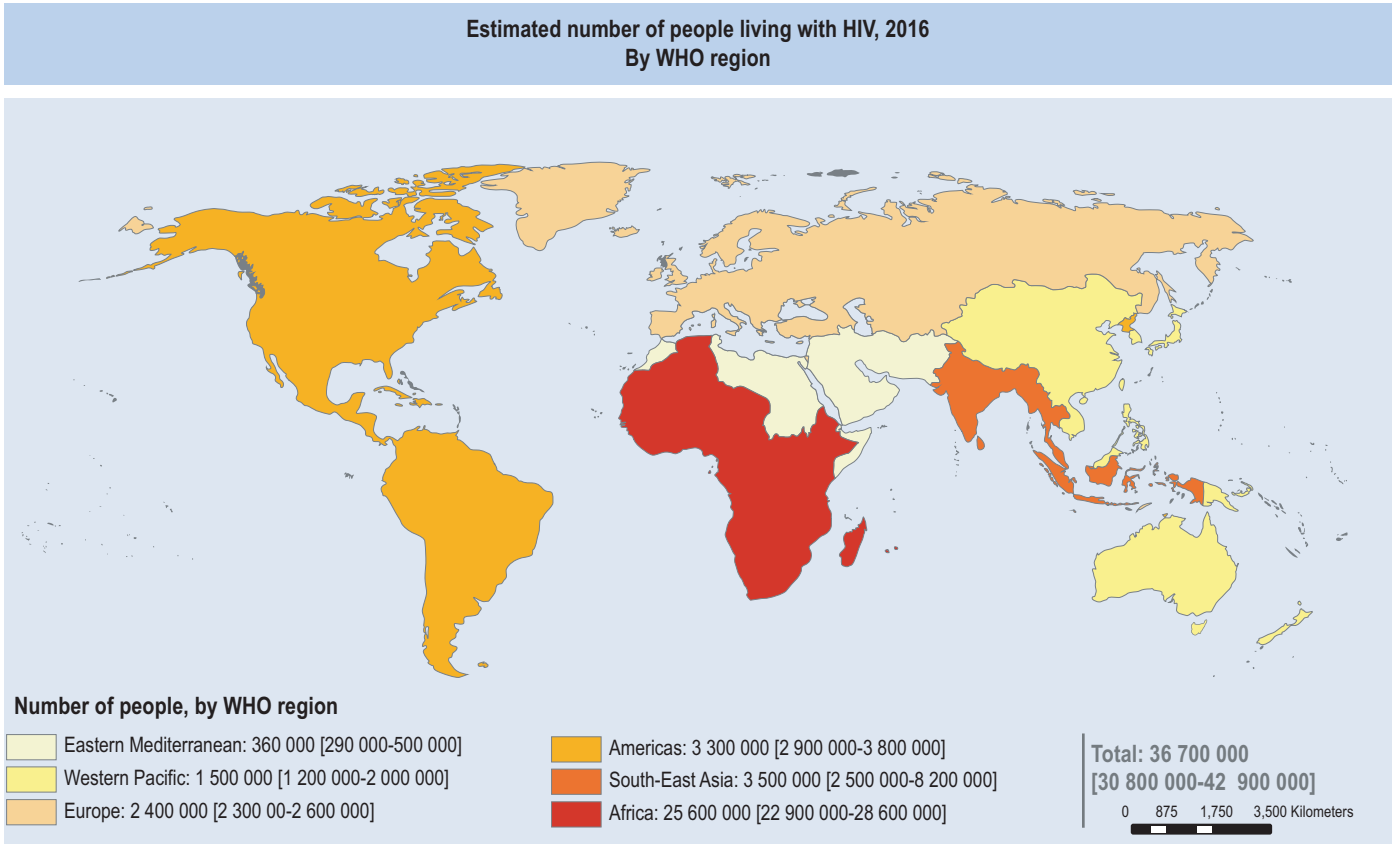


FIG. 41.1 Adults and Children Estimated to be Living with HIV in 2016: by World Health Organization (WHO) Region, 2016. *HIV*, Human immunodeficiency virus. (Also available at https://www.who.int/images/default-source/maps/hiv_all_2016.png?sfvrsn=792e6588_0.)

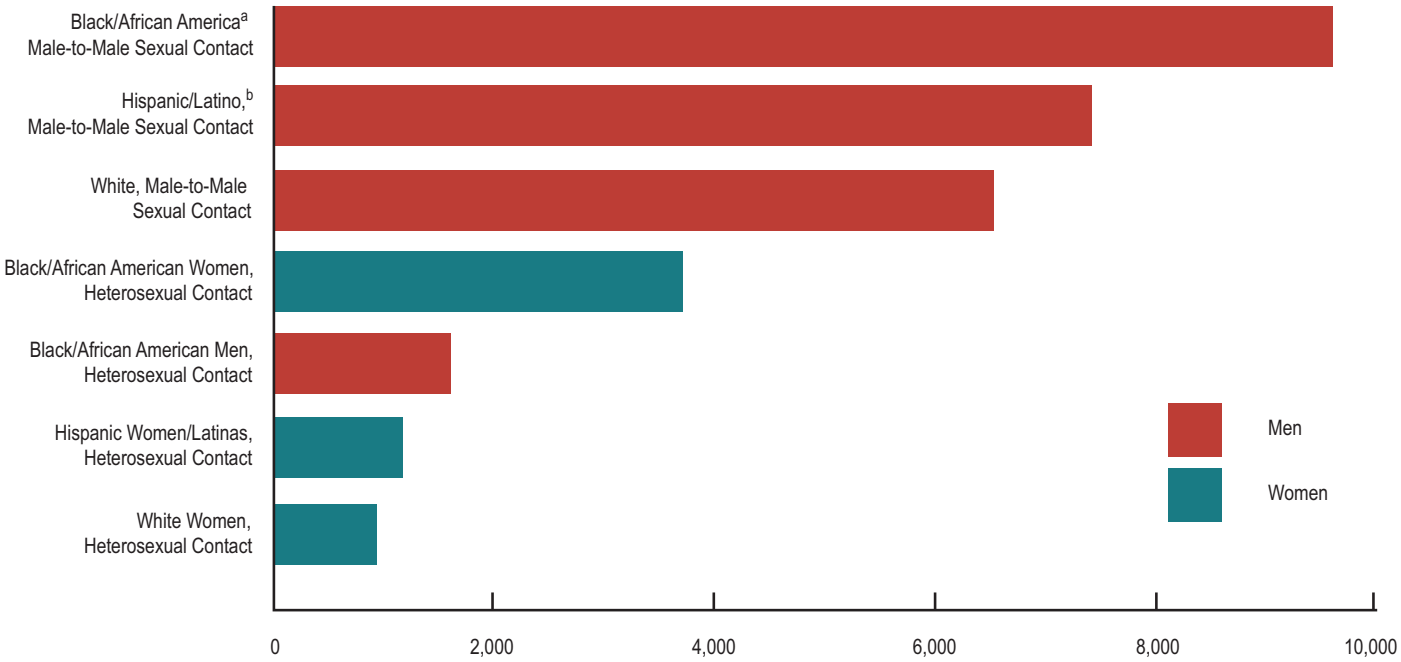


FIG. 41.2 New Diagnoses in the United States and Dependent Areas for the Most Affected Subpopulations, 2018. (https://www.cdc.gov/hiv/statistics/overview/ataglance.html.)

[a]Black refers to people having origins in any of the Black racial groups of Africa. African American is a term often used for people of African descent with ancestry in North America.

[b]Hispanic/Latino people can be of any race.

HIV IMMUNOPATHOGENESIS

HIV Life Cycle

HIV is a lentivirus that preferentially targets CD4 T cells by binding the viral glycoprotein (gp) 120 glycoproteins to two cell surface proteins, the CD4 receptor and either the CCR5 or the CXCR4 chemokine receptor (Fig. 41.3).

Other cell targets for HIV are monocytes and dendritic cells, which can be infected even though they present with significantly less expression of these receptors than CD4 T cells.

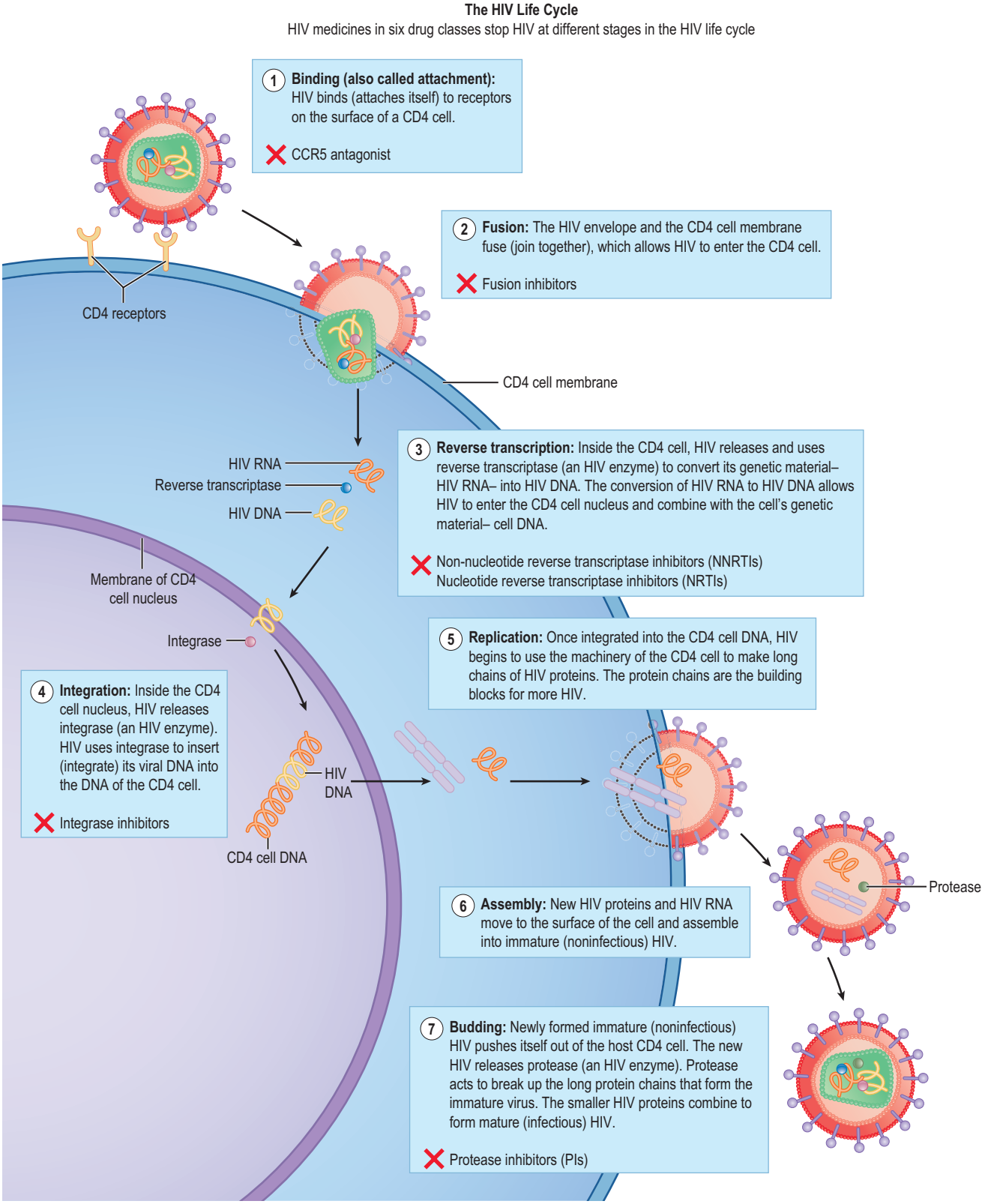


FIG. 41.3 The Human Immunodeficiency Virus *(HIV)* Life Cycle and Stage Susceptibility to Antiretroviral Therapy. (From https://www.info.nih.gov/education-materials/fact-sheets/19/73/the-hiv-life-cycle.)

After the virus is attached to the cell, the viral membrane fuses with the cell membrane and viral RNA enters the cytoplasm. It is transcribed by the viral reverse transcriptase (RT) into double-stranded DNA, which is translocated to the cell nucleus and is subsequently integrated into the cell genome as a provirus, mediated by virus integrase enzyme. Viral proteins Tat and Nef, in addition to intrinsic T-cell activation factors, induce active transcription and expression of viral RNA and proteins to produce new virus particles that exit the host cell to infect other target cells. CD4-independent viral entry has been demonstrated in B cells, astrocytes, and kidney epithelial cells; however, efficient viral replication is not likely to occur in these cells.

HIV Infection Through Mucosa

The mucosa of the gastrointestinal tract and the genital tract are the most significant ports of entry of HIV,[3] with viral infections occurring through the anorectal mucosa more efficiently than through the female genital tract, most likely due to anatomical differences in these tissues. The mucosa in the cervix and vagina has a multilayered epidermis and is covered with thick mucosa. In contrast, the anorectal mucosa has a monolayer of cylindrical epithelial cells. The diversity of HIV virion mutations within an individual shortly after anorectal infection is larger in number than after vaginal infection, supporting the protective role of the mucus and the multilayered epithelia of the female lower genital tract. HIV infection might be favored with mucosal damage due to trauma or due to inflammation secondary to other sexually transmitted diseases (Fig. 41.4).

Cells of the innate and adaptive immune system are found in mucosal tissues developing immune responses, immune tolerance, and/or inflammation when in contact with microbes or foreign substances. Anti-HIV–specific CD4 T cells and CD8 T cells are found in the cervical and vaginal mucosa of women with chronic HIV infection, although simultaneous significant viral shedding suggests that the presence of these T cells is not sufficient to control the spread of the virus.

The gastrointestinal tract contains the largest amount of lymphoid tissue in the body. HIV replication can be detected in the germinal centers of lymph nodes within a week, and viremia is found within 21 days of infection. Activation and destruction of CD4 T cells lead to massive depletion of CD4 T cells, and an increase of serum inflammatory cytokines, which might explain the flu-like syndrome that patients experience during acute HIV infection.[4] Simultaneously, HIV reservoirs get established with viral infection and integration of viral DNA in resting cells. Studies in people with chronic HIV with different degrees of viremia suggest that the presence of HIV-specific cytotoxic T cells in the gastrointestinal mucosa is the immune parameter most associated with preservation of mucosal HIV-specific CD4 T cells and viremia control. Mucosal Th17 cells are particularly susceptible to HIV infection and their depletion is associated with reduced secretion of interleukin (IL)-21, interleukin-11 (IL-11) an interleukin needed for the integrity of the mucosal barrier.[5] Natural killer (NK) cells, γδ T cells, and innate lymphoid cells in the mucosa are activated, contributing to persistent inflammation. In chronic HIV infection, these cells are eventually depleted from tissues.[6]

Chronic Immune Activation

Progressive T-cell depletion in HIV infection is induced in part by a state of chronic immune activation, which also contributes to noninfectious inflammatory complications such as cardiovascular disease.[4,7] In individuals with HIV, immune activation is caused by increased microbial translocation in the gut, direct Toll-like receptor (TLR) stimulation by HIV, and coinfection with other pathogens, resulting in high serum and tissue levels of inflammatory cytokines, such as IL-6. Lipopolysaccharide levels are also increased in peripheral blood, which in turn can activate the coagulation cascade and promote thrombosis. Low or normal levels of regulatory T cells are observed in long-term non-progressors (LTNPs), suggesting that these cells play a role in anti-HIV immunity, most specifically in modulating CD8 T-cell immunity. It has also been suggested that regulatory T cells may play a beneficial role by reducing chronic immune activation.[8]

HIV Latency and HIV Reservoirs

The most difficult challenge in the efforts to achieving cure of HIV infection is the presence of HIV reservoirs, defined as resting or nonreplicating cells infected with HIV.[9,10] Anti-HIV

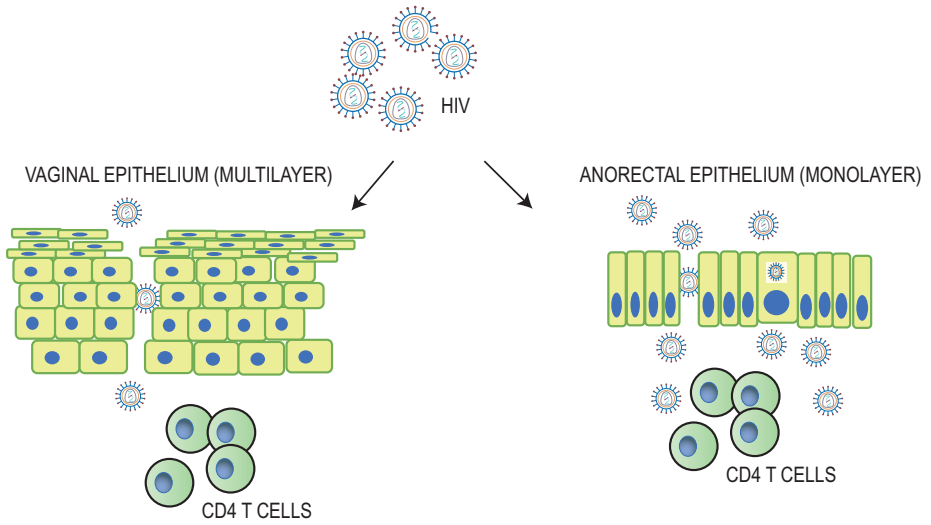


FIG. 41.4 Anorectal and Vaginal Epithelium Are Ports of Human Immunodeficiency Virus *(HIV)* Entry.

drugs are able to efficiently suppress HIV replication; however, when these drugs are stopped, activation of resting cells results in production of new virus.[11] Most proviral DNA is detected in memory CD4 T cells in lymphoid tissues and in peripheral blood. Infected T cells producing very low levels of viral RNA and proteins are most likely to avoid immunosurveillance. Molecular mechanisms resulting in reduced or near absent HIV transcription include the balance of negative and positive transcription factors, transcription interference, and DNA methylation or other epigenetic modifications. T-cell activation leads to HIV replication by changing the dynamics of transcription factors, such as Nuclear Factor Kappa B (NFκB), promoting gene expression, including expression of HIV genes. HIV reservoirs are established during the acute phase of the infection: thus the recommendation of early intensive treatment to reduce the population of latently infected CD4 T cells.[9] The reservoir increases in size with persistent HIV replication; and even though viremia decreases after successful antiretroviral treatment, HIV genetic diversity studies suggest that the number of T cells harboring proviruses remains large.[12] Based on the natural decay of HIV infected, resting T cells, which have a half-life of 44 months, it is estimated that it would take a lifetime to eliminate the HIV reservoir in a patient receiving an optimal ARV regimen.[13] Nonreplicating monocytes, astrocytes, and glial cells are other latently infected cells that form part of the HIV reservoir.

KEY CONCEPTS

HIV Reservoirs

- HIV reservoirs are groups of cells that are infected by the virus, and remain inactive for a period of months or years. When these cells are activated, viral replication occurs.
- HIV reservoirs have been difficult to target due to reduced or absent viral expression.
- HIV reservoirs are established during the early stages of infection.
- The size of HIV reservoirs is proportional to the degree of HIV viremia.

ANTI-HIV IMMUNITY

Immunity against HIV depends mostly on specific cytotoxic CD8 T cells, which recognize infected cells by their expression of viral proteins in the context of human leukocyte antigen (HLA) class I molecules (Fig. 41.5). This interaction results in CD8 T-cell activation with release of cytokines, enzymes, and cytotoxic granules, leading to death of the infected cell.[4,14] These antiviral cells are most efficient when certain combinations of HLA and virus strain occur in the host, such as the presence of cells bearing the HLA-B27 allele and infected with a clade B viral strain.

However, HIV infection almost always results in global T-cell destruction and exhaustion, unless ARV therapy is administered. HIV reservoirs remain practically ignored by CD8 T cells, in part due to the low level of viral protein expression, and pharmacological agents are being investigated to activate quiescent cells with integrated proviral DNA.[10] Of note, the presence of anti-HIV CD8 T cells decreases after the start of ARV therapy, presumably due to the reduced viremia and the low viral protein expression of HIV-infected resting cells. In addition, viral escape mutants not recognized by CD8 T cells become predominant in individuals with chronic HIV infection.

Characterization of the anti-HIV antibody response is of priority to optimize the design of vaccines and monoclonal antibodies for active and passive immunization, respectively. These vaccines are aimed to induce an immune response that reduces or prevents HIV infection at the mucosa by interfering with cell infection, as well as helping to reduce viral load at infected tissues. Non-human primate models of HIV infection have proven that this protective effect of anti-HIV antibodies can be achieved; however, it has yet to be demonstrated in humans.[15] Antibody responses are initially directed against the gp41 portion of the Env protein, and over time develop against other HIV proteins. Several studies in individuals with HIV have identified antibodies that target epitopes common to several HIV strains, known as broadly neutralizing antibodies (Fig. 41.6).[16]

Similarly important antibody functions are mediating antibody-dependent cytotoxicity and enhancing phagocytosis. These responses and subsequent anti-HIV antibodies might

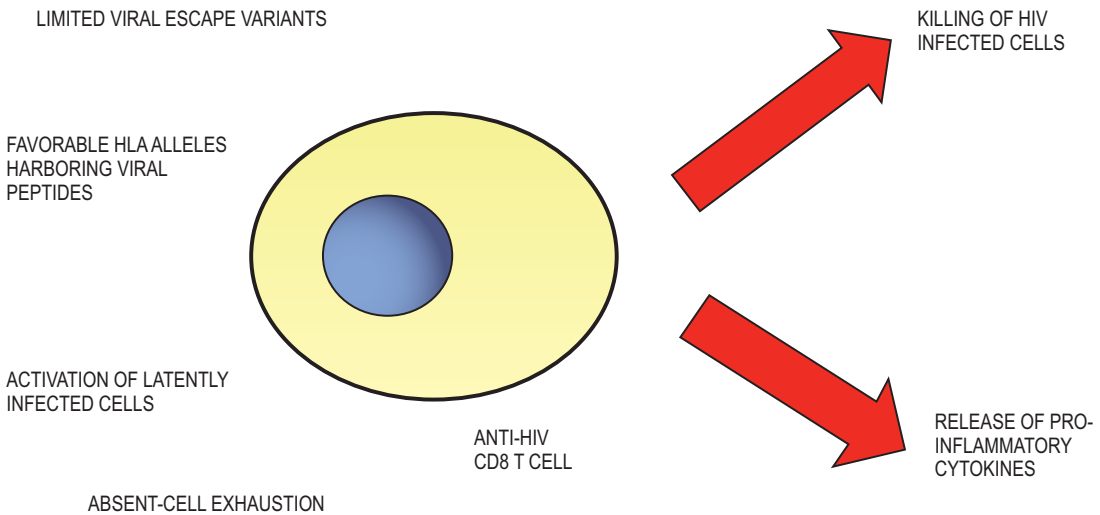


FIG. 41.5 Anti-Human Immunodeficiency Virus CTL Function (Factors Affecting CD8 T-Cell Activity). *CTL,* Cytotoxic T lymphocyte; *HIV,* human leukocyte antigen.

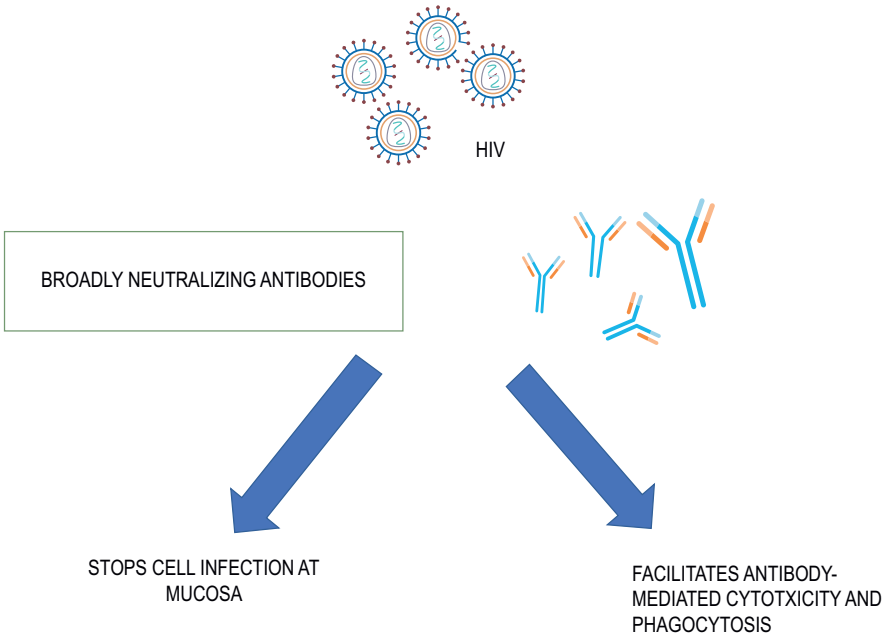


FIG. 41.6 Broadly Neutralizing Antibodies. *HIV,* Human immunodeficiency virus.

have reduced efficacy because of the rapid production of escape mutants, with subsequent generation of new antibody specificities. This is the most compelling argument explaining why antibody-based approaches are not likely to influence chronic infections. Clinical trials to test the protection against HIV infection conferred by a cocktail of broadly neutralizing antibodies are currently under way.

INNATE IMMUNITY IN HIV INFECTION

Innate immunity mechanisms are active in HIV infection with a significant role in controlling HIV viremia and optimizing the adaptive anti-HIV immunity. Pattern recognition receptors, such as TLR and retinoic acid-inducible gene I (RIG-I)-like receptors, recognize viral proteins and nucleic acids leading to the activation of immune cells, most significantly induced by type I and type II interferon pathways, through interferon regulatory factor (IRF)3 and IRF7 (Fig. 41.7).[17]

One of these proteins is interferon-inducible protein 6 (IFI6), which binds HIV DNA after reverse transcription, activating the inflammasome and inducing cell death. Cyclic GMP-AMP synthase (cGAS) is another DNA-binding protein that signals activation of innate immunity cells.

KEY CONCEPTS

Innate Immunity in HIV Infection

- Anti-HIV innate immunity mechanisms play a significant role in the control of viremia, and the optimization of antiviral adaptive immunity.
- Innate immunity includes the expression of several antiviral proteins, some of which have been linked to long-term viremia control.
- NK cells, monocytes, macrophages, and epithelia are innate immunity cells that respond to HIV infection with inducing cell death and increasing secretion of cytokines, leading to activation of the adaptive immune response and inflammation pathways.

The HIV gp120 protein can activate TLR2 and TLR4 in epithelial cells, triggering local inflammatory responses and recruiting lymphocytes to the infection site. TLR7, TLR8, and RIG-I are intracellular proteins that recognize HIV RNA and activate antiviral mechanisms. TLR8 activates the NLRP3 inflammasome and the release of IL-1β. Anti-HIV proteins that are expressed after activation of pattern recognition receptors include apolipoprotein B mRNA-editing enzyme catalytic subunit-like 3G (APOBEC3G), tripartite motif (TRIM) 5a, sterile alpha motif and histidine aspartate domain containing protein 1 (SAMHD1), tetherin, schlafen 11 (SLF11), interferon-inducible transmembrane protein (IFITM), and myxovirus-resistance protein 2 (MX2). These proteins are also called restriction factors and are the first line of defense against HIV, with inhibitory molecular mechanisms ranging from uncoating, to inhibition of reverse transcription, to interfering with viral release. Cells of the innate immune system, such as monocytes, macrophages, and NK cells, are activated as a result of cytokine release. These cytokines mediate the complex interplay of the innate immunity with the adaptive immunity, increasing viral peptide expression for immune surveillance recognition, and favoring Th1 polarization of CD4 T cells, among other responses.[18]

CLINICAL FEATURES

If HIV infection is left untreated, its natural history involves the progression through three clinical phases: acute retroviral syndrome, chronic or latent infection, and finally AIDS. Each clinical phase correlates with specific events in the interaction between HIV and the host immune system. A small percentage of patients become long term nonprogessors (LTNP), and an even smaller percentage become elite controllers (see Long-Term Non-Progressors/Elite Controllers).

Acute HIV Infection

Soon after infection, unopposed by effective host immune responses, HIV rapidly replicates and disseminates to lymphoid tissues (see Immunopathogenesis: Gastrointestinal System). During this time, HIV virus cannot be detected in plasma. As virus rapidly expands in gut-associated lymphoid tissue and

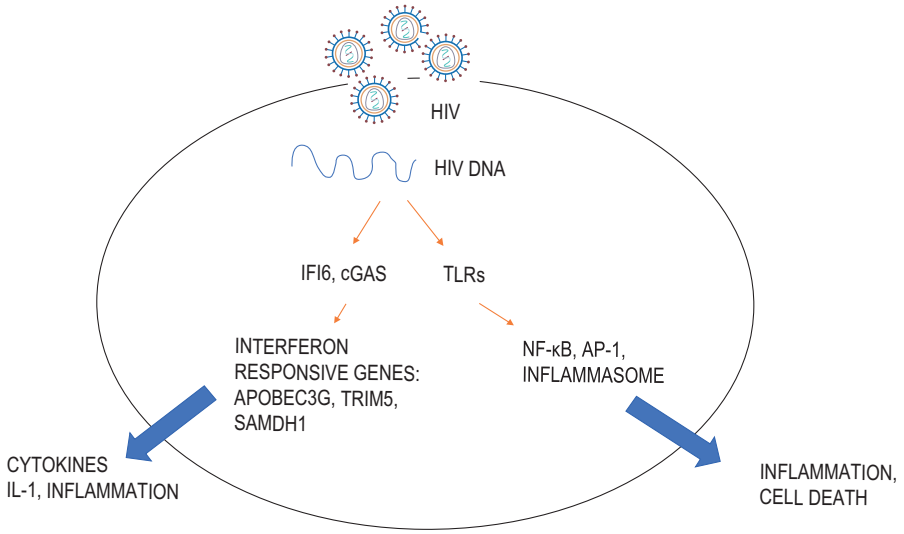


FIG. 41.7 Activation of Innate Immunity Mechanisms. *cGAS*, Cyclic GMP-AMP synthase; *HIV*, human immunodeficiency virus; *IFI6*, interferon-inducible protein 6; *IL-1*, interleukin-1; *NF-kB*, nuclear factor kappa B; *TLRs*, Toll-like receptors. For other abbreviations see text.

then spreads into the systemic circulation, there is a sharp rise in plasma viral RNA, reaching as high as 10 million copies per milliliter. This high level of viremia causes irreversible destruction of reservoirs of helper T cells and establishes viral latency, the silent integration of HIV-1 DNA into the genomes of resting T cells. Plasma viremia typically peaks 3 to 4 weeks after transmission, and then, as a result of the depletion of susceptible CD4 T cells and HIV-specific immune responses, the virus load precipitously declines, followed by more gradual decline for several weeks before reaching a set point.

The presentation of the acute retroviral syndrome associated with acute HIV infection can be highly variable, ranging from asymptomatic to severe illness requiring hospital admission. The most commonly described signs and symptoms of acute infection are nonspecific and resemble an influenza or mononucleosis-like illness, with associated fever, fatigue, myalgia, rash, and headache (Table 41.1).[19]

CLINICAL PEARLS

Acute Infection Is an Opportunity for Early HIV Diagnosis

- Acute HIV infection often causes a nonspecific viral syndrome, often described as being similar to influenza or infectious mononucleosis.
- Many patients present to a clinician at this stage, but most are not recognized as HIV infection.
- Irreparable damage to the host immune system occurs during this stage of HIV infection, resulting in chronic immune activation and the eventual collapse of the immune system.
- HIV viral latency is established during the acute infection as HIV DNA integrates with the host genome. Integrated viral genetic material makes cure of HIV impossible even after prolonged viral suppression with antiretroviral therapy (ART).
- ART initiated during the acute infection can halt the destruction of the body's memory T cells in the gut-associated lymphoid tissue (GALT) and lead to better long-term outcomes for the patient.
- Acute and early HIV infection is associated with high viral loads and increased infectiousness. ART initiated during the acute infection can lead to suppression of viral replication, lowering the viral load and decreasing the risk of onward infection.

The onset of symptoms typically occurs 2 to 4 weeks after transmission, coinciding with the period of high plasma viremia, and dissemination of virus and a burst of inflammatory cytokines. As the host develops HIV-specific immunity and the virus load decreases, CD4 and CD8 T cells recover, and the symptoms of the acute infection resolve. Although up to 90% of patients seek medical care for this illness, the nonspecific nature of the symptoms makes diagnosis of acute infection difficult. And absent a high index of suspicion, most newly infected individuals are not diagnosed until much later. Diagnosis of HIV in the acute or early phase enables the initiation of ART. Early treatment has many potential benefits including decreasing the severity of the acute infection, lowering the viral set point, and slowing disease progression. The public health implications of the acute HIV infection are enormous because the risk of transmission from individuals with acute infection appears to be much higher than that from those with established infection, in part because of the high viral load in blood and genital secretions during the acute phase.

Chronic HIV Infection

The acute infection is followed by a prolonged latent period that may last 8 to 10 years in adults but is much shorter in children. During this time, the HIV viral load fluctuates around a relatively stable set point.[20] The viral set point is a major determinant of infectivity and risk of disease progression, with higher viral loads being associated with more likely viral transmission, more rapid disease progression, and greater risk of death. The host immune response is insufficient to eradicate the infection but may be enough to contain viral replication for many years. Although commonly thought to represent a stalemate between viral replication and CD4 T-cell production, this period is actually characterized by a steady and inexorable decline of CD4 T cells (50 to 75 cells/μL of peripheral blood per year). As the infection progresses, most individuals develop clinical symptoms. The ability of the immune system to contain viral replication is overcome, and the viral load begins to increase. There is usually an inflection point in the CD4 T-cell curve marking the start of a period of more rapid decline in CD4 T-cell counts. As these counts fall, immunodeficiency, symptomatic disease, and AIDS eventually occur (Fig. 41.8).

TABLE 41.1 Clinical Signs and Symptoms of Acute HIV Infection

Features	Overall ($n = 375$) %	SEX		ROUTE OF TRANSMISSION	
		Male ($n = 355$) %	Female ($n = 23$) %	Sexual ($n = 324$) %	Injection Drug Use ($n = 34$) %
Fever	75	74	83	77	50
Fatigue	68	67	78	71	50
Myalgia	49	50	26	52	29
Skin rash	48	48	48	51	21
Headache	45	45	44	47	30
Pharyngitis	40	40	48	43	18
Cervical adenopathy	39	39	39	41	27
Arthralgia	30	30	26	28	26
Night sweats	28	28	22	30	27
Diarrhea	27	27	21	28	23

Centers for Disease Control and Prevention: US Public Health Service: Preexposure prophylaxis for the prevention of HIV infection in the United States—2017 Update: a clinical practice guideline. https://www.cdc.gov/hiv/pdf/risk/prep/cdc-hiv-prep-guidelines-2017.pdf.
Daar ES, Pilcher CD, Hecht FM. Clinical presentation and diagnosis of primary HIV-1 infection. Curr Opin HIV AIDS. 2008;3(1):10–15

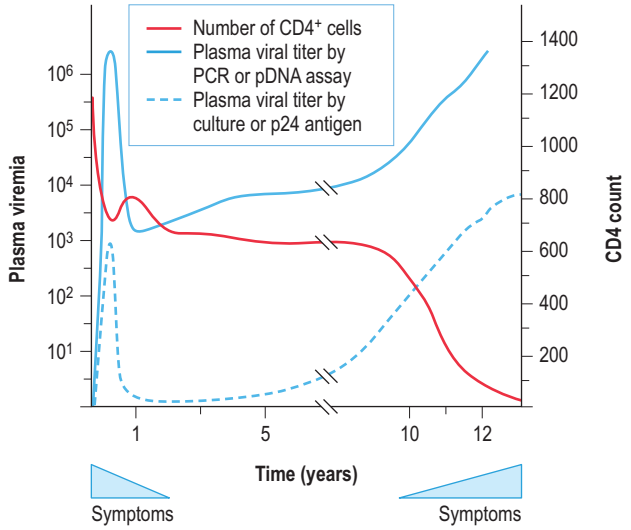


FIG. 41.8 Natural Course of Human Immunodeficiency Virus. The relationship between CD4 T-cell counts and viral loads over the course of infection in the absence of treatment. *PCR*, Polymerase chain reaction; *pDNA*, plasmid DNA. (From Baliga CS, Shearer WT. HIV/AIDS. In: Fireman P, ed. *Atlas of Allergy*. 3rd ed. Philadelphia, PA: Elsevier Science [USA]; 2005: 351–367.)

End-Stage HIV Infection: AIDS

As the peripheral blood CD4 T-cell count drops to less than 200 cells/μL, the immune system's ability to fight infection is compromised to the extent that AIDS-defining illnesses begin to appear. On the basis of the CD4 T-cell count, prophylaxis for opportunistic infections is administered (Table 41.2). Without treatment, most patients will succumb to opportunistic infections within 2 years of developing AIDS.

Long-Term Non-Progressors/Elite Controllers

A minority of individuals with HIV infection are called LTNPs because they do not progress to AIDS after a defined period in the absence of ARV therapy. A subset of these individuals are elite controllers because in addition to non-progression of disease, they are able to maintain undetectable viral loads throughout infection. Understanding the mechanisms involved in their ability to control the infection may be helpful in elucidating factors required for a functional cure of HIV.[20]

DIAGNOSIS AND MONITORING OF HIV INFECTION

Diagnostic Tests

The diagnosis of HIV infection depends on the detection of viral and host biological markers, which appear in a chronology that is typically consistent among individuals.[21] After infection there is an "eclipse period" during which no available diagnostic test is capable of detecting HIV. The earliest diagnostic markers, HIV RNA and HIV p24, can be measured 9 to 12 days and 14 to 19 days after infection, respectively (Fig. 41.9). HIV immunoglobulin is detectable approximately 3 weeks after and persists throughout infection. The time between infection and detection of HIV immunoglobulin is referred to as the seroconversion window period. Serological assays for HIV have historically been categorized as first-, second-, third-, and fourth-generation tests with improvement in assay sensitivity resulting in earlier detection of infection with each successive generation. More recently, the use of HIV test "generation" nomenclature is no longer recommended, and tests are now categorized based on the identified marker. First- and second-generation antibody tests are now referred to as IgG-sensitive tests, third-generation assays as IgM/IgG-sensitive tests, and fourth-generation as antigen-antibody immunoassays.

In 2014, the CDC and the Association of Public Health Laboratories (APHL) released an HIV laboratory testing algorithm designed to better detect acute infections and differentiate between HIV-1 and HIV-2.[21] This algorithm was updated in 2018 (Fig. 41.10).[22]

HIV Immunoassays

The CDC testing algorithm starts with an antigen-antibody immunoassay. These tests detect p24 antigen, and anti-HIV-1/HIV-2 IgM and IgG. A negative result is conclusive and generally requires no follow-up testing. The assay will be reactive if either p24 or HIV antibodies are present. Further testing is required to determine which combination of markers have been detected.

TABLE 41.2 Opportunistic Infection Prophylaxis and Treatment in Adolescents and Adults

Risk Factor	Agent	Prophylactic Medication
CD4 cell count <200 cells/μL	*Pneumocystis jiroveci*	Trimethoprim–sulfamethoxazole (TMP-SMX) or dapsone plus or minus pyrimethamine and leucovorin or aerosolized pentamidine or atovaquone
	Coccidioidomycosis	In endemic areas: fluconazole or itraconazole
CD4 T-cell count <100 cells/μL	*Toxoplasma gondii*	TMP-SMX or dapsone plus pyrimethamine plus leucovorin or atovaquone plus pyrimethamine plus leucovorin
	Histoplasmosis	In endemic areas: itraconazole
CD4 T-cell count <50 cells/μL	*Mycobacterium avium* complex (MAC)	Macrolide (clarithromycin or azithromycin) or rifabutin
	Cryptococcosis	In endemic areas: fluconazole or itraconazole
Purified protein derivative (PPD) >5 mm induration or recent tuberculosis (TB) contact but no active TB and no history of treatment for active or latent TB	*Mycobacterium tuberculosis*	Isoniazid (INH) + pyridoxine for 9 months. If unlikely to complete 9-month course and on highly active antiretroviral therapy (HAART): rifabutin plus pyrazinamide for 2 months
Contact with chickenpox or shingles in varicella-zoster seronegative individuals	Varicella-zoster	Varicella-zoster immunoglobulins (VZIGs)
Human immunodeficiency virus (HIV)–infected	*Streptococcus pneumoniae*	Pneumovax
	Meningococcus—for youth attending the military or college and consider for unvaccinated adults	Menactra
Negative anti-hepatitis B core antibodies (HBc) and previously unimmunized or underimmunized to hepatitis B	Hepatitis B	Recombivax-HB or Engerix-B
Negative anti-hepatitis A serology	Hepatitis A	Havrix

For additional information see the current US guidelines at https://clinicalinfo.hiv.gov.
From Baliga CS, Shearer WT. HIV/AIDS. In: Fireman P, ed. *Atlas of Allergy*. 3rd ed. Philadelphia, PA: Elsevier Science (USA); 2005:351–367.

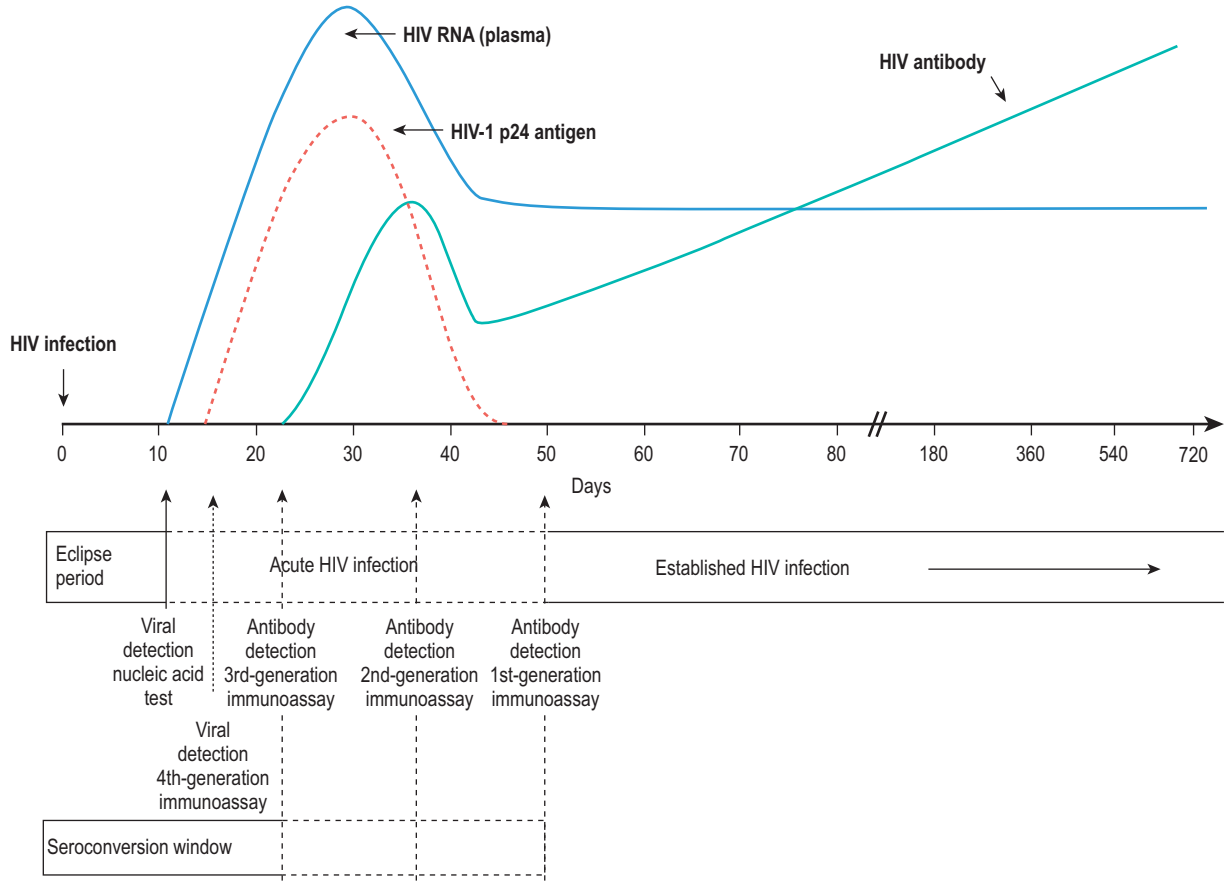


FIG. 41.9 Sequence of appearance of laboratory markers for human immunodeficiency virus type 1 *(HIV-1)* infection. (From Centers for Disease Control and Prevention and Association of Public Health Laboratories. *Laboratory Testing for the Diagnosis of HIV Infection: Updated Recommendations*. 2014. Available from https://doi.org/10.15620/cdc.23447).

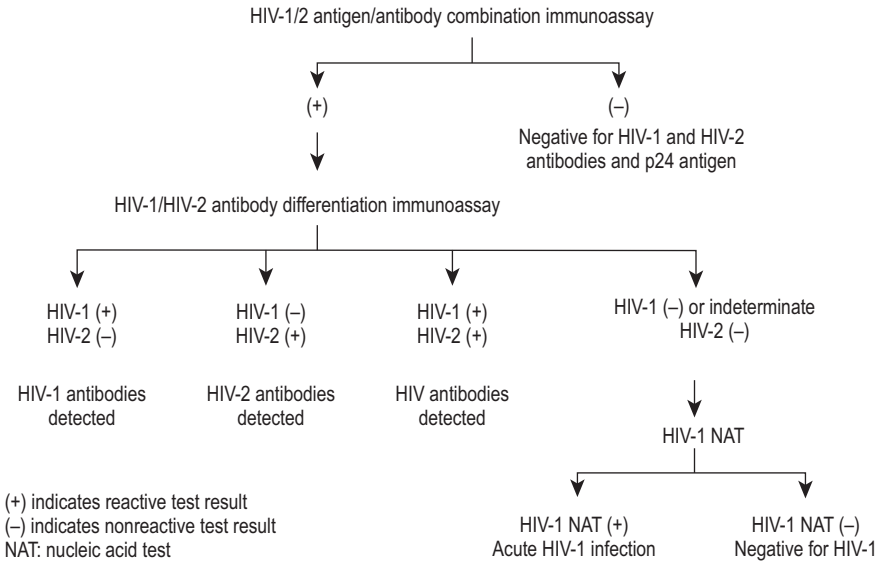


FIG. 41.10 Recommended Laboratory Human Immunodeficiency Virus *(HIV)* Testing Algorithm for Serum or Plasma Specimens. (From https://www.cdc.gov/hiv/pdf/guidelines testing recommendedlabtestingalgorithm.pdf)

To determine whether HIV-1, HIV-2, or both HIV-1 and HIV-2 antibodies are present, an HIV-1/HIV-2 differentiation assay is performed. Detection of either HIV-1 or HIV-2 antibodies confirms HIV-1 or HIV-2 monoinfection. Detection of both HIV-1 and HIV-2 antibodies confirms HIV-1/HIV-2 dual infection. If the assay is unable to confirm presence of antibodies to either HIV-1 or HIV-2 then there is a possibility of acute HIV infection and sample is then tested for HIV-1 nucleic acid. If HIV-1 nucleic acid is not detected, then HIV infection is excluded. If HIV-1 nucleic acid is detected, HIV infection in the acute stage has been diagnosed. HIV RNA detection plays a valuable role in identifying early infection before seroconversion and in confirming reactive screening tests.

HIV Nucleic Acid Amplification Tests

Nucleic acid amplification tests are used to detect HIV RNA or DNA in biological samples. HIV RNA polymerase chain reaction (PCR) is used to identify the presence of HIV RNA using qualitative assays or to quantify the virus load or the extracellular viral RNA in plasma and serum. The HIV viral load is used to monitor the effectiveness of ART in suppressing viral replication. HIV DNA PCR is a qualitative assay used to detect HIV viral DNA in peripheral blood mononuclear cells. Detecting HIV DNA often from whole blood samples collected on filter paper as dried blood spots (DBS) has allowed the early diagnosis of HIV in perinatally exposed infants and is still used for this purpose in international settings. HIV DNA PCR is not typically available for HIV diagnosis in the United States, having been largely replaced by qualitative HIV RNA PCR.

Point-of-Care Diagnostic Tests

Point-of-care (POC) HIV diagnostic tests, often described as "rapid" tests, are the most commonly used HIV screening tests. POC HIV tests are self-contained enzyme immunoassays that detect HIV antibodies in whole blood or oral fluid specimens and can be used at the POC, able to be performed by non-laboratory staff, and yield results within 30 minutes. Because results of rapid tests are available quickly, individuals can learn of their results in a single encounter. POC tests are slightly less sensitive but are as specific as laboratory assays. POC tests, often performed by lay counselors, have enabled widespread testing in resource-limited settings, such that, in 2017, 75% of people living with HIV knew their HIV status.[23]

Monitoring Tests

Once infection is confirmed, specific laboratory tests are performed at baseline and then periodically to monitor efficacy of ART, HIV disease status, and progression, to inform treatment decisions, and to identify end-organ toxicity.[24] CD4 T-lymphocyte (CD4) cell count and HIV RNA (viral load) are the two markers of ART responses and HIV disease progression.

CD4 T-Cell Count

The CD4 T-cell count is used to assess immune function, is essential for identifying those with severe immunosuppression, and is predictive of potential for disease progression and mortality. The CD4 T-cell count determines whether prophylactic medications for opportunistic infections are needed. Once a patient is on ART, an increasing CD4 T-cell count also helps confirm the efficacy of the therapy.[24] In children younger than 5 years of age, CD4 percentage is preferred because it typically remains stable in the setting of age-related changes in absolute CD4 count in this age group.[25]

HIV Viral Load

Particularly for patients on ART, plasma HIV RNA (viral load) is the most important indicator of response to therapy. Optimal viral suppression is generally defined as a viral load persistently below the level of detection (<20 to 75 copies/mL, depending on the assay used) and is usually achieved within 12 to 24 weeks of effective ART.[24] Failure to achieve maximal viral suppression or detectable virus after a period of maximal suppression may indicate virological failure attributable to drug resistance or nonadherence to ART. Higher HIV viral loads also correlate with greater risk for onward infection.

Drug Resistance: HIV Genotype Versus Phenotype

Viral resistance to ARV agents can be assessed by either HIV genotypic or phenotypic assays. Genotypic testing involves sequencing of the protease, RT, and integrase regions of the HIV genome to identify the presence of key mutations that confer anti-HIV drug resistance. Genotype testing is the preferred method of resistance testing to guide initial therapy and also to inform regimen changes in patients with suboptimal virological responses to ART.[24]

Phenotype assays assess the ability of HIV to replicate in vitro in the presence of various concentrations of ARV agents. The assay is performed by isolating certain key regulatory genes from HIV, usually protease, RT, and integrase, inserting them into standardized viral constructs containing an indicator cassette, and infecting cell lines in the presence of ARV agents. The results are compared against control viral isolates and expressed as a fold-change in viral susceptibility. Phenotyping assays remain very expensive, and large studies have failed to conclusively prove a clinical advantage of phenotype assays over genotype assays. The phenotype assay quantifies susceptibility and is typically used by experts evaluating individuals who have accumulated resistance and failed multiple regimens.[24]

KEY CONCEPTS

Other HIV Tests

- Genotyping: helps to guide the choice of antiretroviral medications in patients with resistant virus by sequencing the viral genetic code and identifying mutations that confer resistance to specific agents or classes of agents.
- Phenotyping: similar information to genotyping but not widely used; based on growing engineered viruses with a patient's virus' genes in the presence of antiretrovirals to determine their resistance.

These assays present data as if there is only one virus strain in the body; in reality, there are numerous viral strains at any one time, with many more archived in cells; these assays detect the dominant strain in the circulation, neglecting the other strains, which may account for up to 20% of the circulating viral particles. Given this limitation, when genotyping results are utilized for changing therapy, old treatment regimens and previous genotype results must be taken into consideration.

Testing for Viral Tropism and Abacavir Hypersensitivity

Additional assays should be performed before the initiation of specific ARV medications. A viral tropism assay should be performed before initiation of a CCR5 antagonist. HLA-B*57:01 testing is indicated before initiation of abacavir, as this HLA phenotype is associated with abacavir hypersensitivity in 5% to 8% of patients early in the course of treatment.

TREATMENT

Antiretroviral Therapy: Attacking the Life Cycle of HIV

Combinations of ARV medications are used to maximally inhibit HIV replication and to reduce HIV-associated morbidity and mortality. Combination ART refers specifically to a combination of at least three ARV medications inhibiting the HIV life cycle (Figs. 41.3 and 41.11).

Current US guidelines recommend initiation of ART-naïve persons with a combination of ARV medications that includes drugs from at least two classes and often involves the use of two nucleoside reverse transcriptase inhibitors (NRTIs) plus an integrase strand transfer inhibitor (INSTI), a nonnucleoside reverse transcriptase inhibitor (NNRTI), or a protease inhibitor (PI) with a pharmacokinetic enhancer (cobicistat or ritonavir). In ART-experienced patients, modification of ARV regimens and use of other classes of ARVs is guided, in part, by consideration of a number of factors, including viral resistance patterns, potential side effects, available medication formulations, pill burden, frequency of dosing, tolerability, short-term and long-term adverse event profiles, desire for pregnancy, and desire to preserve subsequent treatment options.[24]

When to Start Therapy

There are significant data to support treating all individuals with HIV infection with ART and starting treatment as soon as possible after diagnosis to reduce morbidity and mortality and to prevent transmission of HIV (see also HIV Prevention). Randomized controlled trials have demonstrated that ART should be initiated in all patients with HIV infection, regardless of disease stage. The urgency to initiate ART is greatest for patients with lower CD4 counts, in whom the absolute risk of opportunistic infections, non-AIDS morbidity, and death is highest. Standard of care suggests that rapid initiation of ART, defined as initiating immediately or within days of diagnosis, is indicated to improve survival, reduce reservoirs, and decrease long-term morbidity caused by persistent inflammation. Trials have shown that risk of development of cardiovascular, kidney and liver disease, and malignancy may be reduced by reducing viral replication and inflammation.[26–29] Earlier ART initiation appears to increase the probability of restoring normal CD4 counts, a normal CD4/CD8 ratio, and lower levels of immune activation and inflammation, as summarized in the US adult treatment guidelines.[24]

Recommendations for initiating therapy in infants and children have always been aggressive due to rapid disease progression in infants born with HIV.[25] In children less than 1 year of age, the health and survival benefit of rapid ART initiation has been demonstrated by clinical trials. Growth and development, including neurodevelopment, have been shown in trials to be significantly better in children who initiated treatment early.

Antiretroviral Agents

There are more than 20 approved ARV drugs, which are classified into six classes based on their chemical structure or the viral life-cycle step that they inhibit (see Fig. 41.11). The classes of ARV agents currently available include NRTIs, NNRTIs, PIs, fusion inhibitors, integrase inhibitors, attachment inhibitor, post-attachment inhibitors, and CCR5 antagonists.

Reverse Transcriptase Inhibitors, Protease Inhibitors, and Integrase Inhibitors

Modified versions of cellular nucleosides, NRTIs, once triphosphorylated in vivo, are incorporated into the proviral DNA by HIV RT and induce premature chain termination, thereby inhibiting successful conversion of the viral RNA to DNA. NNRTIs bind to RT and induce a conformational change such that RT is unable to bind with nucleotides. PIs act on viral protease, preventing the cleaving of the posttranslational viral polyproteins necessary for the maturation and infectivity of viral particles. Integrase inhibitors prevent strand transfer of

NRTI • Abacavir (ABC) • Emtricitabine (FTC) • Lamivudine (3TC) • Tenofovir DF (TDF) • Tenofovir alafenamide (TAF) • Zidovudine (AZT, ZDV) **NNRTI** • Doravirine (DOR) • Efavirenz (EFV) • Etravirine (ETR) • Nevirapine (NVP) • Rilpilvirine (RPV)	**PI** • Atazanavir (ATV) • Darunavir (DRV) • Fosamprenavir (FPV) • Lopinavir (LPV) • Saquinavir (SQV) • Tipranavir (TPV) **INSTI** • Bictegravir (BIC) • Dolutegravir (DTG) • Elvitegravir (EVG) • Raltegravir (RAL)	**FUSION INHIBITOR** • Enfuvirtide (ENF, T-20) **CCR5 ANTAGONIST** • Maraviroc (MVC) **ATTACHMENT INHIBITOR** • Foctemsavir **POST-ATTACHMENT INHIBITORS** • Ibalizumab-uiyk **PHARMACOKINETIC (PK) ENHANCER** • Ritonavir (RTV) • Cobicistat (COBI)
ONE PILL REGIMEN	• **BIC/FTC/TAF** • **DOR/3TC/TDF** • **DRV/COBI/FTC/TAF** • **DTG/ABC/3TC; Only if HLA-B*57:01 negative** • **DTG/RPV** • **EFV/TDF/FTC** • **EVG/COBI/TDF/FTC** • **RPV/TDF/FTC (if HIV RNA <100,000 copies/mL and CD4 >200 cells/μL)** • **RPV/TAF/FTC (if HIV RNA <100,000 copies/mL and CD4 >200 cells/μL)**	

FIG. 41.11 Antiretroviral (ARV) medications approved by the US Food and Drug Administration (FDA). *INSTI*, Integrase strand transfer inhibitor; *NNRTI*, nonnucleoside reverse transcriptase inhibitor; *NRTI*, nucleoside reverse transcriptase inhibitor; *PI*, protease inhibitor.

viral DNA and thus block the incorporation of the completed HIV DNA copy into the host-cell DNA. A long-acting, injectable formulation combining an NNRTI and an integrase inhibitor is being studied.

Entry Inhibitors: Fusion Inhibitors, CCR5 Blockers, Attachment and Post-Attachment Inhibitors

Fusion inhibitors and CCR5 antagonists inhibit HIV entry into host cells. Fusion inhibitors bind to viral gp41 and block the conformational changes necessary to induce fusion of the viral particle with the host cell. CCR5 antagonists bind to the CCR5 chemokine co-receptor on host cells, inducing a conformational change that impedes CCR5 interaction with HIV gp120, thereby preventing HIV entry into host cells. Fostemsavir is a first-in-class HIV attachment inhibitor that works by attaching directly to HIV gp120 and, as a result, blocking HIV from attaching to CD4 on the host cell. Ibalizumab is a CD4-directed post-attachment inhibitor. Use of agents in this class of medications is typically limited to treatment of individuals with multidrug resistant infection.

IMMUNORECONSTITUTION AFTER THERAPY

Return of T Cells: Memory T Cells, Then Naïve T Cells

To varying degrees, the immune system is able to recover following initiation of therapy, a process called immunoreconstitution.[30] Upon start of ART in patients who are compliant and able to tolerate the regimen, the initial CD4 T-cell count is the best predictor of a successful outcome. Rapid reduction in the viral load, often to an undetectable level, is one of the earliest changes following initiation of ART, reflecting the ability of combination ART to rapidly suppress viral replication. Lagging behind the drop in viral load is the rise in CD4 T-cell concentration. An initial increase in circulating CD4 T cells occurs in 3 to 6 months as a result of a decrease in immune activation and subsequent migration of memory T cells ($CD4^+$, $CD45RO^+$) out of the lymphoid compartment. A more gradual rise in total CD4 T cells occurs over the course of 3 to 5 years with the appearance of new, naïve ($CD4^+$, $CD45RA^+$, $CD62L^+$) and memory T cells. Interestingly, a substantial minority of patients never reach a normal level of CD4 T cells, but, instead, reach a plateau at lower levels. Primary drug prophylaxis and some secondary drug prophylaxis for opportunistic infections may be discontinued in patients once the CD4 T-cell count reaches greater than 200 cells/μL and is maintained for more than 3 to 6 months. Cellular and humoral responses to most pathogens also recover with rising CD4 T-cell counts. Of interest, a low CD4 T-cell count at the time of initiating therapy predicts a poor response to bacterial vaccines even after recovery of CD4 T-cell levels, suggesting a lag in the return of naïve CD4 T cells.

Immune Reconstitution Inflammatory Syndrome

Immune reconstitution inflammatory syndrome (IRIS) is a well-known, if incompletely understood, response in patients with AIDS after initiating ART.[30] IRIS is characterized by an acute paradoxical worsening of inflammatory symptoms of treated opportunistic infections or the unmasking of previously subclinical, untreated infections related to the recovery of immune responses to opportunistic pathogens. IRIS occurs within weeks of ART initiation as the memory and effector antigen-activated

CD4 T-cell population recovers. A recent systematic review found that IRIS developed in 13% of patients after initiation of ART.[30] The most predictive risk factor for the development of IRIS was a low CD4 T-cell count at the start of ART, with the incidence of IRIS increasing exponentially as the CD4 T-cell count declined. IRIS develops more commonly in patients with cytomegalovirus (CMV) retinitis, cryptococcal meningitis, progressive multifocal leukoencephalopathy, and tuberculosis. Studies reported that as many as 4% of patients with IRIS died, but the proportion was much higher if the syndrome was associated with cryptococcal meningitis.[30] Fig. 41.12 shows the several models of inflammation that could result in IRIS as the immune system improves when AIDS is treated with ARV medications.[30]

Hyperallergenic State Associated With Immunoreconstitution

Another complication possibly associated with IRIS is the appearance of asthma in children who were perinatally infected with HIV and received combination ART since infancy.[31] This condition may be mediated by CD4 T-cell activation, release of Th2-type cytokines, and loss of regulatory T cells (Tregs) and tolerance. In support of this concept, Gingo et al. reported at least a 20% prevalence of asthma in adults with HIV infection compared with that of 8.8% in the general population.[32] In a subsequent pediatric study in which pulmonary function testing was objectively measured by spirometry, the following findings emerged: nonreversible obstructive pulmonary disease was present in youths who had been perinatally infected with HIV and possibly in those exposed to HIV but who were uninfected. This pulmonary disorder contains elements of asthma and chronic obstructive pulmonary disease (COPD) and closely resembles the asthma-COPD overlap syndrome. Chronic infections and immune dysregulation appear to play a significant role in this complication of HIV infection.[33,34]

PREVENTION

Prevention of Mother-to-Child Transmission

More than 90% of children living with HIV worldwide were infected through mother-to-child transmission during pregnancy, around the time of birth, or through breastfeeding.[1] Efforts to prevent this transmission hold the most promise in reducing the number of children infected with HIV, and these efforts include (i) early identification of HIV infection in pregnant women through routine antenatal testing; (ii) provision of ARV medications to both the pregnant woman and her infant; (iii) delivery by elective cesarean section, when indicated; (iv) complete avoidance of breastfeeding when safe and sustainable alternatives are available; (v) widespread availability of educational programs addressing HIV infection; and (vi) HIV counseling and testing services.[35]

Prevention of Sexual Transmission

Several biomedical interventions have the potential for radically changing the patterns and rates of HIV transmission. These include male circumcision and expanded use of ART in infected individuals to prevent ongoing infection (treatment as prevention) or prophylactically in uninfected individuals either before or after potential exposure to HIV to prevent acquisition of infection (PrEP and postexposure prophylaxis).

Male Medical Circumcision

The penile foreskin contains HIV-susceptible cells and is a potential portal of viral entry. Randomized controlled trials

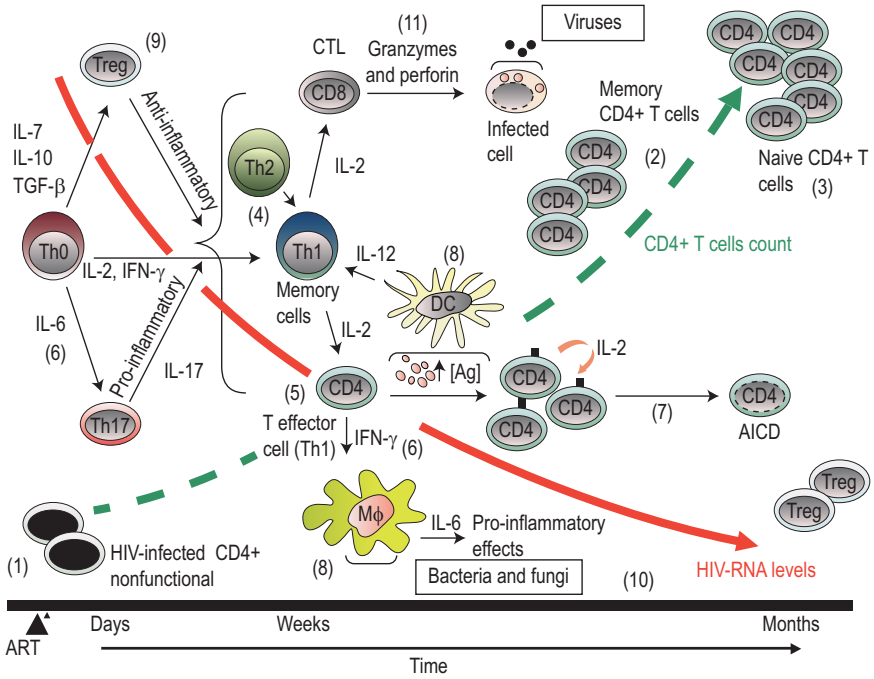


FIG. 41.12 Classical model of immunopathogenesis of immune reconstitution inflammatory syndrome. *ART*, Antiretroviral therapy. (From Manzardo C, Guardo AC, Letang E, et al. Opportunistic infections and immune reconstitution inflammatory syndrome in HIV-1-infected adults in the combined antiretroviral therapy era: a comprehensive review. *Expert Rev Anti Infect Ther*. 2015;13[6]:751–767.)

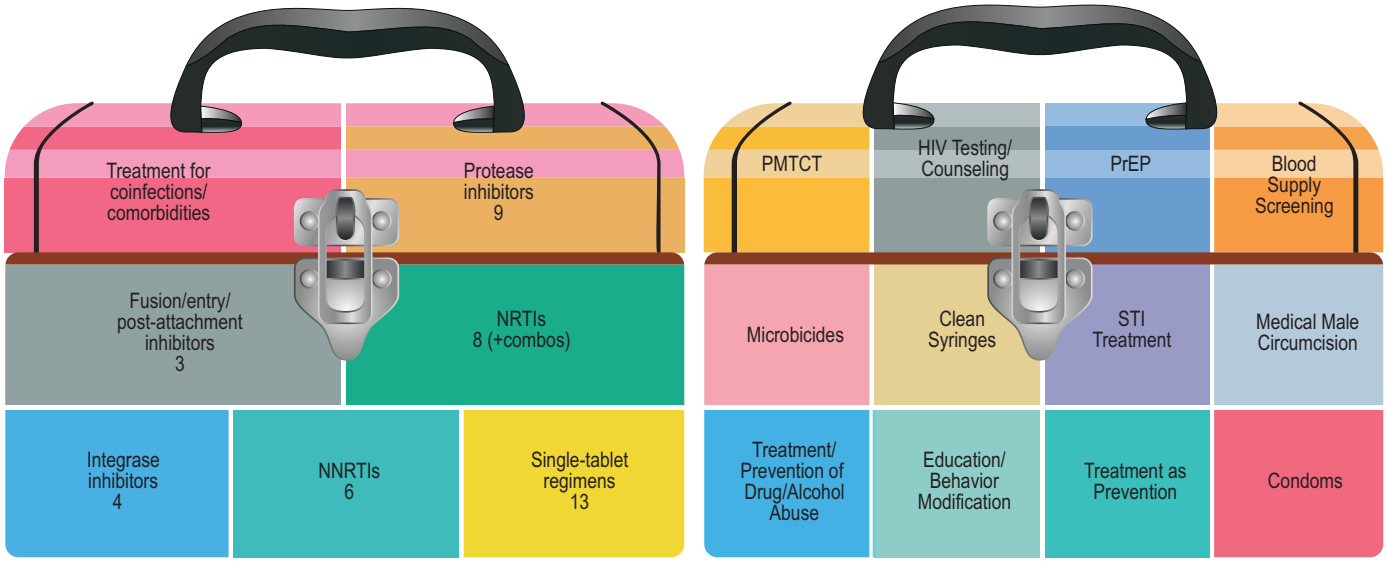


FIG. 41.13 Human Immunodeficiency Virus Treatment and Prevention Toolkits. *NNRTI*, Nonnucleoside reverse transcriptase inhibitor; *NRTI*, nucleoside reverse transcriptase inhibitor; *PMTCT*, prevent mother-to-child transmission; *PrEP*, pre-exposure prophylaxis. (From Eisinger RW, Folkers GK, Fauci AS. Ending the human immunodeficiency virus pandemic: optimizing the prevention and treatment toolkits. *Clin Infect Dis*. 201969[12]:2212–2217.)

in several African countries have indicated that male medical circumcision reduces the risk of heterosexually acquiring HIV infection by 50% to 60%.[36] The WHO recommends male circumcision as part of a comprehensive HIV prevention package; however, condom use and other prevention modalities also remain important in HIV prevention. There was only a modest benefit for the female partners of the circumcised men. In fact, in some circumstances, the risk of transmission to women from men with HIV infection who have undergone circumcision may be increased, perhaps through exposure to infected blood attributed to resumption of sexual activity before the circumcision site had fully healed.[36]

Pre-Exposure Prophylaxis

PrEP is the use of an ARV medication to prevent the acquisition of HIV infection by uninfected persons. PrEP can either be taken orally, using an ARV drug available for treatment of HIV infection (tenofovir plus emtricitabine), or topically as a vaginal gel containing tenofovir. The efficacy of oral PrEP has been shown in randomized controlled trials and is high when the drug is used as directed. Moderate efficacy of a PrEP gel has been shown in one trial. The European Medicines Agency recently approved the use of an ARV-containing vaginal ring for prevention for women in high HIV burden settings. People at substantial risk of HIV infection should be offered PrEP as an additional prevention choice, as part of comprehensive prevention. Other formulations of ART are being studied for use as PrEP. A long-acting injectable integrase inhibitor shows promise in preventing HIV.

Expanded Treatment With Antiretroviral Therapy

The expanded use of ART in individuals with HIV infection has been shown to reduce HIV transmission to uninfected partners. Both "Treatment as Prevention" and "Test and Treat" strategies involve the use of ART in individuals with the express purpose of treating HIV early and reducing transmission to others. In both approaches, ART is initiated regardless of CD4 count or viral load to reduce the viral load in the genital secretions of individuals who have HIV infection and thereby reduce HIV transmission to partners. The efficacy of the "Treatment as Prevention" model is demonstrated in the multinational HIV Prevention Trials Network (HPTN 052) clinical trial that examined the effectiveness of ART to prevent the sexual transmission of HIV in serodiscordant couples.[37] Serodiscordant couples (1763 in number) were randomly assigned to have the individual with HIV either start ART immediately upon enrollment or to defer ART until immunological or clinical criteria were met. Of 28 genetically linked infections that occurred during the trial, only 1 infection occurred in couples assigned to receive immediate treatment, representing a 96% reduction in the risk of HIV transmission. There were also fewer morbidity and mortality events in the early-treatment group, suggesting a therapeutic benefit from early treatment as well. Subsequent studies have found no linked transmissions in a total of more than 150,000 condomless sex acts when the partner with HIV was on ART and viral load was below detectable levels.[38] HIV treatment and prevention tools are illustrated in Fig. 41.13.

HIV VACCINES: CLINICAL TRIALS

Preventive Vaccines

The production of an effective HIV vaccine has been thwarted by the genetic variability of HIV, extreme rate of mutation in the virion, and the sequestration of the virus in impenetrable reservoirs, predominantly the nonreplicating CD4 T cell. Also, glycan shielding may occur. HIV shields its envelope proteins with a blanket of molecules called glycans, which resist penetration by antibodies. More than 30 HIV vaccines have been tested in human trials, including those with recombinant env gp120 proteins with adjuvants, HIV DNA plasmids, viral vectors,

and prime-boost designs.[39] Most vaccines work by inducing the B cells to make antibodies against the infection that the vaccine mimics. Antibodies attack pathogens and either destroy them directly, or "tag" infected cells so they can be destroyed by other parts of the immune system. Broadly neutralizing antibodies (bNAbs) overcome the genetic variability and glycan shield in that they are active against a wide number of different viral strains and react to and attach to the most conserved parts of HIV by penetrating the glycan shield.

So far, however, the results for HIV vaccines have been disappointing. Some HIV vaccines can induce anti-HIV antibody responses, but they have proved to be either ineffective (e.g., the HVTN 702 study) or only marginally effective (e.g., the RV 144 vaccine study). The phase III trial RV 144 in Thailand (ALVAC-HIV vCP1521 + AIDSVAX gp120 B/E) showed possible protection against HIV infection in heterosexual men and women.[39]

In addition to the goal of developing an HIV vaccine that elicits neutralizing antibodies, the search for a vaccine that stimulates a protective CD8 cytotoxic T-cell response continues. Association of certain major histocompatibility complex (MHC) molecules with HIV disease progression is clearly linked to the cytotoxic T-cell responses. Unfortunately, the HIV vaccine-induced cytotoxic CD8 T-cell response is insufficient to halt the progression of acute or chronic HIV disease. This was clearly indicated in the STEP clinical trial, in which the CD8 T-cell effects were the same between HIV-infected vaccinees and sham HIV-infected vaccinee controls. Qualities of effector and central memory CD8 T cells that would be protective include (i) production of cytotoxic cytokines (e.g., interferon gamma (IFN)-γ and IL-2); (ii) rapidly replicating capacity; (iii) cytotoxic potential; (iv) high affinity for HIV antigens; (v) inhibition of HIV replication; (vi) recognition of specific HIV epitopes restricted by protective HLA-B antigens; (vii) central memory cells with long life spans; and (viii) rapid-attack memory cells at mucosal HIV entry sites. A collaboration of scientists has proposed the modification of the partially successful RV 144 vaccine with the goal of producing a new HIV vaccine that will broadly neutralize HIV and variants of HIV that emerge under selective pressure.[40] These modifications include changes in the viral epitopes, vaccine adjuvants, and use of a different clade as the construct of the virus. The goal of this new proposal is to use the information of many previous HIV trials to produce an ideal HIV vaccine that will prevent the spread of HIV infection in children and adults.

Several large-scale trials of vaccines in late-phase development are ongoing in southern Africa: HIV Vaccine Trials Network (HVTN) 702 (Uhambo; NCT02968849) is a phase 2b/3 trial of safety and efficacy of the prime-boost ALVAC HIV vaccine plus bivalent gp120 protein adjuvanted with MF59 regimen in 5400 adult men and women in South Africa; HVTN 705 (Janssen; NCT03060629) is a phase 2b/3 trial of safety and efficacy of the prime-boost Ad26-mosaic vaccine plus gp140 protein vaccine in 2600 adult women in southern Africa and a second study of this strategy in MSM (HVTN 706; NCT03964415); and PrEPVacc (NCT04066881) is a phase 2b trial to assess the combination of an HIV vaccine (DNA, modified Vaccinia Ankara Virus, and Env protein plus adjuvant) and PrEP using an adaptive trial design.[41] A small phase I safety trial of a gene-transfer protocol using a DNA vector vaccine coding for the production of the bNAb VRCO7, a bNAb targeting the CD4 binding site of the HIV-1 envelope glycoprotein, showed promising results.[42]

Therapeutic Vaccines

A therapeutic vaccine is one in which the vaccine is used after infection occurs, aiming to induce antiviral immunity to alter the course of disease. This would be accomplished by controlling viremia or reducing the viral set point in infected patients. Primate models suggest that just such a result is possible, especially with cellular immunity–inducing vaccines. To date, however, data from human studies have not shown any conclusive benefit in using therapeutic vaccines alone. Using a therapeutic vaccine in combination with ART is another approach currently under investigation. A small study of a therapeutic vaccine randomized 15 subjects to receive vaccine and 16 to receive placebo. The intervention consisted of priming with a plasmid DNA (pDNA) vaccine containing genes encoding multiple HIV proteins, followed by boosting with an attenuated viral vector expressing a single HIV gene.[43] All participants had viral rebound and immune responses to HIV were only marginally enhanced.

Future for HIV Vaccines

There is general belief that more basic research exploring vaccine design and trials in animals will lead to important clues for human study, but newer approaches have been employed as well (Fig. 41.14). Illustration of clinical trials in ARV-treated macaques may be applied to humans in future by Byrareddy et al.[44] and Nishimura et al.[45] Monoclonal antibody specific for CD4 T-cell-surface integrin ($\alpha_4\beta_7$) disrupts cellular trafficking of CD4 T cells with gastrointestinal tissue mucosal vascular addressing cell adhesion molecule (MAdCAM1). The CD4 T-cell counts remained steady, CD8 T-cell immunity sharply increased, and HIV replication became undetectable for up to 2 years.[44,45] As illustrated by these HIV vaccine studies in animals, understanding the immune correlates of vaccine efficacy is the usual approach for judging the success of an HIV vaccine. However, broadly neutralizing antibodies (bNAbs) are being explored in another way. Investigators are hypothesizing that bNAb formation will correlate with immunity and are attempting to design a vaccine to induce the correlate. Large passive transfer studies of bNAbs for prevention will inform the feasibility of this approach. The challenge is to take an epitope on the viral envelope that might induce a broadly neutralizing antibody, clone the antibody, and show binding to the epitope, then make an immunogen that induces the bNAb.

ON THE HORIZON

- Production of newer antiretroviral drugs, including those used for post-exposure prophylaxis with greater specificity for interrupting events in the viral life cycle and with fewer side effects for patients.
- Investigation of new microbicidal drugs that can be safely applied before exposure for protection against human immunodeficiency virus (HIV) transfer.
- Development of preventive and therapeutic vaccines for HIV/acquired immunodeficiency syndrome (AIDS) that induce strong viral neutralizing antibody power and strong CD8 T-cell cytotoxic responses.
- Testing of gene construct–modified autologous hematopoietic stem cells capable of halting HIV replication.

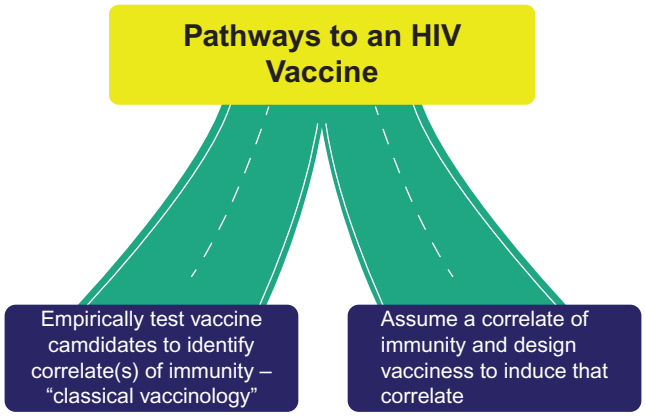


FIG. 41.14 Pathways to an Human Immunodeficiency Virus (HIV) Vaccine; Classical Vaccinology Versus a Newer Approach Being Used to Study Vaccines Designed to Induce Broadly Neutralizing Antibodies. The newer approach assumes a correlate of immunity and involves designing vaccines to induce that correlate. (From Eisinger, RW, Folkers GK, Fauci AS. Ending the human immunodeficiency virus pandemic: optimizing the prevention and treatment toolkits. *Clin Infect Dis*. 2019;69[12]:2212–2217.)

TRANSLATIONAL RESEARCH NEEDS AND CONCLUSIONS

Current ARV therapeutic agents for treatment of HIV are usually administered in simplified regimens of one pill per day, and these regimens are potent with minimal toxicity. When taking these ARVs regularly, people living with HIV can expect a near-normal life expectancy. However, the eradication of HIV/AIDS can be approached practically only with an effective preventive vaccine. The approach to cure with gene therapy is perhaps the most sophisticated translational research venture. Using viral vectors to place gene constructs within nuclear DNA to prevent HIV replication is the goal of such research. Arguably, the most advanced form of this genetic engineering to halt HIV replication is the zinc finger endonuclease approach to disrupt specific genes necessary for the life cycle of HIV. Adoptive transfer of autologous zinc finger–treated stem cells with infinite replication capacity may be an attractive future for individuals already infected with HIV. The extraordinary experiment of HLA-matched and CCR5-δ35 deletion of hematopoietic stem cell immunoreconstitution of a patient with HIV infection ("Berlin Man") is a proof of concept of molecular and genetic engineering to cure HIV infection, but is a technique totally impractical for the millions of patients with HIV infection worldwide. Nevertheless, this "one in a million" chance experiment has demonstrated the survival advantage of lymphocytes that cannot become infected with HIV.

HIV, a type 1 retrovirus, contains merely nine genes, but those nine genes have so far thwarted all scientific efforts toward finding a cure of its infection in humans. Optimism is warranted, however, because of the enormous knowledge base the study of HIV has generated in understanding the many arms of innate and adaptive immunity protecting humans and the promise of a curative treatment or vaccine for HIV. Perhaps no other disease has caused so much to be learned so fast. In much of the world, HIV causes chronic infection, rather than certain death, thanks, in large part, to the use of ARV drugs. More novel drugs are in development as a result of the new-found understanding of the molecular biology of HIV.

The HIV/AIDS pandemic has also brought the sobering realization that other new and potentially deadly pathogens could yet emerge to strike at humanity. Severe acute respiratory syndrome (SARS) coronavirus-2 (CoV-2), which causes the novel coronavirus disease that emerged in 2019 (COVID-19), has caused over 20 million infections worldwide and resulted in hundreds of thousands of deaths. The SARS CoV-2 pandemic has wreaked havoc on economies, interrupted supply chains of goods and services, and brought health services in the United States to the brink of failure. The novel pathogen will bring untold morbidity among survivors of severe disease. Regarding HIV prevention and treatment, the new pandemic has disrupted HIV treatment and prevention programs worldwide and has brought to the forefront health and healthcare disparities and inequalities reminiscent of the early days of the HIV pandemic.

ACKNOWLEDGMENTS

Supported in part by NIH Grants HD052102, AI069536 AI36211, and AI082978, and the Pediatric AIDS Fund of Texas Children's Hospital.

REFERENCES

1. Joint United Nations Programme on HIV/AIDS (2019). *UNAIDS Data*. Geneva, Switzerland; 2019. Available from https://www.unaids.org/en/resources/documents/2019/2019-UNAIDS-data
2. Centers for Disease Control and Prevention. *HIV Surveillance Report, 2018 (Updated 2020)*.Available from https://www.cdc.gov/hiv/statistics/overview/index.html
3. Shacklett BL. Mucosal immunity in HIV/SIV infection: T cells, B cells and beyond. *Curr Immunol Rev*. 2019;15(1):63–75.
4. Maartens G, Celum C, Lewin SR. HIV infection: epidemiology, pathogenesis, treatment, and prevention. *Lancet*. 2014;384(9939):258–271.
5. Kanwar B, Favre D, McCune JM. Th17 and regulatory T cells: implications for AIDS pathogenesis. *Curr Opin HIV AIDS*. 2010;5(2):151–157.
6. Mudd JC, Brenchley JM. ILC you later: early and irreparable loss of innate lymphocytes in HIV infection. *Immunity*. 2016;44(2):216–218.
7. Deeks SG, Lewin SR, Havlir DV. The end of AIDS: HIV infection as a chronic disease. *Lancet*. 2013;382(9903):1525–1233.
8. Chevalier MF, Weiss L. The split personality of regulatory T cells in HIV infection. *Blood*. 2013;121(1):29–37.
9. Katlama C, Deeks SG, Autran B, et al. Barriers to a cure for HIV: new ways to target and eradicate HIV-1 reservoirs. *Lancet*. 2013;381(9883):2109–2117.
10. Siliciano JD, Siliciano RF. Recent developments in the search for a cure for HIV-1 infection: targeting the latent reservoir for HIV-1. *J Allergy Clin Immunol*. 2014;134(1):12–19.
11. Cary DC, Fujinaga K, Peterlin BM. Molecular mechanisms of HIV latency. *J Clin Invest*. 2016;126(2):448–454.
12. Maldarelli F. The role of HIV integration in viral persistence: no more whistling past the proviral graveyard. *J Clin Invest*. 2016;126(2):438–447.
13. Murray AJ, Kwon KJ, Farber DL, et al. The latent reservoir for HIV-1: how immunologic memory and clonal expansion contribute to HIV-1 persistence. *J Immunol*. 2016;197(2):407–417.
14. Jones RB, Walker BD. HIV-specific CD8(+) T cells and HIV eradication. *J Clin Invest*. 2016;126(2):455–463.
15. Barin F, Braibant M. HIV-1 antibodies in prevention of transmission. *Curr Opin HIV AIDS*. 2019;14(4):273–278.
16. Richardson S, Moore PL. The antibody response in HIV-infected donors. *Curr Opin HIV AIDS*. 2019;14(4):233–239.

17. Altfeld M, Gale Jr. M. Innate immunity against HIV-1 infection. *Nat Immunol*. 2015;16(6):554–562.
18. Bergantz L, Subra F, Deprez E, et al. Interplay between intrinsic and innate immunity during HIV infection. *Cells*. 2019;8(8):922.
19. Centers for Disease Control and Prevention. *Preexposure Prophylaxis for the Prevention of HIV Infection in the United States—2017 Update: A Clinical Practice Guideline. 2018.* Available from https://www.cdc.gov/hiv/pdf/risk/prep/cdc-hiv-prep-guidelines-2017.pdf. Accessed June 30, 2020.
20. Hersperger AR, Martin JN, Shin LY, et al. Increased HIV-specific $CD8^+$ T-cell cytotoxic potential in HIV elite controllers is associated with T-bet expression. *Blood*. 2011;117(14):3799–3808.
21. Centers for Disease Control and Prevention, Association of Public Health Laboratories. *Laboratory Testing for the Diagnosis of HIV Infection: Updated Recommendations 2014*. Available from https://doi.org/10.15620/cdc.23447. Accessed June 30, 2020.
22. Centers for Disease Control and Prevention, Association of Public Health Laboratories. *Laboratory Testing for the Diagnosis of HIV Infection: Updated Recommendations.* Atlanta, GA; 2018. Available from https://stacks.cdc.gov/view/cdc/50872
23. Joint United Nations Programme on HIV/AIDS (UNAIDS). *Knowledge Is Power: Know Your Status, Know Your Viral Load.* 2018. Available from https://www.unaids.org/sites/default/files/media_asset/jc2940_knowledge-is-power-report_en.pdf. Accessed June 30, 2020.
24. Panel on Antiretroviral Guidelines for Adults and Adolescents. *Guidelines for the Use of Antiretroviral Agents in HIV-1-infected Adults and Adolescents*. Bethesda, MD: NIH Office of the AIDS Research Advisory Council; 2020. Available from https://aidsinfo.nih.gov/contentfiles/lvguidelines/adultandadolescentgl.pdf. Accessed June 30, 2020.
25. Panel on Antiretroviral Therapy and Medical Management of HIV-Infected Children. *Guidelines for the Use of Antiretroviral Agents in Pediatric HIV Infection.* Available from https://aidsinfo.nih.gov/contentfiles/lvguidelines/pediatricguidelines.pdf. Accessed June 30, 2020.
26. Achhra AC, Mocroft A, Ross M, et al. Impact of early versus deferred antiretroviral therapy on estimated glomerular filtration rate in HIV-positive individuals in the START trial. *Int J Antimicrob Agents*. 2017;50(3):453–460.
27. Dharan NJ, Neuhaus J, Rockstroh JK, et al. Benefit of early versus deferred antiretroviral therapy on progression of liver fibrosis among people with HIV in the START randomized trial. *Hepatology*. 2019;69(3):1135–1150.
28. Ghislain M, Bastard JP, Meyer L, et al. Late antiretroviral therapy (ART) initiation is associated with long-term persistence of systemic inflammation and metabolic abnormalities. *PLoS One*. 2015;10(12):e0144317.
29. Lundgren JD, Borges AH, Neaton JD. Serious non-AIDS conditions in HIV: benefit of early ART. *Curr HIV/AIDS Rep*. 2018;15(2):162–171.
30. Manzardo C, Guardo AC, Letang E, et al. Opportunistic infections and immune reconstitution inflammatory syndrome in HIV-1-infected adults in the combined antiretroviral therapy era: a comprehensive review. *Expert Rev Anti Infect Ther*. 2015;13(6):751–767.
31. Siberry GK, Leister E, Jacobson DL, et al. Increased risk of asthma and atopic dermatitis in perinatally HIV-infected children and adolescents. *Clin Immunol*. 2012;142(2):201–208.
32. Gingo MR, Wenzel SE, Steele C, et al. Asthma diagnosis and airway bronchodilator response in HIV-infected patients. *J Allergy Clin Immunol*. 2012;129(3):708–714. e8.
33. Shearer WT, Jacobson DL, Yu W, et al. Long-term pulmonary complications in perinatally HIV-infected youth. *J Allergy Clin Immunol*. 2017;140(4):1101–1111. e7.
34. Attia EF, Jacobson D, Yu W, et al. Immune imbalance and activation are associated with lower lung function in youth with perinatally acquired HIV. *J Allergy Clin Immunol*. 2020;145(5):1473–1476.
35. Panel on Treatment of HIV-Infected Pregnant Women and Prevention of Perinatal Transmission. *Recommendations for Use of Antiretroviral Drugs in Pregnant HIV-1-Infected Women for Maternal Health and Interventions to Reduce Perinatal HIV Transmission in the United States*. 2020. Available from https://aidsinfo.nih.gov/contentfiles/lvguidelines/perinatalgl.pdf. Accessed July 9, 2020.
36. Davis SM, Hines JZ, Habel M, et al. Progress in voluntary medical male circumcision for HIV prevention supported by the US President's Emergency Plan for AIDS Relief through 2017: longitudinal and recent cross-sectional programme data. *BMJ Open*. 2018;8(8):e021835.
37. Cohen MS, Chen YQ, McCauley M, et al. for the HPTN 052 Study Team. Antiretroviral therapy for the prevention of HIV-1 transmission. *N Engl J Med*. 2016 Sept 1;375:830–839.
38. Eisinger RW, Folkers GK, Fauci AS. Ending the human immunodeficiency virus pandemic: optimizing the prevention and treatment toolkits. *Clin Infec Dis*. 2019;69(12):2212–2217.
39. Huang Y, Follmann D, Nason M, et al. Effect of rAd5-vector HIV-1 preventive vaccines on HIV-1 acquisition: a participant-level meta-analysis of randomized trials. *PLoS One*. 2015;10(9):e0136626.
40. Corey L, Gilbert PB, Tomaras GD, et al. Immune correlates of vaccine protection against HIV-1 acquisition. *Sci Transl Med*. 2015;7(310):310rv7.
41. Bekker LG, Tatoud R, Dabis F, et al. The complex challenges of HIV vaccine development require renewed and expanded global commitment. *Lancet*. 2020;395(10221):384–388.
42. Casazza JP, Narpala S, Novik L, et al. Durable HIV-1 antibody production in humans after AAV8-mediated gene transfer. Abstract #41. Conference on Retroviruses and Opportunistic Infections (CROI) Conference. Boston, MA, March 8–11, 2020. Available from https://www.croiconference.org/abstract/durable-hiv-1-antibody-production-in-humans-after-aav8-mediated-gene-transfer/
43. Sneller MC, Justement JS, Gittens KR, et al. A randomized controlled safety/efficacy trial of therapeutic vaccination in HIV-infected individuals who initiated antiretroviral therapy early in infection. *Sci Transl Med*. 2017;9(419):eaan8848.
44. Byrareddy SN, Arthos J, Cicala C, et al. Sustained virologic control in SIV+ macaques after antiretroviral and alpha4beta7 antibody therapy. *Science*. 2016;354(6309):197–202.
45. Nishimura Y, Gautam R, Chun TW, et al. Early antibody therapy can induce long-lasting immunity to SHIV. *Nature*. 2017;543(7646):559–563.

42

Autoantibody-Mediated Phenocopies of Primary Immunodeficiency Diseases

Jennifer D. Treat, Sarah K. Browne, and Christa S. Zerbe

Anticytokine autoantibodies have been identified in health and disease, with their role in pathogenesis ranging from none to directly causal. It is becoming increasingly recognized that neutralizing, high-titer autoantibodies cause a variety of potentially life-threatening illnesses. Their manifestations are diverse, and their clinical presentation is generally a reflection of the cytokine pathways that are rendered functionally deficient.

Examples include pulmonary alveolar proteinosis (PAP) caused by anti-granulocyte macrophage-colony-stimulating factor (GM-CSF) autoantibodies;[1] disseminated nontuberculous mycobacterial (NTM) and other opportunistic infections caused by anti-interferon-γ (IFN-γ) autoantibodies;[2] chronic mucocutaneous candidiasis (CMC) caused by anti–interleukin (IL)-17A, anti-IL-17F, or anti-IL-22 autoantibodies;[3,4] and bacterial infection caused by anti-IL-6 autoantibodies[5] (Table 42.1). In each of these cases a high-titer, biologically active autoantibody has been identified in association with a unique clinical syndrome. Although this list is not exhaustive, these diseases are of particular interest in that the clinical manifestations are often similar to those in patients with genetic defects in the same cytokine-associated pathways. The Mendelian defects that confer a similar phenotype to their anti-cytokine autoantibody counterpart also provide a strong biological rationale for establishing the autoantibodies as causal. Although a number of other anticytokine-autoantibody associated syndromes such as pure red-cell aplasia (antierythropoietin autoantibodies)[6] and severe osteoporosis in celiac disease (antiosteoprotegerin autoantibodies)[7] have been described, this chapter focuses on those that increase susceptibility to infection.

There are numerous commonalities and distinctions worth highlighting among anticytokine autoantibody-associated immune deficiency. Shared in common is the presence of high-titer, neutralizing anticytokine autoantibodies of the immunoglobulin G (IgG) isotype. PAP caused by anti-GM-CSF autoantibodies appears to be primarily a lung disease largely resulting from disruption of GM-CSF-dependent pulmonary surfactant catabolism. *In vitro* studies using primary human cells and mouse work have demonstrated that anti-GM-CSF autoantibodies cause immune dysfunction largely through transcription factor PU.1, which may explain pulmonary and extrapulmonary infections observed in PAP, particularly those opportunists known to be controlled by neutrophils or macrophages.[1] PAP caused by anti-GM-CSF autoantibodies is similar to severe immune deficiency caused by anti-IFN-γ autoantibodies in that both diseases have adult onset. On the basis of broad screening for other autoantibodies, it appears affected patients make high-titer neutralizing autoantibody against only one cytokine and do not demonstrate an increased frequency of other autoantibodies or forms of autoimmunity.[1,2] Patients with autoantibodies against IL-17A, IL-17F, IL-22 who also have chronic mucocutaneous candidiasis (CMC) are different in that, so far, all have had an underlying primary diagnosis of either autoimmune polyendocrinopathy with candidiasis and ectodermal dysplasia (APECED, also known as autoimmune polyendocrinopathy syndrome type 1 [APS-1]) or thymic neoplasia.[3,4] Thus in patients with APECED, a mendelian defect in the gene *AIRE*, which is responsible for negative selection for autoreactive T cells in the thymus, the onset of anti-cytokine autoantibodies and CMC occurs earlier in life. Occasionally, CMC is the initial presentation of thymoma (just as myasthenia gravis caused by anti-acetylcholine receptor autoantibodies must prompt evaluation for thymoma), highlighting the fact that for all anticytokine autoantibody-associated syndromes, the timeline for development of anticytokine autoantibodies relative to the observation of clinical disease is largely unknown. Additionally, patients with thymoma and APECED can develop multiple anti-cytokine autoantibodies, although they have a proclivity for certain ones, including type I IFNs, IL-17A, IL-17F, IL-22, and, in the case of thymoma (but not APECED), IL-12. Given the presence of multiple neutralizing autoantibodies, and the fact that T-cell intrinsic defects or thymic insufficiency can also confer infection susceptibility, it is challenging to prove that a particular anticytokine autoantibody is the necessary and sufficient agent of disease pathogenesis. In the case of anti-IL-6, four patients have

KEY CONCEPTS

Anticytokine Autoantibody-Associated Primary Immunodeficiency Diseases

- Anticytokine autoantibodies have been identified in a group of patients with immunodeficiency diseases characterized by the development of high-titer, neutralizing autoantibodies to cytokines.
- The clinical manifestations in patients with anticytokine autoantibodies are similar to those seen in patients with genetic defects in the pathway of the target cytokine.
- Immunodeficiency syndromes described include those associated with anti-granulocyte macrophage-colony-stimulating factor (GM-CSF) autoantibodies and pulmonary alveolar proteinosis (PAP); anti-interferon (IFN)-γ autoantibodies and severe immunodeficiency; anti-interleukin (IL)-17A, anti-IL-17F, and anti-IL-22 autoantibodies and chronic mucocutaneous candidiasis (CMC); and anti-IL-6 autoantibodies and bacterial infection.
- Management of these syndromes involves treating the consequences of the autoantibody (*e.g.*, infectious manifestations, lung disease) or targeting the autoantibody itself.

TABLE 42.1 Anticytokine Autoantibody–Associated Syndromes and Their Phenotypically Similar Genetic Counterparts

Cytokine Target	Genetic Phenocopies	Clinical Manifestations	Laboratory and Radiological Manifestations	*In vitro* Evidence for Biological Activity	Infections	Comments
GM-CSF	GM-CSF receptor α or β subunits	PAP Insidious and progressive respiratory failure Newly described cases of isolated cryptococcal meningitis and *Nocardia* infection in individuals without human immunodeficiency virus (HIV) infection[17–19]	BAL fluid contains large foamy macrophages or monocyte-like macrophages and elevated levels of surfactant proteins Characteristic computed tomography (CT) of chest demonstrates ground-glass opacities with thickening of intralobular septae, "crazy paving." PFTs demonstrate restrictive and diffusion defects[1]	Anti-GM-cerebrospinal fluid autoantibodies inhibit pSTAT-5 production,[17] PU.1 expression, Macrophage inflammatory protein 1α (MIP-1α) production	Pulmonary and extrapulmonary infections with *Nocardia, Aspergillus, Proteus, Histoplasmosis*, and *Cryptococcosis*[1]	Mutations in surfactant proteins have also been described SP-B, Surfactant protein C (SP-C), or ATP-binding cassette subfamily A-3 (ABCA3), but are considered clinically distinct
IFN-γ[h]	Mutations in IFN-γR1; IFN-γR2; STAT1; IL-12Rβ1; IL-12Rβ2; IL-12p40; NEMO; IRF8; ISG15	Chronic infections with intracellular pathogens, particularly lymphadenitis, skin, soft tissue, and bone infections; can be multiple organisms simultaneously or sequentially; reactive dermatoses; constitutional symptoms common	Elevated erythrocyte sedimentation rate, C-reactive protein-reactive protein (CRP), β_2 microglobulin, anemia, hypergammaglobulinemia. CT imaging may demonstrate abscess formation or osteomyelitis	Anti-IFN-γ autoantibodies inhibit phospho-STAT1 production(2, 11), IFN-γ inducible gene expression[29] and IL-12p70 and TNF-α protein production[2,11]	Nontuberculous mycobacteria, tuberculosis, nontyphoidal *Salmonella, Histoplasma, Penicillium, Cryptococcus, Burkholderia pseudomallei*, varicella-zoster virus[2]	Functional testing to evaluate downstream inhibitory effects of autoantibody important, (*i.e.*, plasma inhibition of IFN-γ–induced phospho-STAT1 production)
IL-17A IL-17F IL-22	IL-17RA IL-17F	Recurrent candidal infections of mucosal surfaces, nails, and skin Infection may become resistant to antifungals	(APECED patients only) In patients with pneumonitis, common CT findings include bronchiectasis, ground glass opacities or mosaicism.	Anti–IL-17 autoantibodies inhibit IL-17–induced IL-6[4]	*Candida* spp.	Reported only in the context of APECED and thymoma so far
IL-6	STAT3	Recurrent bacterial infections.	Undetectable CRP	Anti–IL-6 autoantibodies prevent IL-6–induced CRP mRNA[5]	*Staphylococcus aureus, Escherichia coli* and *Streptococcus intermedius*	Four cases reported to date

GM-CSF, Granulocyte macrophage–colony stimulating factor; *PAP,* pulmonary alveolar proteinosis; *BAL,* bronchoalveolar lavage; *PFT,* pulmonary function test; *pSTAT,* phosphorylated signal transducer and activator of transcription test; *MIP,* macrophage inflammatory protein; *SP,* surfactant protein; *ABC,* adenosine triphosphate-binding cassette; *IFN,* interferon; *NEMO,* nuclear factor (NF)-κB essential modulator; *IRF,* interferon regulatory factor; *TNF,* tumor necrosis factor; *APECED,* autoimmune polyendocrinopathy with candidiasis and ectodermal dysplasia.

been identified with neutralizing anti-IL-6 IgG antibodies. All four presented with severe bacterial infections but had low C-reactive protein (CRP), an IL-6-driven inflammatory marker.[5,8,9]

OVERVIEW OF PATHOPHYSIOLOGY

The pathophysiology of infection susceptibility is generally thought to involve a functional deficiency in the cytokine that is being neutralized (Chapter 14). It is believed that a high-titer autoantibody binds its respective cytokine target, thereby blocking downstream signaling and biological activity. For each anticytokine–autoantibody pair, it has been demonstrated that plasma or purified IgG from a patient with the anticytokine autoantibody prevents the activity of the targeted cytokine at the levels of signal transduction, gene transcription, and/or protein expression. In the case of anti–IFN-γ autoantibodies, it has been demonstrated that antibody levels track with disease activity;[10,11] however, for anti-GM-CSF autoantibodies, the results have been conflicting. It is also possible, but not yet proven, that antibody-binding avidity may influence the degree of disease severity as well. Thus, it may be possible to have high-titer, lower avidity anti-cytokine autoantibody leading to a similar disease phenotype to low-titer, high-avidity anti-cytokine autoantibody.

The events that lead to the generation of anti-cytokine autoantibodies are poorly understood and are likely disease specific. Nonetheless, by comparing and contrasting these diseases, we may begin to understand some key factors. Although a large cohort of patients with PAP have been described in Japan, this disease is seen worldwide across all ethnicities and not within families, suggesting that if there is a genetic component, it is a complex one. No familial clustering has been identified in over 130 reported cases of anti–IFN-γ autoantibodies and opportunistic infection;[2,12–15] however, the disease is mostly seen in Asian-born Asians, suggesting that there may be an environmental trigger in the context of a common genetic background.

The fact that anti-cytokine autoantibodies are both IgG and high-affinity implicates the T-helper (Th) lymphocyte-dependent processes of class switching and affinity maturation. Interestingly, anti-IL-17A, -IL-17F, and -IL-22 autoantibodies appear directly linked to either the genetic *AIRE* deficiency of APECED or the acquired *AIRE* deficiency observed in patients with thymoma.[16] In both cases, thymic-driven disease appears to be leading to extensive B-lymphocyte dysregulation in the form of many autoantibodies beyond just anti-cytokine autoantibodies. However, given that B cells may play a primary role in the development of autoimmunity in *AIRE* deficiency, the mechanisms underlying B-cell autoreactivity are likely complex.[16] Furthermore, evidence in mouse models of rheumatologic disease suggests that peripheral B-lymphocyte lineages leading to autoantibodies may fundamentally differ from those leading to development of protective antibodies.[17] Thus a common phenomenon of anticytokine autoantibody production may, in fact, be a reflection of a convergence of multiple differing mechanisms.

ANTI-GM-CSF AUTOANTIBODIES AND PULMONARY ALVEOLAR PROTEINOSIS

GM-CSF is a hematopoietic stem cell (HSC) growth factor that binds the GM-CSF receptor, which is widely expressed on many cell lineages, including neutrophils, macrophage precursors, dendritic cells (DCs), erythrocyte progenitors, and megakaryocytes. The GM-CSF receptor is composed of two α and two β subunits, which together bind two GM-CSF molecules with high affinity and induce signal transduction and activator of transcription (STAT)5 phosphorylation, nuclear translocation, and induction of expression of the master transcription factor PU.1. PAP was first described in 1958 by Rosen et al. as an idiopathic syndrome of respiratory failure, histopathologically characterized by alveoli filled with acellular periodic acid-Schiff-positive proteinaceous material.[1] The pathogenesis of PAP has since been linked to congenital or acquired defects in the GM-CSF signaling pathway.

The first clues to the etiological mechanism of PAP surfaced in 1994 and 1995 when GM-CSF$^{-/-}$ and GM-CSF receptor β$^{-/-}$ mice, respectively, demonstrated pulmonary disease that was virtually identical to human PAP. Shortly thereafter, mechanisms involving disruption of GM-CSF signaling were linked to PAP in humans.[1] Primary PAP results from mutations in either the GM-CSF receptor subunits α or β and generally leads to severe respiratory failure and usually presents early in life. Autoimmune PAP results from neutralizing anti-GM-CSF autoantibodies, can also cause respiratory failure, and shares the same pulmonary histopathology as the primary form (Fig. 42.1). In contrast to primary PAP, the autoimmune form is typically diagnosed in adulthood,[1] and its clinical course and severity are highly variable, ranging from progressive respiratory decline to spontaneous resolution. A secondary form of PAP caused by qualitative or quantitative deficiency of alveolar macrophages, generally in the context of hematologic malignancies, iatrogenic immunosuppression, or inhaled toxins, has also been recognized.

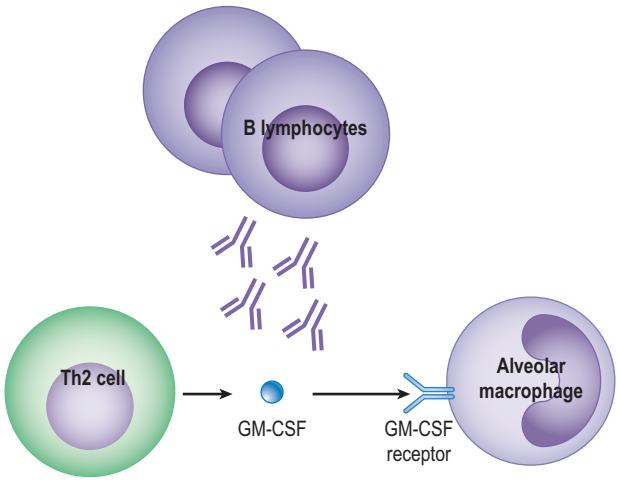


FIG. 42.1 Anti-Granulocyte Macrophage-Colony-Stimulating Factor (GM-CSF) Autoantibody Associated Pulmonary Alveolar Proteinosis (PAP). Autoimmune PAP results from impairment of GM-CSF dependent catabolism of surfactant and its subsequent overaccumulation in pulmonary alveoli due to neutralizing anti-GM-CSF autoantibodies, leading to respiratory failure. Th2 cell, type 2 T-helper cell.

Although the primary pathological process relates to impairment of GM-CSF-dependent catabolism of surfactant and its subsequent overaccumulation in pulmonary alveoli, there has long been an association with opportunistic infections.[1] Beyond its role in pulmonary surfactant homeostasis, the GM-CSF receptor is widely expressed on immune cells, including neutrophils, monocytes, DCs, megakaryocytes, and erythrocyte progenitor cells, and influences cell differentiation, proliferation, and immune activation. GM-CSF has been shown in both humans and mice to facilitate not only terminal differentiation of monocytes to alveolar macrophages but also innate immune responses, primarily via induction of transcription factor PU.1. Cells isolated from bronchoalveolar lavage (BAL) fluid from patients with autoimmune PAP show decreased PU.1 messenger RNA (mRNA), which is thought to be central to the pathogenesis of PAP. PU.1 has been shown to regulate Toll-like receptor (TLR) signaling, adhesion, phagocytosis, and microbicidal activity, thus providing a mechanism for the increased infection susceptibility that is observed in PAP. Beyond the macrophage, defects have also been shown in neutrophil adhesion, phagocytosis, oxidative burst, and bacterial killing from both the blood of human patients with PAP and GM-CSF$^{-/-}$ mouse bone marrow.[18]

Patients with acquired PAP are vulnerable to typical respiratory pathogens as well as opportunistic infections. Although the high incidence of respiratory infections may be partially attributable to their underlying chronic lung disease, the opportunistic infections are generally caused by organisms controlled by macrophages, including *Nocardia, Histoplasma*,[1] nontuberculous mycobacteria,[19] and *Cryptococcus*.[20] Although Witty et al. reported eight cases of PAP and *Mycobacterium avium* complex cultured from BAL fluid,[19] these patients did not receive antimycobacterial drugs and appeared to fare no worse than uninfected patients. Furthermore, most of these infections were described before the recognition of anti-GM-CSF autoantibodies as being causal, so these reports may be confounded by heterogeneity within the underlying diagnosis.

Extrapulmonary infections have been observed with some frequency, which could be ascribed to the systemic effects of anti-GM-CSF autoantibodies identified in the circulation. In support of this systemic effect, several case reports describe patients with PAP who had extrapulmonary infections, including cases of central nervous system (CNS) *Nocardia* infection, septic arthritis, perinephric abscess, and *Nocardia* dissemination to skin; CNS *Aspergillus* and *Proteus;* and disseminated histoplasmosis.[1] A few case series have also reported opportunistic infections associated with high-titer neutralizing anti-GM-CSF autoantibodies without concurrent PAP. Anti-GM-CSF antibodies have been reported in previously healthy adults uninfected by HIV who had cryptococcal meningitis specifically caused by *Cryptococcus gattii*.[20,21] Neutralizing anti-GM-CSF autoantibodies were also found in five of seven patients with disseminated/extrapulmonary *Nocardia* infections.[22] However, it remains to be seen if these infections are the only manifestations in these patients, or if they may eventually develop PAP, as seen in the two of seven cases of cryptococcal meningitis associated with anti-GM-CSF autoantibodies.[20] Why the same autoantibody may produce different clinical phenotypes is unknown. Opportunistic infections have not been reported as a complication of congenital PAP. Reasons for this could include the extreme rarity of this condition or limited opportunity for exposure to environmental opportunists as a result of medical debilitation.

Treatment includes whole lung lavage (WLL) to evacuate the proteinaceous material contained in the alveoli. Unfortunately, WLL is an invasive procedure that only temporizes the symptoms of PAP without treating the underlying cause, often resulting in the need for repeat procedures. Alternatively, inhaled or subcutaneous GM-CSF has been effective, either by saturating the antibody or by inducing tolerance.[23,24] In two large studies, one using subcutaneous GM-CSF and one using inhaled GM-CSF, a clinical response was not associated with a reduction in plasma or BAL concentrations of anti-GM-CSF autoantibodies, thereby providing possible support for the former mechanism.[23,24] B-cell targeted therapy using the anti-CD20 chimeric monoclonal antibody (mAb) rituximab has been used to treat a small number of patients and has shown encouraging clinical results.[25]

ANTI IFN-Γ AUTOANTIBODIES AND SUSCEPTIBILITY TO INTRACELLULAR PATHOGENS

Interferon gamma, produced predominantly by activated Th1 cells and natural killer (NK) cells, is central to host defense against intracellular pathogens (Chapter 26). The IFN-γ receptor (IFN-γR), expressed primarily on monocytes, is composed of two subunits in duplicate, IFN-γR1 and IFN-γR2. Binding of IFN-γ to its receptor leads to Janus kinase 2 (JAK2) and then JAK1 phosphorylation on the intracellular portions of IFN-γR2 and IFN-γR1, respectively. Subsequent STAT1 docking, phosphorylation, homodimerization, and nuclear translocation lead to transcription of IFN-γ responsive genes. Macrophage activation, differentiation, and elaboration of inflammatory mediators, such as tumor necrosis factor-α (TNF-α) and IL-12, ensue. Defects in the IFN-γ–IL-12 axis lead to mendelian susceptibility to mycobacterial disease as well as those caused by other intracellular pathogens, including listeriosis, salmonellosis, histoplasmosis, melioidosis, and penicilliosis.[26] The list of genetic mutations involving this pathway that result in increased susceptibility to mycobacteria or other intracellular pathogens continues to expand and, to date, includes mutations in IFN-γR1, IFN-γR2, STAT1, IL-12p40, IL-12Rb1, nuclear factor-κB (NF-κB) essential modulator (NEMO), IFN regulatory factor (IRF) 8, and IFN-stimulated gene (ISG) 15.[26] Neutralizing autoantibodies against IFN-γ represent an alternative mechanism by which the IFN-γ–IL-12 metabolic pathway is disrupted, with the first cases described in 2004[12] (Fig. 42.2). We reported 85 patients identified in a 6-month period,[2] and the number of patients is still growing,[10–15] suggesting that immunodeficiency caused by anti–IFN-γ autoantibodies is probably underappreciated.

The infections in patients with anti-IFN-γ autoantibodies mimic many of those seen in patients with inborn errors in the IL-12–IFN-γ signaling pathways and include mycobacterial, particularly nontuberculous environmental mycobacteria, nontyphoidal *Salmonella, Penicillium, Histoplasma, Cryptococcus, Burkholderia pseudomallei*, and additionally varicella-zoster virus (VZV), both dermatomal and disseminated.[2] Infections have been noted in all organ systems, although lymph nodes, skin, soft tissue, and bone appear to be preferentially affected. Up to 50% of patients develop sterile reactive dermatoses, most commonly neutrophilic dermatosis, but also erythema nodosum, pustular psoriasis, and exanthematous pustulosis.

Although patients with Mendelian defects tend to present in childhood, all reported cases of patients with anti–IFN-γ were

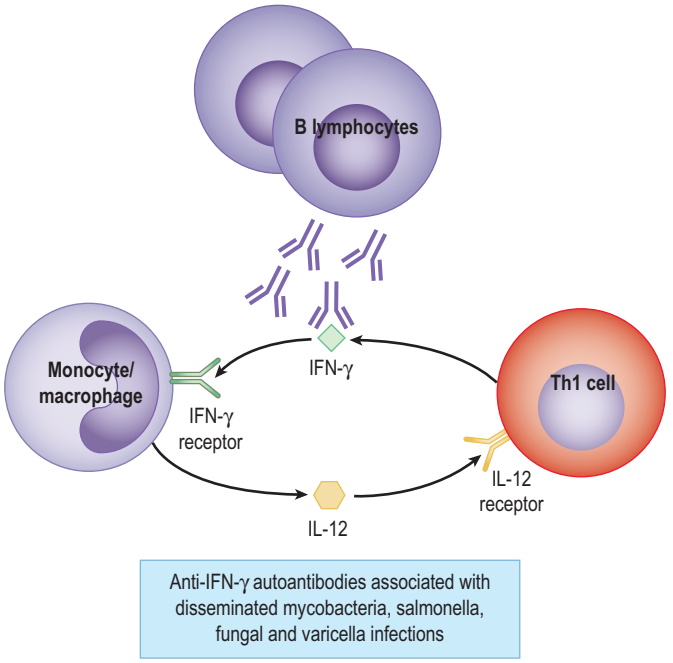


FIG. 42.2 Anti-Interferon (IFN)γ Autoantibody Associated Susceptibility to Intracellular Pathogens. Neutralizing autoantibodies against IFN-γ disrupt the IFN-γ-IL-12 metabolic pathway. Infections in patients with anti-IFN-γ autoantibodies mimic many of the infections seen in patients with inborn errors in the interleukin (IL)-12-IFN-γ signaling pathways and include mycobacterial, nontyphoidal *Salmonella, Penicillium, Histoplasma, Cryptococcus, Burkholderia pseudomallei*, and varicella zoster virus. Th1 cell, type 1 T-helper cell.

previously healthy adults, the vast majority of whom were Asian-born Asians. There is now a single case report in a juvenile diagnosed with disseminated NTM in the presence of anti-IFN-γ autoantibodies.[27] A recent association with human leukocyte antigen–antigen D related (HLA-DR)*15:02/16:02 and HLA-DQ*05:01/05:02 was shown in 78 patients with anti–IFN-γ autoantibodies,[28] further implicating a genetic predisposition to development of disease. However, no familial clustering has been observed, and Asians born outside of Asia have yet to be reported with this syndrome, suggesting complex genetics and possibly an environmental contribution to autoantibody pathogenesis. The publication of an American-born, Caucasian female with this syndrome also adds to the complexity of its pathogenesis.[15]

A recent study compared 74 Thai patients with anti-IFN-γ autoantibodies to an existing US cohort of 19 patients and found several distinctions. *Mycobacterium abscessus* was the most commonly isolated NTM species in the Thai group while in the US, *Mycobacterium avium* complex was most commonly identified. The most commonly involved sites also varied amongst the groups with lymph nodes and skin being common in the Thai cohort while bone, lung and central nervous system involvement was more common in the US cohort. Sweet syndrome was a common disease-associated condition at presentation in Thailand, a condition uncommonly seen in the US, but which should nevertheless trigger consideration for autoantibodies. In both groups, tracking of autoantibody titers suggested that the largest decrease in titers over time portends a better prognosis.[11]

Notable laboratory features include anemia of chronic disease and elevation in inflammatory markers, such as erythrocyte sedimentation rate (ESR), CRP, and β_2 microglobulin. Immunologically, patients commonly demonstrate polyclonal hypergammaglobulinemia, but they otherwise have normal lymphocyte phenotyping, including CD4 T lymphocytes, monocyte numbers, and IFN-γR1 expression.

Treatment has focused mainly on managing the infections with targeted antimicrobials. Interestingly, a severe paradoxical inflammatory reaction similar to immune reconstitution syndrome, manifesting as lymphadenopathy, cavitary lung lesions, and lytic bone lesions during antituberculosis treatment, has been reported and should be distinguished from true failure of antimicrobial treatment.[29] In cases refractory to anti-infective agents, some have attempted to overcome the antibody with IFN-γ administration or to drive down antibody levels with plasmapheresis and cyclophosphamide[30] or rituximab.[10] There are anecdotal cases of treatment with bortezomib, a proteasome inhibitor, and daratumumab, a CD38-directed monoclonal antibody, however there are not enough data on these newer treatment modalities. It is unclear what factors predict response to antimicrobials alone *versus* a need for further immunomodulation; the efficacy of these immune modulatory approaches remains to be evaluated formally in clinical trials.

ANTI-IL-17 AND ANTI-IL-22 AUTOANTIBODIES AND CHRONIC MUCOCUTANEOUS CANDIDIASIS

IL-17A and IL-17F are proinflammatory cytokines that can combine as either homodimers or heterodimers with each other (Chapter 14). These dimerized combinations signal via IL-17RA and IL-17RC heterodimeric receptor complexes, ultimately activating NF-κB. IL-22, produced by T lymphocytes and NK cells, also signals via a heterodimeric receptor composed of IL-22R1 and IL-10R2 subunits, expressed mainly on epithelial and other nonimmune cells. IL-17A/F and IL-22 cooperate to induce proinflammatory cytokines involved in granulopoiesis and neutrophil recruitment as well as antimicrobial peptides, such as β defensins and S100 proteins, which are thought to be important in mucosal immunity. Clinical evidence for a protective role of IL-17 in CMC arose from the observation of this infectious complication in diseases with varying degrees of Th17 impairment, including *STAT3*-deficient hyper-IgE syndrome (HIES, or Job's syndrome), dectin-1 deficiency, CARD9 deficiency, and, to a lesser degree, IL-12 receptor β_1 deficiency.[32] Strong support of this hypothesis came via instances of two families that demonstrated inherited susceptibility to mucosal candidiasis: one with an autosomal recessive mutation in IL-17RA, and another with an autosomal dominant negative mutation in IL-17F.[31]

APECED leads to a clinical triad of hypoparathyroidism, adrenal insufficiency, and CMC.[16] Other endocrine glands, including gonads, thyroid, endocrine cells in the gut, and pancreatic islet cells, are variably affected. Many other autoimmune phenomena, including Sjögren syndrome, rheumatoid arthritis, hepatitis, keratitis, vitiligo, pernicious anemia, and alopecia, as well as autoantibodies to type I IFNs, have been described.[16] The biological consequences of these anti–IFN-α autoantibodies are unclear, but autoantibodies to other tissue antigens in APECED, such as autoantibodies to glutamic acid decarboxylase, thyroid peroxidase, and 21-hydroxylase, are clearly pathological. Another important consideration in patients with APECED is the association of autoimmune pneumonitis. A recent study found that up to 40% of APECED patients, especially those with

autoantibodies against BPIFB1 (BPI fold containing family B member 1) and KCNRG (potassium channel regulator), have clinical and/or radiographic findings associated with pneumonitis. If left untreated, these patients may develop further complications such as hypoxemic respiratory failure and death. However, if discovered early enough, these patients respond well to treatment with T and B lymphocyte-directed immunomodulation.[32] Providing an explanation for the preponderance of autoimmunity seen in APECED was the observation that the transcriptional activity of *AIRE* in medullary thymic epithelial cells (mTECs) promotes expression of tissue-specific genes, thereby facilitating intrathymic destruction of autoreactive T cells and fostering self-tolerance.[16]

Unlike the above-described Mendelian disorders, the mechanism of CMC in APECED is not directly linked to the genetic deficiency itself but, rather, to the strong genetic predisposition to autoimmunity, which, in this case, includes production of neutralizing anti-IL-17A, anti-IL-17F and anti-IL-22 autoantibodies[3,4] (Fig. 42.3). Puel et al. identified antibodies to IL-17A, IL-17F, or IL-22 in 33 patients with APECED, 29 of whom also had CMC, compared with healthy controls who had neither autoantibodies nor CMC.[4] Kisand et al. found autoantibodies against IL-17A, IL-17F, or IL-22 in up to 90% of 162 cases of APECED, also strongly associated with CMC.[3] They also identified anti-IL-17 and anti-IL-22 autoantibodies in two patients with thymoma and CMC, but in none of the 33 patients with thymoma who did not have CMC. There were a few instances of autoantibodies without CMC, and many examples of CMC independent of APECED, suggesting that autoantibodies to IL-17 and IL-22 are not absolutely necessary for development of CMC. However, their identification, particularly in light of the CMC disease seen in patients with IL-17RA and IL-17F mutations, provides support for a causal relationship. Interestingly, thymoma tissue also demonstrates decreased *AIRE* expression, further linking these diseases beyond a shared tendency toward autoimmunity and anti-cytokine autoantibody production. What is less clear however, is the exact role *AIRE* is playing in pathogenesis of organ-specific autoimmunity (including anti-cytokine autoantibody formation), since the syndromes diverge considerably in this regard.[16]

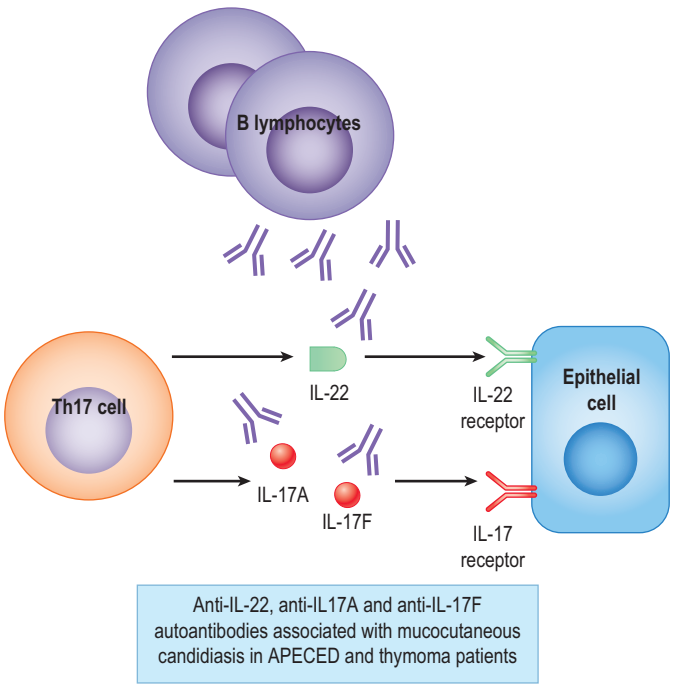


FIG. 42.3 Anti-Interleukin (IL)-17 and Anti-IL-22 Autoantibodies Associated Chronic Mucocutaneous Candidiasis (CMC). IL-17A/F and IL-22 induce proinflammatory cytokines involved in granulopoiesis and neutrophil recruitment as well as important mucosal antimicrobial peptides, such as β defensins and S100 proteins. Neutralizing antibodies against IL-17A, IL-17F, and IL-22 in autoimmune polyendocrinopathy with candidiasis and ectodermal dysplasia (APECED) syndrome constitute the likely mechanism of CMC in these patients.

ANTI-IL-6 AUTOANTIBODIES AND BACTERIAL INFECTIONS

IL-6 is produced by many immune and nonimmune cells, including B cells, T cells, macrophages, synovial cells, endothelial cells, and hepatocytes, and is involved in both acute and chronic inflammation, ranging from sepsis to rheumatoid arthritis. IL-6 binds a heterodimeric receptor composed of IL-6Rα and a shared receptor chain gp130. IL-6Rα confers ligand specificity, whereas gp130 mediates signal transduction. IL-6 regulates the acute phase response in the liver, with its hallmark induction of CRP and elevated ESR. Anti-IL-6 autoantibodies were first identified in a Haitian boy who had two episodes of severe staphylococcal cellulitis, one complicating chickenpox infection, the other following mosquito bites.[5] Treatment included supportive care and anti-infective agents. Undetectable CRP, despite severe infection, suggested impairment in IL-6 activity. Since that time, there have been three other reported cases of anti-IL-6 autoantibodies in patients with severe bacterial infections without elevation in CRP: a 67-year-old man with thoracic empyema with *Escherichia coli* and *Streptococcus intermedius*, a 56-year-old woman with multiple subcutaneous abscesses with *Staphylococcus aureus*, and a 20-month-old female with septic shock presumed to be caused by *Staphylococcus aureus*.[8,9] All three cases demonstrated low levels of CRP despite the severity of their respective infections, prompting the search for anti-IL-6 autoantibodies. *STAT3* is the critical signal transduction molecule for IL-6 and IL-10, and autosomal dominant mutations in *STAT3* also lead to recurrent infections due to staphylococcus, streptococcus and hemophilus,[33] suggesting a common potential mechanism for this phenotypic profile of infection susceptibility. The limitations to this association are that the infections in the first reported case resolved without any apparent change in the anti-IL-6 autoantibody titers, and there are still a very limited number of reported cases to date.

MANAGEMENT

Therapies directed at the syndromes of pathogenic autoantibodies have focused either on treating the disease consequences or targeting the autoantibody itself. Therapeutic BAL removes proteinaceous alveolar material from the lungs of patients with PAP,[1] and anti-mycobacterials treat the disseminated NTM infection in patients with autoantibodies to IFN-γ.[12,34] Approaches ranging from physical removal of the antibody, to immunomodulatory therapy, to induction of immune tolerance or suppression of the population of cells that produce the pathogenic autoantibodies have also been employed.

PAP and disseminated NTM associated with anti–IFN-γ autoantibodies have been treated with exogenous GM-CSF[23,24] and IFN-γ,[35] respectively, resulting in clinical improvement, although PAP has been studied more rigorously in this regard. Autoantibody levels after therapy were not routinely evaluated across diseases, nor were they measured in a standardized fashion, although, at least in the PAP cohorts, it appears they did not change in response to exogenous cytokine administration.[23,24]

In cases that were refractory to treatment, attempts have been made to alleviate the blockade by reducing anti-cytokine autoantibody levels. One patient with anti-IFN-γ autoantibodies underwent plasmapheresis and cyclophosphamide therapy in addition to receiving antimycobacterials.[31] Rituximab, the mouse-human chimeric mAb that targets the human B-cell marker CD20, is currently approved for treatment of B-cell lymphoma and rheumatoid arthritis. Since B cells ultimately differentiate into antibody-producing cells, it has been applied to a number of autoantibody-mediated diseases, such as myasthenia gravis (Chapter 65) and pemphigus vulgaris (a blistering skin disease caused by autoantibodies that recognize desmoglein 3, a keratinocyte cell-surface protein; Chapter 63). Successful rituximab therapy has also been reported in PAP[25] and immunodeficiency caused by anti-IFN-γ autoantibodies[10] with a specific reduction of anticytokine–autoantibody titers.

A new approach in pure red blood cell (RBC) aplasia caused by anti-erythropoietin autoantibodies is to bypass the autoantibody with a erythropoietin receptor synthetic peptide agonist peginesatide (Hematide; Affymax) that does not share homology with the eosinophil peroxidase (EPO) ligand.[35] Although this approach has not been used in cases of anti-cytokine autoantibodies causing immune deficiency, it underscores the range of novel treatment approaches that might be explored for anti-cytokine autoantibody associated syndromes.

ON THE HORIZON

- Anti-cytokine autoantibodies are an emerging mechanism underlying the pathogenesis of immune deficiency in previously "healthy" adults, and perhaps children.
- The clinical manifestations of these diseases, like pulmonary involvement in anti-granulocyte macrophage-colony-stimulating factor (GM-CSF) associated pulmonary alveolar proteinosis (PAP), may provide insight to signaling pathways and target cells of the involved cytokine.
- Recent discovery of many anti-cytokine autoantibody associated diseases suggests that other idiopathic immunodeficiency diseases may be the result of an autoantibody affecting an as-yet unknown signaling pathway or cell type.
- Much research needs to be done on the management of patients with these syndromes, since besides just treating the symptomatic consequences of the autoantibody, targets to control disease may include the autoantibody itself or cells or pathways leading to its generation.
- Profiling anti-cytokine autoantibodies in patients with immunodeficiency could help personalize their management by potentially predicting disease manifestations and optimizing treatment options.

CONCLUSIONS

Anti-cytokine autoantibodies have been identified in healthy adults as well as in those with different diseases, suggesting that their occurrence may range from a normal homeostatic mechanism, to epiphenomena, to being directly pathogenic. In PAP, the identification of anti-GM-CSF autoantibodies came over 40 years after the initial description of the syndrome, suggesting that other currently idiopathic diseases may someday be explained by the identification of neutralizing or agonizing autoantibodies. Furthermore, it was not intuitive that systemic autoantibodies to a hematopoietic growth factor should result in disease confined mainly to the lung. The identification of anti-GM-CSF autoantibodies in cryptococcal meningitis implicates the GM-CSF pathway in host defense of this infection, much as Mendelian disorders have done for NTM and the IFN-γ–IL-12 pathway. The observation of high-titer, neutralizing anti-cytokine autoantibodies in an expanding number of diseases, beyond those characterized only by immune deficiency,[7] combined with the opportunity for novel therapeutic approaches to their diagnoses, mandates that their presence be not merely considered but rigorously sought.

ACKNOWLEDGMENTS

This work was supported by the Divisions of Intramural Research of National Institute of Allergy and Infectious Diseases (NIAID) of the National Institutes of Health (NIH), Bethesda, MD.

REFERENCES

1. Seymour JF, Presneill JJ. Pulmonary alveolar proteinosis: progress in the first 44 years. *Am J Respir Crit Care Med*. 2002;166:215–235.
2. Browne SK, Burbelo PD, Chetchotisakd P, et al. Adult-onset immunodeficiency in Thailand and Taiwan. *N Engl J Med*. 2012;367:725–734.
3. Kisand K, Boe Wolff AS, Podkrajsek KT, et al. Chronic mucocutaneous candidiasis in APECED or thymoma patients correlates with autoimmunity to Th17-associated cytokines. *J Exp Med*. 2010;207:299–308.
4. Puel A, Doffinger R, Natividad A, et al. Autoantibodies against IL-17A, IL-17F, and IL-22 in patients with chronic mucocutaneous candidiasis and autoimmune polyendocrine syndrome type I. *J Exp Med*. 2010;207:291–297.
5. Puel A, Picard C, Lorrot M, et al. Recurrent staphylococcal cellulitis and subcutaneous abscesses in a child with autoantibodies against IL-6. *J Immunol*. 2008;180:647–654.
6. Casadevall N, Dupuy E, Molho-Sabatier P, et al. Autoantibodies against erythropoietin in a patient with pure red-cell aplasia. *N Engl J Med*. 1996;334:630–633.
7. Riches PL, McRorie E, Fraser WD, et al. Osteoporosis associated with neutralizing autoantibodies against osteoprotegerin. *N Engl J Med*. 2009;361:1459–1465.
8. Nanki T, Onoue I, Nagasaka K, et al. Suppression of elevations in serum C reactive protein levels by anti-IL-6 autoantibodies in two patients with severe bacterial infections. *Ann Rheum Dis*. 2013;72:1100–1102.
9. Bloomfield M, Parackova Z, Cabelova T, et al. Anti-IL6 Autoantibodies in an Infant With CRP-Less Septic Shock. *Front Immunol*. 2019;10:2629.
10. Browne SK, Zaman R, Sampaio EP, et al. Anti-CD20 (Rituximab) therapy for anti-IFN-γ autoantibody-associated nontuberculous mycobacterial infection. *Blood*. 2012;119:3933–3939.
11. Hong G, Ortega-Villa A, Hunsberger S, et al. Natural History and Evolution of Anti-Interferon-γ Autoantibody-Associated Immunodeficiency Syndrome in Thailand and the United States. *Clin Infect Dis*. 2019; Aug 20. https://doi.org/10.1093/cid/ciz786. [Epub ahead of print].
12. Hoflich C, Sabat R, Rosseau S, et al. Naturally occurring anti-IFN-γ autoantibody and severe infections with Mycobacterium cheloneae and Burkholderia cocovenenans. *Blood*. 2004;103:673–675.
13. Kampitak T, Suwanpimolkul G, Browne S, et al. Anti-interferon-γ autoantibody and opportunistic infections: case series and review of the literature. *Infection*. 2011;39:65–71.
14. Patel SY, Ding L, Brown MR, et al. Anti-IFN-γautoantibodies in disseminated nontuberculous mycobacterial infections. *J Immunol*. 2005;175:4769–4776.
15. O'Connell E, Rosen LB, LaRue RW, et al. The first US domestic report of disseminated Mycobacterium avium complex and anti-interferon-γ autoantibodies. *J Clin Immunol*. 2014;34:928–932.

16. Kisand K, Peterson P. Autoimmune polyendocrinopathy candidiasis ectodermal dystrophy. *J Clin Immunol*. 2015;35:463–478.
17. Huang H, Benoist C, Mathis D. Rituximab specifically depletes short-lived autoreactive plasma cells in a mouse model of inflammatory arthritis. *Proc Natl Acad Sci USA*. 2010;107:4658–4663.
18. Uchida K, Beck DC, Yamamoto T, et al. GM-CSF autoantibodies and neutrophil dysfunction in pulmonary alveolar proteinosis. *N Engl J Med*. 2007;356:567–579.
19. Witty LA, Tapson VF, Piantadosi CA. Isolation of mycobacteria in patients with pulmonary alveolar proteinosis. *Medicine (Baltimore)*. 1994;73:103–109.
20. Rosen LB, Freeman AF, Yang LM, et al. Anti-GM-CSF autoantibodies in patients with cryptococcal meningitis. *J Immunol*. 2013;190:3959–3966.
21. Saijo T, Chen J, Chen SC, et al. Anti-granulocyte-macrophage colony-stimulating factor autoantibodies are a risk factor for central nervous system infection by Cryptococcus gattii in otherwise immunocompetent patients. *MBio*. 2014;5:e00912–e00914.
22. Rosen LB, Rocha Pereira N, Figueiredo C, et al. Nocardia-induced granulocyte macrophage colony-stimulating factor is neutralized by autoantibodies in disseminated/extrapulmonary nocardiosis. *Clin Infect Dis*. 2015;60:1017–1025.
23. Kavuru MS, Sullivan EJ, Piccin R, et al. Exogenous granulocyte-macrophage colony-stimulating factor administration for pulmonary alveolar proteinosis. *Am J Respir Crit Care Med*. 2000;161:1143–1148.
24. Tazawa R, Trapnell BC, Inoue Y, et al. Inhaled granulocyte/macrophage-colony stimulating factor as therapy for pulmonary alveolar proteinosis. *Am J Respir Crit Care Med*. 2010;181:1345–1354.
25. Kavuru MS, Malur A, Marshall I, et al. An open-label trial of rituximab therapy in pulmonary alveolar proteinosis. *Eur Respir J*. 2011;38:1361–1367.
26. Bustamante J, Boisson-Dupuis S, Abel L, et al. Mendelian susceptibility to mycobacterial disease: genetic, immunological, and clinical features of inborn errors of IFN-gamma immunity. *Semin Immunol*. 2014;26:454–470.
27. Liew W-K, Thoon K-C, Chong C-Y, et al. Juvenile-Onset Immunodeficiency Secondary to Anti-Interferon-Gamma Autoantibodies. *J Clin Immunol*. 2019;39:512–518.
28. Ku CL, Lin CH, Chang SW, et al. Anti-IFN-γ autoantibodies are strongly associated with HLA-DR*15:02/16:02 and HLA-DQ*05:01/05:02 across Southeast Asia. *J Allergy Clin Immunol*. 2016;137:945–948. e8.
29. Xie YL, Rosen LB, Sereti I, et al. Severe paradoxical reaction during treatment of disseminated tuberculosis in a patient with neutralizing anti-IFN-γ autoantibodies. *Clin Infect Dis*. 2016;62:770–773.
30. Baerlecken N, Jacobs R, Stoll M, et al. Recurrent, multifocal Mycobacterium avium-intercellulare infection in a patient with interferon-γ autoantibody. *Clin Infect Dis*. 2009;49:e76–e78.
31. McDonald DR. TH17 deficiency in human disease. *J Allergy Clin Immunol*. 2012;129:1429–1435, quiz 36–7.
32. Ferre EMN, Break TJ, Burbelo PD, et al. Lymphocyte-driven regional immunopathology in pneumonitis caused by impaired central immune tolerance. *Sci Transl Med*. 2019;11(495).
33. Bergerson JRE, Freeman AF. An Update on Syndromes with a Hyper-IgE Phenotype. *Immunol Allergy Clin North Am*. 2019;39:49–61.
34. Kampmann B, Hemingway C, Stephens A, et al. Acquired predisposition to mycobacterial disease due to autoantibodies to IFN-gamma. *J Clin Invest*. 2005;115:2480–2488.
35. Macdougall IC, Rossert J, Casadevall N, et al. A peptide-based erythropoietin-receptor agonist for pure red-cell aplasia. *N Engl J Med*. 2009;361:1848–1855.

Allergic Airway Diseases

David B. Corry, Evan Li, and Amber U. Luong

The allergic airway diseases comprise a large and disparate group of respiratory disorders that are characterized by airway and parenchymal inflammation that impairs sinus and lung function. The physiologic importance of the airways combined with the need to respond immunologically to an extremely broad range of infectious and irritant particulates, aerosols, and gases explains the diverse nature of airway immune disorders and their disproportionately large effect on human health. The allergic respiratory tract immune disorders covered in this chapter are among the most common of human afflictions. Non-allergic lung disorders are considered in Chapters 72 and 73.

Allergic disorders have a common immune phenotype of characteristic cellular, humoral, biochemical, and molecular components. Eosinophils, neutrophils, and tissue mast cells are easily seen on conventional hematoxylin and eosin staining of pathologic specimens. Less obvious on histochemical staining, but equally important to allergic diseases, are B cells that secrete the antibody isotypes immunoglobulin E (IgE) and IgG4, and T helper 2 (Th2) and Th17 cells that secrete a repertoire of cytokines, including interleukin-4 (IL-4), IL-5, IL-9, IL-10, IL-13, and IL-17A that coordinate and activate other allergic effector cells such as eosinophils, mast cells, B cells, and innate lymphoid cells (ILCs) in addition to target cells of the airway such as airway epithelium. The allergic airway diseases are typically chronic and occasionally fatal; although spontaneous remissions are not uncommon, they are rarely curable. However, recent insights have vastly improved prospects for improved therapy.

CLINICAL PRESENTATION OF ALLERGIC AIRWAY DISEASE

Although the diverse effector cells and molecules that characterize allergic inflammatory exudates can be seen anywhere along the respiratory tract, the functional impact of allergic disease is quite different in the upper and lower airways.

Chronic Rhinitis and Rhinosinusitis

Epidemiology and Clinical Presentation

The major upper airway inflammatory disorders are rhinitis and chronic rhinosinusitis (CRS). Rhinitis involves inflammation of the nasal mucosa categorized as either allergic, an IgE-mediated inflammation, or non-allergic rhinitis. Typical rhinitis symptoms include clear rhinorrhea, postnasal drip, and nasal congestion with sneezing and nasal pruritus, often associated with ocular symptoms such as conjunctivitis and tearing, differentiating allergic rhinitis (AR) from non-allergic rhinitis. Based on symptom duration, AR is currently classified as "intermittent" or "persistent" with severity of symptoms noted as "mild" or "moderate-severe."

KEY CONCEPTS

Classification of Allergic Rhinitis

Frequency of Symptoms	Severity of Symptoms
Intermittent	**Mild**
Symptoms present for less than 4 days to a week Or for less than 4 weeks	No presence of sleep disturbance Or impairment of daily activities, leisure and/or sport Or impairment of school or work Or troublesome symptoms
Persistent	**Moderate-Severe**
Symptoms present for more than 4 days to a week Or for more than 4 weeks	Presence of sleep disturbance And/or impairment of daily activities, leisure and/or sport And/or impairment of school or work And/or troublesome symptoms

In contrast to rhinitis, rhinosinusitis affects both the sinuses and nasal mucosa. Rhinosinusitis can be subdivided into acute and CRS based on duration of symptoms. CRS is reported to affect approximately 29 million Americans and is defined by symptoms persisting longer than 12 weeks. Clinically, CRS is broadly categorized based on the absence or presence of nasal polyps. Within CRS with nasal polyps (CRSwNP), there are several subtypes, including allergic fungal rhinosinusitis (AFRS), aspirin-exacerbated respiratory disease (AERD), and cystic fibrosis.

Signs and symptoms associated with rhinosinusitis in adults include facial pain and pressure, headaches, nasal congestion with or without obstruction, frontal or postnasal drainage, generalized malaise, and cough. In contrast, symptoms in children are age related and require the caretaker to recognize them. Young children often present with a chronic cough, and irritability rather than facial pain. The prevalence of rhinosinusitis in children is inversely related to age.[1]

Diagnosis

The diagnosis of AR is based on a history of typical symptoms and common physical exam findings. Several physical findings include "allergic shiners," a darkening of the infraorbital

skin resulting from chronic venous pooling, and a persistent horizontal crease across the nasal bridge resulting from constantly wiping the front of the nose. Bilateral conjunctivitis may be present in patients with ocular involvement.

On anterior rhinoscopy, engorged, boggy, and pale inferior turbinates with clear discharge coating the nasal cavity are supportive of AR. Examination of the oropharynx often reveals cobblestoning of the mucosa, indicating chronic postnasal drainage.

Although not necessary to make the diagnosis of AR, two tests commonly used to demonstrate IgE-mediated allergic reactions are immediate-hypersensitivity skin testing and measurement of serum allergen-specific IgE.

The diagnosis of CRS requires the presence of at least two major or one major and at least two minor clinical symptoms persisting for longer than 12 weeks in conjunction with evidence of inflammation within the sinus cavity. Major symptoms include facial pain/pressure, nasal obstruction, nasal drainage, and hyponosmia or anosmia. Minor symptoms are headaches, halitosis, fatigue, dental pain, cough, and ear pain or pressure. The most specific symptom for rhinosinusitis is the presence of discolored rhinorrhea. Evidence of inflammation on nasal endoscopy or imaging is also necessary to make a CRS diagnosis. On nasal endoscopy, inflammation is suggested by edema and/or drainage from the middle meatus; a diagnosis of CRSwNP is made when nasal polyps are visualized (Fig. 43.1). In patients with a history of CRS but normal nasal endoscopy, computed topography (CT) scanning of the sinuses is necessary to evaluate for mucosal thickening and/or fluid within the sinuses (Fig. 43.2).

Clinical CRSwNP subtypes are defined by additional criteria. For AFRS, the widely accepted criteria include five characteristics: nasal polyps; type I (immediate) hypersensitivity to fungi; radiographic imaging consistent with AFRS; eosinophilic mucin with evidence of fungi; and a lack of evidence of fungal invasion into the surrounding tissue.

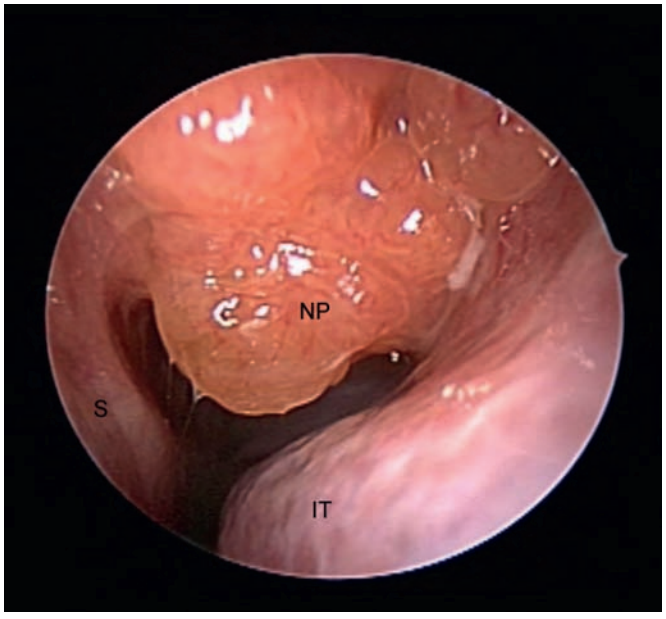


FIG. 43.1 Nasal Polyposis. Seen on this nasal endoscopy are nasal polyps *(NP)* emanating from the sinus cavity into the nasal cavity between the septum *(S)* and the inferior turbinate *(IT)* of a patient with chronic rhinosinusitis.

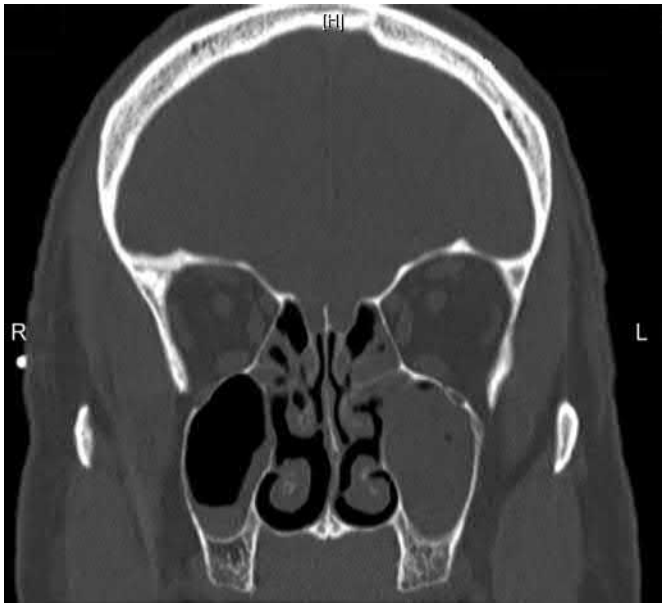


FIG. 43.2 Coronal Sinus Computed Tomography Image From a Patient With Chronic Rhinosinusitis. The maxillary sinuses (lateral to the nasal cavity) and ethmoid sinuses (medial to the orbital cavities) exhibit mucosal thickening and accumulation of obstructed secretions consistent with inflammatory changes within the paranasal sinuses.

KEY CONCEPTS

Diagnosis of Chronic Rhinosinusitis

Chronic Rhinosinusitis

- Presence of at least two of the following symptoms
  - Facial pressure or pain
  - Nasal obstruction or congestion
  - Anterior or posterior nasal drainage
  - Hyposmia or anosomia
- Edema or discolored drainage within the sinus cavity or middle meatus
- *Or* CT sinus showing fluid or mucosal thickening within sinus cavities

Chronic Rhinosinusitis Wthout Nasal Polyps	Chronic Rhinosinusitis With Nasal Polyps
No evidence of nasal polyps within the middle meatus as noted on nasal endoscopy in a patient with no history of previous sinus surgery	Presence of nasal polyps within the middle meatus as noted on nasal endoscopy History of nasal polyps within sinus cavity in a previously operated patient with CRS diagnosis

The diagnosis of AERD requires a history of respiratory symptoms exacerbated by oral intake of aspirin or other cyclooxygenase-1 (COX-1) inhibitors on at least two occasions or a positive reaction on aspirin challenge.[2]

Cystic fibrosis is typically diagnosed in early childhood as a result of pulmonary symptoms. However, nasal polyps in anyone younger than 18 should prompt evaluation for cystic fibrosis, confirmed by either a sweat test or a genetic test evaluation for a mutation in the cystic fibrosis transmembrane conductance regulator gene.[2]

Therapy

The most effective and widely used pharmaceutical approaches to controlling nasal and ocular symptoms of AR are intranasal

glucocorticosteroids and antihistamines. Other drugs potentially useful in rhinitis management include oral steroids, oral leukotriene (LT) receptor antagonists, intranasal chromones, and intranasal ipratropium bromide.

Unlike drugs, allergen immunotherapy is offered to change the immunologic response to antigens with chronic controlled exposure. This therapy typically requires 3 to 5 years of treatment. Two types of immunotherapy are available: subcutaneous immunotherapy (SCIT) and sublingual immunotherapy (SLIT). A concern associated with SCIT is anaphylaxis; this risk is much lower with SLIT.[3]

The treatment of CRS is more controversial than that of AR. In general, treatment consists of medical and/or surgical therapy. CRS is primarily considered to be an inflammatory rather than infectious disorder; as such, antibiotics are less frequently advocated and usually replaced with intranasal and oral steroids. In addition, daily sinus irrigations with physiologic saline represents a critical therapeutic component to counteract mucociliary dysfunction and promote antigen clearance.[4]

Patients with persistent symptoms despite medical therapy are candidates for functional endoscopic sinus surgery (FESS). In AFRS and AERD with sinuses impacted with thick eosinophilic mucus or with serious complications such as acute vision loss from expanded sinus cavities, endoscopic sinus surgery is mandatory and may be followed by medical therapy.

In those undergoing FESS, revision sinus surgery for recurrent disease despite medical therapy occurs in 16% to 25% of CRS patients. Asthma, elevated serum and/or tissue eosinophils, and noncompliance with postoperative medical therapy are associated with higher risk of nasal polyp recurrence and need for revision sinus surgeries. The newest advance in CRS treatment is anti–IL-4/IL-13 receptor antibody (dupilumab) for CRSwNP (Table 43.1). Several other biologics are undergoing clinical evaluation for management of CRS.[4]

Asthma

Epidemiology and Clinical Presentation

After several decades of rising incidence, asthma is the most common chronic disease of childhood and one of the most common disorders of children and adults in the United States. Although most frequently diagnosed initially in childhood, asthma can be first diagnosed at any age. The prevalent and incurable nature of asthma will continue to ensure it as one of the most expensive of medical afflictions both in terms of medical expenditure and time lost from work and school. Asthma is a lower respiratory tract disease characterized by dyspnea and other symptoms including cough, chest tightness, chest pain, and wheezing. Persons with mild disease often present only with a chronic cough. Symptoms are usually intermittent and are characteristically relieved by bronchodilator and antiinflammatory therapy.

Asthma patients are classified into distinct subtypes according to characteristic environmental or occupational exposures that elicit symptoms, the presence or absence of concomitant atopy, temporal expression of symptoms, and responsiveness to antiinflammatory therapy. Respiratory viruses are the most frequently implicated causes of attacks, especially in children, while tobacco smoke, ozone, and particulate air pollution are other major inciting agents. Many, but not all, asthma patients are atopic, with up to 60% of severe asthmatics showing no evidence of environmental sensitization. If atopy is present, patients are referred to as extrinsic, atopic, or allergic asthmatics, whereas those who lack atopy are referred to as having intrinsic or non-allergic asthma. In general, airway constriction occurs and symptoms are provoked when triggering agents are inhaled, representing the clinical expression of airway hyperresponsiveness, the exaggerated tendency of the airway to constrict in response to exposure to provocative agents. Some of these agents (e.g., viruses and pollens) are only intermittently present, causing seasonal asthma. Other agents are encountered continuously (e.g., fungi, dust mites) and cause persistent (or perennial) asthma. Occupational asthma is defined as asthma acquired in the workplace, where dozens of potentially toxic agents have been identified. Numerous additional asthma subsets can be defined according to the factor or factors that commonly elicit attacks of dyspnea. A final category of asthma, steroid-resistant, refers to patients relatively unresponsive to antiinflammatory steroid therapy.

Diagnosis

Asthma is often recognized on clinical grounds alone, with acute attacks marked by obvious dyspnea, wheezing, cough, and use of accessory muscles of respiration. Confirmation of the diagnosis is ensured if attacks reliably resolve with bronchodilator therapy. Spirometry can provide objective evidence of airway obstruction as assessed by reversible decrements in the forced expiratory volume in 1 second (FEV_1) and other measures of airflow such as peak airflow. When the clinical presentation is uncertain, bronchial provocation tests can determine the

TABLE 43.1 Biologic Agents Approved for Use in Allergic Airway Disease Therapy

Biologic Agent	Chemistry	Immune Target	Approved Use	Administration (Adults only)
Omalizumab	IgG1 kappa monoclonal antibody	IgE	Moderate to severe persistent asthma in patients with a positive skin test or in vitro reactivity to a perennial aeroallergen	75–375 mg every 2–4 weeks based on serum IgE level and body weight, subcutaneous
Benralizumab	IgG1 kappa monoclonal antibody	Alpha chain of the IL-5 receptor	Eosinophilic asthma	30 mg every 4 weeks for 3 weeks, then every 8 weeks, subcutaneous
Mepolizumab	IgG1 kappa monoclonal antibody	IL-5	Eosinophilic asthma	100 mg every 4 weeks, subcutaneous
Reslizumab	IgG4 kappa monoclonal antibody	IL-5	Eosinophilic asthma	3 mg/kg every 4 weeks, intravenous
Dupilumab	IgG4 kappa monoclonal antibody	Alpha chain of the IL-4/IL-13 receptor	Eosinophilic or steroid-dependent asthma; CRS with nasal polyps	400 mg followed by 200 mg every other week or 600 mg followed by 300 mg every other week

CRS, Chronic rhinosinusitis; *Ig*, immunoglobulin; *IL*, interleukin.

presence of airway hyperresponsiveness and thereby establish the diagnosis.[5] Additional laboratory data supporting a diagnosis of allergic asthma include peripheral blood eosinophilia, elevated serum total and antigen-specific IgE levels, and positive skin prick test results against one or more allergens.

Therapy

As with rhinitis and rhinosinusitis, asthma therapy is generally nonspecific and directed at improving airflow through bronchodilation and reducing inflammation. Immediate relief of bronchoconstriction and dyspnea is achieved with bronchodilating agents that activate the β_2 adrenergic receptor on airway smooth muscle beta agonists. For long-term asthma control, the most effective agent class is steroids, often combined with long-acting beta agonists (LABAs), which reduce inflammation and suppress airway constriction and dyspnea. For mild to moderate disease, bronchodilating agents and steroids are typically administered by inhalation, which significantly reduces systemic side effects. A secondary agent class used for controlling bronchospasm is anticholinergics that antagonize the muscarinic acetylcholine receptor. Severe disease may also require high-dose treatment with oral or intravenous steroids and high-dose inhaled beta agonists given by nebulizer.

Additional antiinflammatory agents for asthma treatment include LT receptor antagonists, chromones, theophylline, omalizumab (a monoclonal antibody that reduces circulating and mast cell–bound IgE), and more recently anti–IL-5, anti–IL-5 receptor, and anti–IL-4/IL-13 receptor antibodies (see Table 43.1).

Bronchial thermoplasty (BT) is a relatively new bronchoscopic technique in which heat is applied to the airways via a radiofrequency catheter to ablate airway contractile cells. Clinical trials indicate that this technique reduces exacerbation rates long term, but the risk of initially enhancing exacerbations suggests that BT should be reserved for severe asthmatics with refractory disease.

OTHER AIRWAY ALLERGIC DISEASE SYNDROMES

In addition to CRS and asthma, several other allergic airway diseases involving prominent airway and eosinophilia can cause profound morbidity. These disorders are clinically heterogeneous but are believed to share a similar pathophysiology related to inhalation of antigens that provoke airway eosinophil and Th2 responses.

Eosinophilic disorders are organized according to whether there is an *extrinsic* or *intrinsic* cause of the eosinophilia (Table 43.2). Inhaled or ingested extrinsic factors, including medications and infectious agents (e.g., parasites, fungi, mycobacteria) can trigger an eosinophilic response. This may be mild and self-limited, as in Loeffler syndrome. Intrinsic pulmonary eosinophilic syndromes are either idiopathic, malignant, or premalignant diseases, often systemic in nature.

Extrinsic Eosinophilic Syndromes

Tropical Eosinophilic Pneumonias

The tropical eosinophilic pneumonias comprise a group of clinically similar eosinophil-predominant inflammatory disorders characterized by chest pain, wheezing, cough, and airway hyperresponsiveness, often during a debilitating, but transient, febrile illness. Migrating parasites from the genera *Dirofilaria*, *Strongyloides*, *Wuchereria*, and *Brugia* traversing the lungs are considered responsible for most cases of tropical eosinophilic pneumonia. However, in the United States, *Strongyloides* spp. are the most common cause of parasitic infection and tropical eosinophilic pneumonia. Immunocompromised patients, including those taking systemic steroids, may develop *Strongyloides* hyperinfection syndrome, in which recently hatched larvae burrow through the intestine and migrate to the lungs, causing a severe and potentially fatal lung disease that is frequently complicated by sepsis (Fig. 43.3). Therapy of parasite-related pulmonary eosinophilia syndromes is directed at relieving symptoms and eliminating the parasites and other offending agents.

TABLE 43.2 Eosinophilic Lung Disorders

Disease	Causative Agent	Proposed Immune Mechanism
Loeffler syndrome	Inhaled food, infection, or medication	T cell–mediated hypersensitivity reaction
Drug rash with eosinophilia and systemic symptoms (DRESS) syndrome	Drugs: sulfonamides, phenobarbital, sulfasalazine, carbamazepine, and phenytoin	Hypersensitivity reaction to drug
Parasitic infections	*Strongyloides* spp., *Wuchereria bancrofti*, *Brugia malayi*	T-cell and B-cell clonal activation in response to parasite antigens and adjuvant factors.
Allergic bronchopulmonary aspergillosis	*Aspergillus*	IgE and immune complex deposition
Acute eosinophilic pneumonia	Fungal infections, cigarette smoking, post–stem cell transplant	Hypersensitivity response to inhaled antigen (infectious or otherwise)
Chronic eosinophilic pneumonia	Unknown systemic-mediated process	Unknown, but chronic nature evident with T cell–mediated granuloma production
Idiopathic hypereosinophilic syndrome	Infections, systemic diseases and drugs that drive peripheral eosinophilia	Systemic responses caused in part due to excess interleukin (IL)-5 production from clonal expansion of T helper 2 (Th2) cells as well as fusion gene *FIP1L1–PDGFR*
Churg-Strauss syndrome	Autoimmune vasculitis to unknown antigen, associated with asthma	Decreased T-regulatory cell function with diminished IL-10 production

Allergic Bronchopulmonary Aspergillosis

ABPA is a severe pulmonary allergic reaction to *Aspergillus* antigens seen almost exclusively in the setting of preexisting asthma or cystic fibrosis.[6] Diagnostic criteria include asthma with wheezing, peripheral blood eosinophilia, detection of precipitating anti-*Aspergillus* antibodies, elevated serum total IgE, and radiographic evidence of fleeting pulmonary infiltrates often with central bronchiectasis. *Aspergillus* spp. and other fungi can frequently be isolated from airway secretions of ABPA patients, suggesting active fungal growth within the airways, or

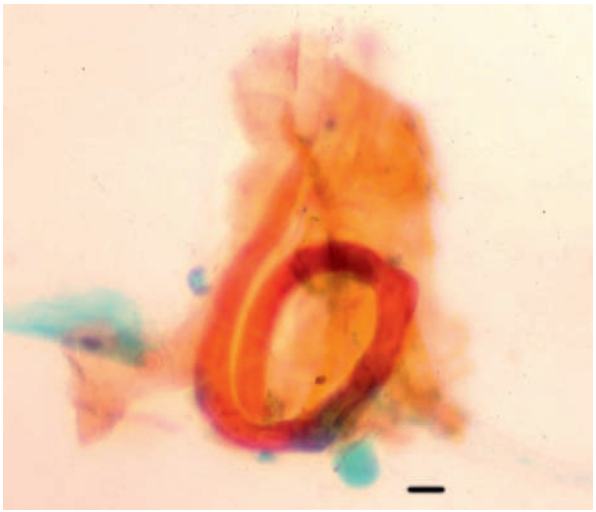


FIG. 43.3 Strongyloidiasis. The coiled larva of *Strongyloides stercoralis* is seen on this Papanicolaou stain of a bronchoalveolar lavage sample from a patient with *Strongyloides* hyperinfection. Original magnification 400×; bar=10 μm.

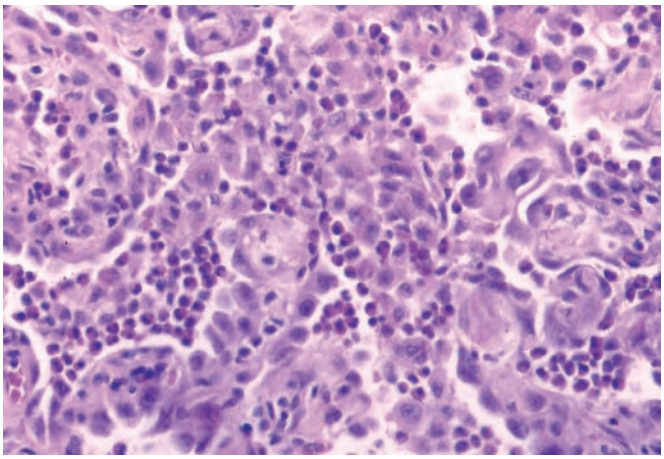


FIG. 43.4 Histology of Chronic Eosinophilic Pneumonia. Lung biopsy specimen from a patient with chronic eosinophilic pneumonia demonstrates a confluent infiltrate with eosinophils filling alveoli together with large, multinucleate macrophages. Original magnification, 200×, hematoxylin and eosin stain.

airway mycosis. Complications of chronic ABPA include severe airway hyperresponsiveness, bronchiectasis, eosinophilic pneumonia, pulmonary fibrosis, and invasive fungal disease. Treatment of ABPA aims to suppress the inflammatory response to the fungus and to control bronchospasm with steroid therapy, the duration of which may be shortened by concomitant use of oral antifungal agents such as itraconazole.[6]

Acute and Chronic Eosinophilic Pneumonia

AEP is an acute, often debilitating eosinophilic inflammatory syndrome exclusively involving the lungs and marked by pulmonary infiltrates, dyspnea progressing to frank respiratory failure, and fever. The diagnosis is dependent on eosinophils exceeding 25% of all inflammatory cells within bronchoalveolar lavage fluid. Increasing evidence suggests association between AEP and respiratory fungal infections and new-onset cigarette smoking.[7,8] AEP has also been reported following allogeneic hematopoietic stem cell transplantation in the setting of graft-versus-host disease (GVHD).[9] Prompt recognition and treatment of GVHD with steroids usually results in rapid improvement.

In contrast, CEP presents more chronically (>6 weeks' duration); although it can occur in isolation, it more frequently presents in association with autoimmune and malignant diseases. Like AEP, CEP can present with striking eosinophilic inflammation of the lung (Fig. 43.4). Granulomas are occasionally seen on biopsy, suggesting an antigen-driven, T cell–mediated process in the chronicity of the disease. Treatment, as for CEP, is centered on steroids, but unlike AEP, relapse occurs frequently after treatment discontinuation.[10]

Hypersensitivity Pneumonitis

Hypersensitivity pneumonitis (HP), also termed extrinsic allergic alveolitis, is a pleiotropic acute, subacute, or chronic pulmonary inflammatory disorder marked by cough, chest tightness, malaise, and, in acute disease, fever and chills. Wheezing is uncommon, as are allergic features such as peripheral blood eosinophilia and atopy. Subacute and chronic disease typically lack fever and malaise, but dyspnea can be debilitating, progressing to hypoxemia and death due to end-stage pulmonary fibrosis if untreated. Histologically, HP presents acutely with interstitial and alveolar neutrophilia that evolves in subacute disease as lymphocytic interstitial lung inflammation often including poorly formed granulomas. Chronic disease is marked by progressively more interstitial fibrosis replacing the lymphocytic inflammation.

HP is caused by inhalation of a variety of microbes, including thermophilic actinomycetes, fungi, and protozoans, often in combination with antigens derived from plants and animals (e.g., soybean proteins; bird dander). Uncommonly, low-molecular-weight chemicals such as isocyanates can also elicit HP. Treatment is directed at removing offending inhalants and administration of steroids and bronchodilators.[11,12]

Intrinsic Eosinophilic Syndromes

Hypereosinophilic Syndrome

This multisystem disease is marked by the massive accumulation of eosinophils in many tissues and almost universally involves the lungs. Expansion of Th2 cells and local and systemic release of IL-4 and IL-5 are frequently seen. The myeloid variant of hypereosinophilic syndrome (HES) results in the clonal expansion of Th2 cells in the absence of known antigen. A deletion on chromosome 4 resulting in a *FIP1L1-PDGFRA* fusion along with a variety of other chromosomal aberrations strongly support the concept that HES is a myeloproliferative disorder involving Th2 cells, although aberrant secretion of IL-5 by both solid and liquid tumors can produce similar syndromes. The presence of $CD3^{-}CD4^{+}$ T cells is the hallmark of the lymphocytic variant of HES; the molecular mechanisms of this variant are largely unknown. When neither a mutation nor aberrant T cells can account for the aberrant eosinophilia, as is seen in approximately 75% of all cases, HES is considered idiopathic. Many organs can be affected, resulting in dysfunction or failure of the gastrointestinal tract, skeletal muscles (leading potentially to respiratory failure), endomyocardial fibrosis, myocarditis, and congestive heart failure. Pulmonary involvement manifests as obstructive

airway disease, pulmonary edema, or pulmonary emboli due to a hypercoagulability. Diagnosis is based on discovery of peripheral blood eosinophilia in the setting of a multisystem disorder, with evidence of aberrant Th2 responses or elevated IL-5 secretion, a defined mutation in the case of the myeloid variant of HES, and the presence of $CD3^{-}CD4^{+}$ T cells in the lymphocytic variant. The most effective therapy for HES is based on tyrosine kinase inhibition using agents such as imatinib mesylate for myeloid variant HES. For lymphocytic and idiopathic variants of HES, the goal is to reduce peripheral eosinophilia with steroids, hydroxyurea, or anti–IL-5 antibodies (see Table 43.1).[13]

Eosinophilic Granulomatosis and Polyangiitis

Eosinophilic granulomatosis and polyangiitis (EGPA), also termed Churg-Strauss syndrome (CSS), is an idiopathic necrotizing vasculitis of medium- and small-caliber vessels characterized by airway obstruction and eosinophilia. The disease has an autoimmune nature, with circulating antimyeloperoxidase and antineutrophil cytoplasmic antibodies (p-ANCA) in 60% to 70% of affected individuals. Because EGPA is usually seen in patients with a history of asthma and allergies and the prominent pathologic feature is necrotizing vascular and tissue granulomas, the term "allergic granulomatosis and angiitis" is used synonymously. Reports linking EGPA with the LT inhibitors zafirlukast and montelukast associated with steroid withdrawal suggest that these agents unmask preexisting EGPA rather than directly causing the disorder. Similar observations have been made with omalizumab treatment. The vasculitis of CSS can affect the sinuses, central and peripheral nervous systems, gastrointestinal tract, kidneys, and heart. Treatment of CSS is based on reinstituting systemic steroids, leading to disease resolution in most patients. Severe steroid-resistant disease may require cyclophosphamide and other immunosuppressants.[14]

IMMUNOLOGIC MECHANISMS OF ALLERGIC AIRWAY DISEASE

The immune mechanisms of disease proposed by Gell and Coombs in the 1960s remain essential to understanding the pathogenesis of allergic airway diseases. Although such mechanisms probably operate in all allergic diseases, their relative importance depends on whether the disease process predominantly affects the upper or lower respiratory tract (Fig. 43.5).

KEY CONCEPTS

Immunopathogenesis of Allergic Airway Disease

- Gell and Coombs type 1 and type 4 hypersensitivity mechanisms contribute to disease expression, especially airway obstruction.
- Innate immune pathways involving interleukin (IL)-33, thymic stromal lymphopoietin (TSLP), IL-25, and complement proteins critically contribute to the development of allergic airway inflammation.
- Environmental agents that are now established as important initiation factors for allergic airway inflammation include proteases, and endotoxin derived from fungi and bacteria.
- Fungi and viruses are established infectious causes of allergic disease of the upper and lower airways.

Type I (Immediate) Hypersensitivity

This form of hypersensitivity involves the activation of basophils and mast cells that release histamine and other inflammatory mediators. Antigen recognition is via IgE antibodies binding to high-affinity receptors (FcεRI) to arm effector cells. Th2 cells

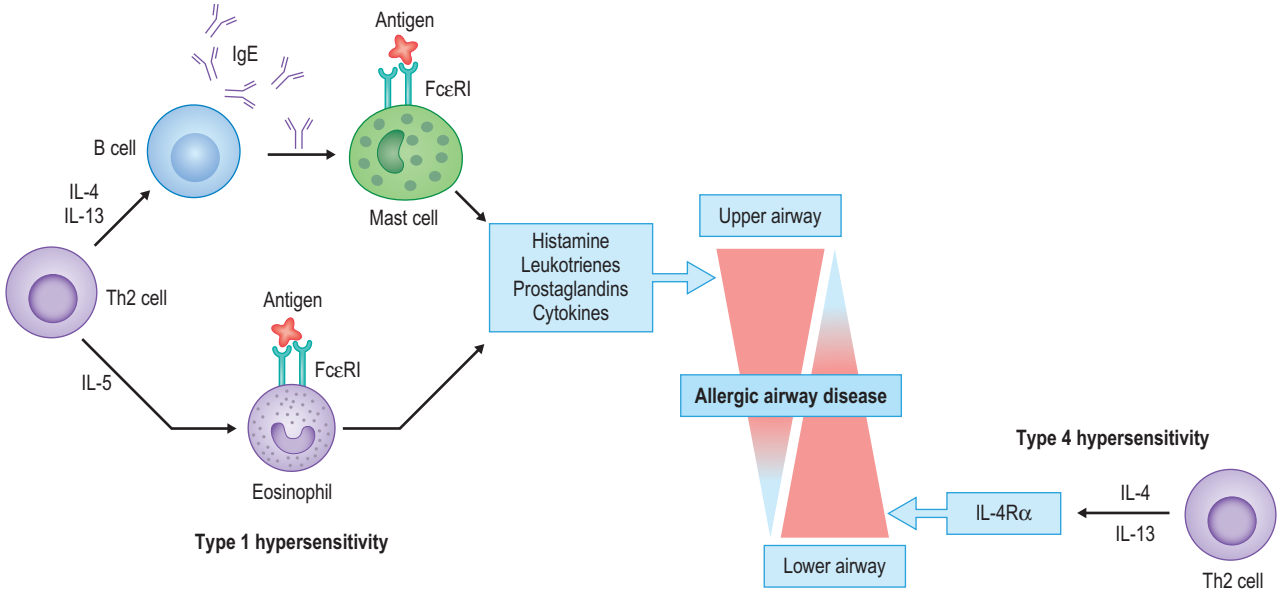


FIG. 43.5 Differential Importance of Allergic Immune Mechanisms According to Airway Level. Type I hypersensitivity *(left)*, mediated by immunoglobulin E *(IgE)*-primed mast cells and eosinophils, is ultimately driven indirectly by the cytokines secreted by T helper 2 *(Th2)* cells. In contrast, type IV hypersensitivity *(right)* is mediated directly by Th2 cytokines, especially interleukin *(IL)*-4 and IL-13, acting through a similar receptor that includes IL-4Rα. Both immune mechanisms are important to the expression of allergic disease at all airway levels, but type I hypersensitivity predominates in the upper airway, whereas type IV hypersensitivity likely assumes a more important role in the lower airway.

coordinate both production of IgE antibodies and activation and recruitment of allergic effector cells to the airway. Antigen-specific IgE bound to the surface of mast cells and basophils is cross-linked upon exposure to relevant antigens, causing cellular activation and release of preformed mediators of inflammation such as histamine, proteases, LTs, numerous cytokines, and other substances. IL-4 released primarily by Th2 cells is an important regulator of type I hypersensitivity reactions because it is required for B-cell maturation and IgE secretion.

There is evidence that IL-4 and IL-13 can mediate distinct effector phenotypes in airway and tissue macrophages and dendritic cells. Two major effector macrophage subtypes include conventionally activated (M1) macrophages that arise under the predominant influence of type I cytokines, especially interferon (IFN)-γ, and alternatively activated macrophages (M2) that arise under the influence of IL-4 and IL-13 in the relative absence of IFN-γ. M2 macrophages express a distinct gene profile including high-level expression *of arginase 1, Ym1, Fizz1 (RELM)*, and *PD-L2*. Current evidence suggests that M2 macrophages promote allergic responses and that redirecting them to the M1 phenotype might be useful.[15] Type I hypersensitivity is most prominently activated during anaphylaxis and AR.

Cell-Mediated Features of Immediate Hypersensitivity

Airway obstruction in allergen-sensitized asthma evolves over several hours after allergen exposure and is seen in two distinct phases. The early phase response is marked by airway constriction that becomes maximal approximately 30 minutes after allergen exposure and is fully relieved after approximately 2 hours (Fig. 43.6). Approximately 50% of asthmatic subjects tested develop a late-phase response in which airway obstruction again develops 4 to 6 hours after allergen exposure. Late-phase reactions are linked to airways infiltration with Th2 cells and eosinophils. The accompanying bronchoconstriction is reversible with bronchodilating agents.[16]

Airway hyperresponsiveness is neurologically mediated through parasympathetic nerves such as the vagus and is fully reversible with bronchodilating agents that either interrupt muscarinic parasympathetic signaling directly (e.g., ipratropium bromide) or activate receptors (e.g., β_2-adrenergic) that antagonize muscarinic bronchoconstrictive pathways. Airway hyperresponsiveness is recognized as episodic bronchoconstriction that is reversed with bronchodilating agents. Late-phase responses after antigen challenge therefore represent a form of airway hyperresponsiveness.

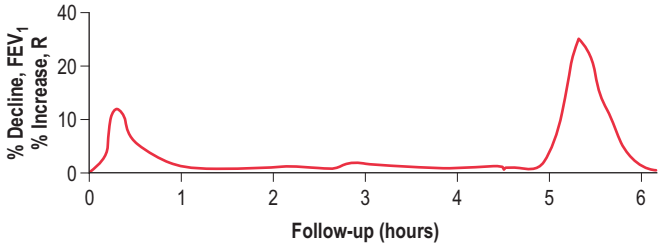


FIG. 43.6 Early- and Late-Phase Airway Changes Following Allergen Challenge. This graph shows the two phases of bronchoconstriction typically seen after allergen inhalation in sensitized, asthmatic subjects. Within 20 to 30 minutes after allergen inhalation, the first (early) phase of bronchoconstriction is seen, as assessed by either a decline in forced expiratory volume in 1 second *(FEV_1)* or increase in airway resistance *(R)*. After quickly subsiding, approximately 4 to 6 hours (h) later, a second (late) phase of bronchoconstriction occurs.

Studies from many species have demonstrated that airway hyperresponsiveness in the setting of allergic inflammation is critically dependent on Th2 cells and type 2 ILCs (ILC2) that have specifically been recruited to the lung. Moreover, it is now clear that IL-13 is the major Th2 and ILC2 cytokine that mediates airway hyperresponsiveness by acting on constitutive airway cells, such as airway smooth muscle cells, that express the IL-13 receptor. However, IL-13 does not directly induce bronchoconstriction. Bronchoconstriction in asthmatics is triggered by diverse exogenous factors in addition to allergens (e.g., altered temperature and humidity, pungent odors, irritating aerosols) and endogenous stimuli (e.g., extreme emotional states) with little apparent connection to immunologic mechanisms. Thus, rather than directly mediating airway obstruction, IL-13 establishes the basis for responding broadly to diverse agents with neurologically mediated bronchoconstriction.[17]

A second and more insidious form of airway obstruction is physical obstruction of the airways due to mucus and fibrin clots accumulating in the airways as tenacious plugs, a phenomenon currently termed plastic bronchitis. Airway obstruction due to plastic bronchitis is not immediately reversible with bronchodilators or other pharmacologic agents and is consequently the major cause of asthma death due to asphyxiation.[18]

Finally, IL-4 and IL-13 further coordinate the recruitment and retention of allergic effector cells to airway epithelium and submucosa, which facilitates rapid responses to inhaled allergens. Acting through a similar receptor that includes the alpha chain of the IL-4 receptor (IL-4Rα), IL-4, and IL-13 signal constitutive airway cells such as airway epithelial cells to induce secretion of a restricted repertoire of chemoattractants that promote the immigration of allergic cells expressing specific cognate receptors from the lung and airway microcirculation (Table 43.3).[19]

Contributing Immune Mechanisms in Allergic Airway Disease

IL-5 contributes to both immediate and cell-mediated hypersensitivity reactions by promoting the growth and differentiation of eosinophils. Although widely viewed as pathogenic and contributing to expression of allergic airway diseases, more recent studies indicate that eosinophils are important in tissue remodeling and in controlling allergic inflammation induced by pathogens such as fungi.

In addition to Th2 cells, mast cells, and eosinophils, ILCs including ILC2, natural killer cells (NK cells; a type of ILC1) and γδ T cells may also contribute to allergic disease expression through their ability to rapidly secrete type 2 and other cytokines.[20]

TABLE 43.3 Chemoattractants Linked to the Recruitment of Allergic Inflammatory Cells

Chemokine	Receptor
CCL1	CCR8
CCL11	CCR3
CCL17, CCL22	CCR4
CX_3CL1	CX_3CR1
Prostaglandin D_2	CRTh2
Leukotriene B_4	BLT1

Numerous additional mediators contribute to the expression of allergic disease. The complement system is especially important; complement proteins C3a and C5a, the major anaphylatoxins, are both essential for the expression and regulation of experimental asthma. C3a signaling through the C3a receptor (C3aR) is required for robust Th2 responses, allergic inflammation, and airway hyper-responsiveness in response to airway allergens. In contrast, C5a, which can signal through two receptors, C5aR and C5L2, appears to inhibit Th2 responses, perhaps acting as a physiologic antagonist of the allergic disease–promoting activity of C3a.[21]

Inflammatory lipid mediators important to allergic airway disease include the LTs and prostaglandins (PGs). The cysteinyl LTs (CysLTC4, CysLTD4, and CysLTE4) signal through at least two receptors to mediate similar allergic features as IL-13, including airway inflammation and hyperresponsiveness. Full expression of experimental allergic disease, in fact, appears to require the concomitant expression of both IL-13 and the CysLTs. However, IL-13 appears to be the dominant allergic mediator, perhaps partly accounting for why LT antagonism alone is less effective than inhaled steroids in treating asthma. Noncysteinyl LTs such as LTB_4 also contribute to the expression of allergic airway inflammation by controlling recruitment of allergic effector cells, including Th2 cells.[22]

Environmental Factors and Allergic Disease Initiation

Formerly, asthma, rhinosinusitis, and rhinitis were believed to result from aberrant immunologic responses to innocuous inhaled antigens, including allergens derived from dust mites, cats, dogs, and plants. More recently, a major exception to the intrinsically innocuous nature of allergens was recognized. Specifically, the allergenic fungi such as *Aspergillus* spp. and other molds, but also yeasts such as *Candida albicans*, are now recognized as ubiquitous respirable agents that can actively infect and grow within the respiratory tract, producing airway mycosis.[23]

A fungal infectious basis for allergic airway disease is demonstrated by (1) the high rate of isolation of filamentous fungi from the airways in CRS, especially AFRS, asthma and ABPA; (2) the almost universal presence of fungus-specific Th2 immunity and fungi isolated from airway mucus in subjects with allergic CRS (Fig. 43.7); (3) the efficacy of antifungal antibiotics when given to fungus-sensitized asthmatics; and (4) experimental validation that filamentous fungi are infectious for the mouse airway and produce allergic airway disease comparable with asthma. Moreover, fungal airway infection is sufficient to induce atopy to innocuous bystander antigens, suggesting that fungal infection could underlie both atopy and respiratory tract allergic disease.[23]

Respiratory viruses and particulate matter are also prominently linked to allergic airway disease. Approximately 70% to 80% of children and adults test positive for human rhinovirus (HRV) during acute disease exacerbations. Other respiratory viruses likely contribute to allergic disease pathogenesis, although the mechanisms remain obscure.[24] Particulate matter such as tobacco smoke, diesel exhaust particles, and other forms of smoke are strongly linked to asthma exacerbation and enhanced atopy, as is exposure to ozone (O_3). A common link tying these various forms of air pollution to allergic disease may be induction of oxidative stress, ultimately leading to enhanced activation of nuclear factor kappa B (NF-κB), enhanced Th2 cytokine release, and increased allergic inflammation.[25]

Research from experimental systems has shed additional light on how allergens initiate inflammation and disease. Although structural features of allergens are not linked to their allergic character, a common biochemical feature strongly associated with allergic disease is protease activity. Proteases as single molecules are as effective as any complex allergen or fungal infection in inducing allergic airway inflammation and hyperresponsiveness when administered to rodents or inhaled by humans. Household proteases that retain enzymatic activity are derived largely from fungi, suggesting again that airway infection due to these organisms, resulting in in situ protease production, may be an important mechanism underlying allergic disease induction. Irritation of the airways through viral infection, ozone exposure and other mechanisms further increases endogenous airway protease activity, especially through induction of thrombin activity.[23]

Analyses of diverse allergenic proteases suggest that they initiate a complex, airway epithelial-centered mechanism in which the cytokines thymic stromal lymphopoietin (TSLP), IL-33, and IL-25 are induced and lead to robust allergic responses. In part, this sequence is initiated by the action of exogenous and endogenous proteinases on fibrinogen. Cleavage of fibrinogen by

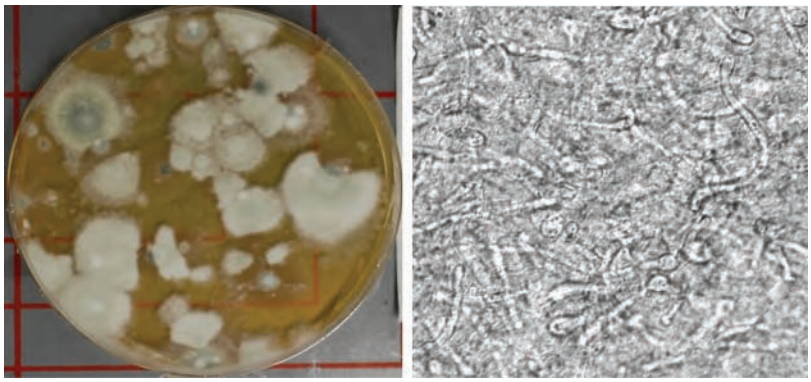


FIG. 43.7 Respiratory Tract Fungi in Allergic Airway Disease. (A) Representative Sabouraud's plate fungal cultures of sinus lavage fluid from a representative subject with chronic rhinosinusitis and asthma. (B) Photomicrograph of unstained sinus lavage mucus from a representative patient with allergic fungal rhinosinusitis showing extensive hyphal network (400× original magnification). (Modified from Porter, P.C., Lim, D.J., Maskatia, Z.K., et al., Airway surface mycosis in chronic T_H2-associated airway disease, *J Allergy Clin Immunol*. 2014;134:327.)

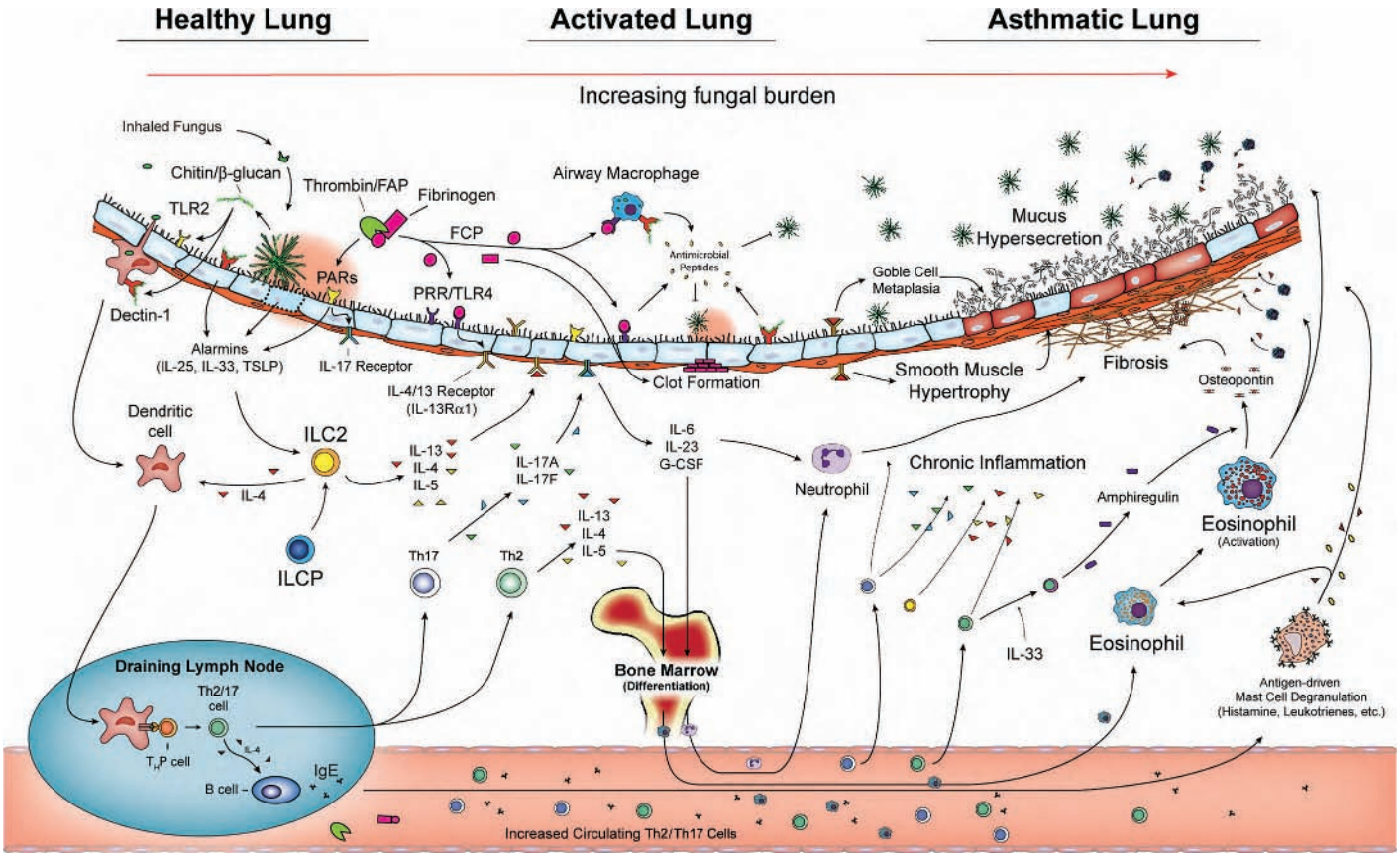


FIG. 43.8 Immune Pathogenesis of Asthma Due to Airway Mycosis. All initially healthy individuals are exposed to inhaled fungi that can produce at least transient airway mycosis. Depending on host and environmental factors, the airway mycosis can lead to airway inflammation that characterizes the activated, but not necessarily diseased, lung, and eventually progresses to clinically significant asthma with attendant immune and structural sequelae. The initial wave of airway mycosis triggers innate allergic responses by activating pattern recognition receptors (*PRRs*) including Toll-like receptor 2 *(TLR2)* and Dectin-1 expressed on airway epithelial and immune cells that detect fungal sugars such as chitin and β-glucan. Epithelial cells then secrete cytokines such as interleukin-25 *(IL-25), IL-33*, and thymic stromal lymphopoietin (TSLP) into the subepithelial space to promote the development of innate lymphoid cells type 2 *(ILC2)* and dendritic cells that migrate to draining lymph nodes and promote development of T helper type 2 (*Th2*) and *Th17* cells and IgE-secreting B cells. At the same time, endogenous proteinases, and potentially fungal proteinases, promote the breakdown of fibrinogen into fibrinogen cleavage products *(FCP)* that signal through TLR4 to initiate airway hyperreactivity, eosinophilia, and mucous hypersecretion. Proteinases, fungal and endogenous, may also interact with protease-activated receptor 2 (PAR2) to either promote or attenuate asthma-like disease. With greater degrees of airway mycosis, the low-grade, innate allergic inflammation that characterizes healthy and activated lungs converts to a chronic lung (asthma) phenotype characterized by more persistent and intense allergic inflammation that includes prominent T helper 2 (Th2) cells and Th17 cells, eosinophils (under the direction of *IL-5*), and mast cells (under the direction of *IL-4* and immunoglobulin E *[IgE]*), and neutrophils (under the direction of *IL-17A* and related cytokines), which promote severe allergic inflammation, airway hyperreactivity, and mucous hypersecretion that characterize various asthma endotypes. Most of these immune elements, including highly activated neutrophils, eosinophils, and antimicrobial peptide-secreting epithelial cells, have an antifungal role. Nonetheless, failure to resolve airway mycosis can result in permanent airway damage including airway fibrosis and bronchiectasis. (Reproduced from Li, E., Knight, J.M., Wu, Y., et al., Airway mycosis in allergic airway disease *Adv Immunol.* 2019;142:106.)

proteases yields fibrinogen-cleaved products (FCPs) that signal through Toll-like receptor 4 (TLR4) to initiate both antifungal responses and innate allergic inflammation, including ILC2 responses (Fig. 43.8).[23]

NOVEL PATHWAYS TO THERAPY IN INFLAMMATORY AIRWAY DISEASE

Intensive research conducted over several decades has markedly improved understanding of the immune and environmental basis of allergic and non-allergic airway diseases. However, currently available therapies for these diseases have failed to match this level of sophistication. Allergic disease is immunologically complex and likely to involve multiple hypersensitivity mechanisms operating in parallel. Recent studies also show that factors other than allergens likely critically influence the airway immune response to inhaled antigens and particulate matter. These adjuvant factors include cell wall products of bacteria and fungi, secreted factors such as proteases, and endogenous factors that result from cells damaged through exposure to allergens and smoke. Evidence further suggests that asthma

and rhinosinusitis may be linked to respiratory tract infections involving viruses, especially HRV, and fungi. Future therapies are therefore likely to focus on both the endogenous factors that coordinate allergic inflammation (e.g., cytokines) and etiologic environmental factors.

Once considered primarily a physical barrier, the respiratory epithelium is now known to actively coordinate immunity to allergens and other environmental challenges. Through release of IL-33, IL-25, and/or TSLP, respiratory epithelial cells can initiate and coordinate the innate and adaptive type 2 immune response (see Fig. 43.8). In addition, respiratory epithelial cells produce several chemokines in response to environmental triggers recruiting dendritic cells, ILC2s, basophils, eosinophils, and mast cell progenitors.[26] A number of triggers listed earlier as well as trauma to airway cells can activate the release of these type 2 cytokines and chemokines.

The epithelial cell cytokines, IL-25, IL-33, and TSLP share many overlapping functions supporting a local type 2 inflammatory response but also possess unique functions. For example, IL-33 and IL-25 activate different types of ILCs, with IL-33 primarily activating ILC2 while IL-25 preferentially stimulates a novel population of type 2 innate cells. TSLP is associated more with activation of the adaptive type 2 inflammatory response via dendritic cells and uniquely supports the hematopoiesis of basophils. Together, these epithelial cell–derived cytokines orchestrate a robust innate and adaptive type 2 inflammatory response in addition to initiating repairs needed for the epithelial barrier (see Fig. 43.8).[26]

ON THE HORIZON

- Clinical trials focused on interrupting key innate immunologic pathways (*i.e.*, IL-33, thymic stromal lymphopoietin [TSLP]) in allergic disease.
- Focused research on the importance of infectious pathogens, especially fungi, in asthma and chronic rhinosinusitis.
- Accelerated development of biomarkers for asthma, chronic rhinosinusitis, and possibly other allergic airway disorders.
- Emergence of clinical trials focusing on antiinflammatory small molecules.

Recent studies have verified the importance of Th2- and ILC-derived cytokines, including IL-5 and IL-13, to the pathogenesis of human allergic disease. The anti–IL-5 antibody mepolizumab and anti–IL-4/IL-13 receptor (dupilumab) antibodies were recently approved for use in eosinophilic asthma and appear to offer substantial benefit (see Table 43.1). In addition to defining diseases that are most likely to benefit from them, a major challenge with all biologic agents is overcoming their prohibitive costs. Therefore the therapy of allergic airway diseases is likely to evolve toward the development of small, relatively easily manufactured and relatively inexpensive molecules that antagonize disease-related pathways established through biologics as essential to disease expression.

The central role that proteases and the proteolytic activation of complement (e.g., C3) and coagulation-related inflammatory pathways (e.g., fibrinogen) play in allergic diseases support targeting of these molecules as part of future clinical trials in allergic airway disease. However, the discovery that severe, life-threatening forms of asthma and severe CRS are ultimately due to fungal infections suggests that novel anti-inflammatory therapies must be combined with antifungal approaches if patients are to derive maximum benefit. Although several positive trials involving antifungals for allergic airway disease have been reported, future studies are likely to involve improved clinical trial designs combined with superior antifungal agents and improved means of delivery, including by aerosol.

REFERENCES

1. Fokkens W, Lund V, Mullol J. European position paper on rhinosinusitis and nasal polyps 2007. *Rhinol Suppl*. 2007;20:1–136.
2. Bauer AM, Turner JH. Personalized medicine in chronic rhinosinusitis: phenotypes, endotypes, and biomarkers. *Immunol Allergy Clin North Am*. 2020;40(2):281–293.
3. Solelhac G, Charpin D. Management of allergic rhinitis. *F1000Prime Rep*. 2014;6:94.
4. Agarwal A, Spath D, Sherris DA, et al. Therapeutic antibodies for nasal polyposis treatment: where are we headed?. *Clin Rev Allergy Immunol*. 201959:141–149.
5. Corry DB, Irvin CG. Promise and pitfalls in animal-based asthma research: building a better mousetrap. *Immunol Res*. 2006;35(3):279–294.
6. Knutsen AP, Bush RK, Demain JG, et al. Fungi and allergic lower respiratory tract diseases. *J Allergy Clin Immunol*. 2012;129(2):280–291.
7. Swartz J, Stoller JK. Acute eosinophilic pneumonia complicating *Coccidioides immitis* pneumonia: a case report and literature review. *Respiration*. 2009;77(1):102–106.
8. Uchiyama H, Suda T, Nakamura Y, et al. Alterations in smoking habits are associated with acute eosinophilic pneumonia. *Chest*. 2008;133(5):1174–1180.
9. Yoshimi M, Nannya Y, Watanabe T, et al. Acute eosinophilic pneumonia is a non-infectious lung complication after allogeneic hematopoietic stem cell transplantation. *Int J Hematol*. 2009;89(2):244–248.
10. Rose DM, Hrncir DE. Primary eosinophilic lung diseases. *Allergy Asthma Proc*. 2013;34(1):19–25.
11. Lacasse Y, Girard M, Cormier Y. Recent advances in hypersensitivity pneumonitis. *Chest*. 2012;142(1):208–217.
12. Sahin H, Kaproth-Joslin K, Hobbs SK. Hypersensitivity pneumonitis. *Semin Roentgenol*. 2019;54(1):37–43.
13. Curtis C, Ogbogu P. Hypereosinophilic syndrome. *Clin Rev Allergy Immunol*. 2016;50(2):240–251.
14. Gioffredi A, Maritati F, Oliva E, Buzio C. Eosinophilic granulomatosis with polyangiitis: an overview. *Front*. 2014;5:549.
15. Sica A, Mantovani A. Macrophage plasticity and polarization: in vivo veritas. *J Clin Invest*. 2012;122(3):787–795.
16. Fireman P. Understanding asthma pathophysiology. *Allergy Asthma Proc*. 2003;24(2):79–83.
17. Li E, Landers CT, Tung HY, et al. Fungi in mucoobstructive airway diseases. *Ann Am Thorac Soc*. 2018;15(suppl 3):S198–S204.
18. Molfino NA, Nannini LJ, Martelli AN, Slutsky AS. Respiratory arrest in near-fatal asthma. *N Engl J Med*. 1991;324(5):285–288.
19. Tomankova T, Kriegova E, Liu M. Chemokine receptors and their therapeutic opportunities in diseased lung: far beyond leukocyte trafficking. *Am J Physiol Lung Cell Mol Physiol*. 2015;308(7):L603–L618.
20. Zhu J. T helper 2 (Th2) cell differentiation, type 2 innate lymphoid cell (ILC2) development and regulation of interleukin-4 (IL-4) and IL-13 production. *Cytokine*. 2015;75(1):14–24.
21. Zhang X, Kohl J. A complex role for complement in allergic asthma. *Exp Rev Clin Immunol*. 2010;6(2):269–277.
22. Kanaoka Y, Boyce JA. Cysteinyl leukotrienes and their receptors; emerging concepts. *Allergy Asthma Immunol Res*. 2014;6(4):288–295.
23. Li E, Knight JM, Wu Y, et al. Airway mycosis in allergic airway disease. *Adv Immunol*. 2019;142:85–140.
24. Newcomb DC, Peebles RS Jr. Bugs and asthma: a different disease? *Proc Am Thorac Soc*. 2009;6(3):266–271.
25. Dozor AJ. The role of oxidative stress in the pathogenesis and treatment of asthma. *Ann NY Acad Sci*. 2010;1203:133–137.
26. Hammad H, Lambrecht BN. Barrier epithelial cells and the control of type 2 immunity. *Immunity*. 2015;43(1):29–40.

Mast Cells and Mast Cell Disorders

Cem Akin, Michelle Al-Hosni and Dilawar Singh Khokar

Mast cells (MCs) are granular innate immune cells that are widely distributed in tissue and are best recognized as the primary effector cell of allergic disease. While there is still much to learn about human MCs, significant advancements have been made in our understanding of human MCs including their functions independent of allergic disease as well as the origin and development of MCs and their progenitors. In this chapter, we will review the origin, development, and functions of MCs, as well as their roles in human health and disease. Basophils are circulating cells containing granules staining similarly to MCs (Fig. 44.1), attributed with overlapping pathogenic roles in allergic disease, although their physiologic function is less understood.

MAST CELL BIOLOGY

Mast Cell Origin and Early Development

MCs are of hematopoietic origin. One proposed model of MC development is that hematopoietic stem cells in the bone marrow give rise to different types of progenitor cells including multipotent and common myeloid progenitors that progress to a mast cell progenitor (MCp), which exits the bone marrow and migrates to different tissue sites to complete differentiation and maturation.[1] Tissue resident MCs persist for extended periods of time and are exposed to a host of stimuli, which allows further differentiation and maturation.[1,2] MC development occurs along a myeloid pathway in distinct stages characterized by various cell surface markers as well as variable expression in transcription factors. In humans, the MCp population exists as CD34/CD117/CD13 positive cells.[3]

Although multiple cytokines influence MC growth and development, stem cell factor (SCF) is considered the most important.[4] SCF is the ligand for the c-kit receptor (CD117), which is a transmembrane receptor with intrinsic tyrosine kinase activity.[4] SCF is best known for facilitating MC survival that is thought to occur via its inhibition of apoptotic proteins.[4] It has also been linked to other MC functions including proliferation, differentiation, maturation, cellular adhesion, and MC chemotaxis.[4] In addition, SCF has been implicated in the augmentation of mast cell activation (MCA) via different receptors including FcεR1 as well as TLR4, and c-kit signaling enhances MC degranulation via calcium influx and induction of changes in transcriptional activity.[4] Gain of function mutations in the c-kit receptor gene are central to the pathogenesis of mastocytosis.[5] Mice with hypofunctioning KIT/SCF pathway have a profound MC deficiency. These findings demonstrate that SCF/KIT is a key signaling pathway for MCs, although there are likely additional factors contributing to progenitor development.[6] *In vitro*, MCs can be cultured from $CD34^+$ sorted progenitors in hematopoietic stem cell media supplemented with SCF alone.[4] It has been generally thought that MCs share a common progenitor with myeloid cells, although a MC specific unique progenitor directly branching from the multipotential progenitor cell has also been demonstrated.[7]

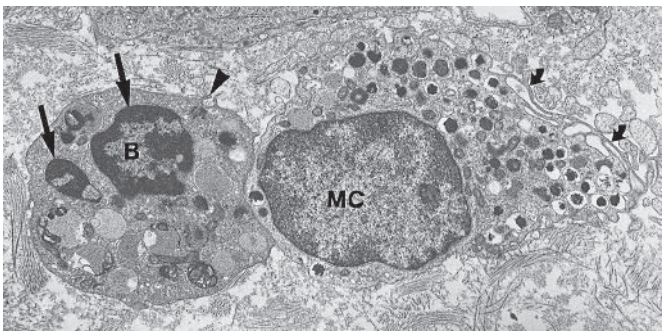


FIG. 44.1 Ultrastructures of Mast Cell and Basophil. A basophil *(B)* adjacent to a mast cell *(MC)* in the ileal submucosa of a patient with Crohn disease. The basophil exhibits a bilobed nucleus *(solid arrows)* whose chromatin is strikingly condensed beneath the nuclear membrane. The basophil surface is relatively smooth with a few blunt processes *(arrowhead)*. The mast cell nucleus is larger and its chromatin less condensed than that of the basophil. The mast cell's granules are smaller, more numerous, and more variable in shape and content than those of the basophil. The mast cell surface has numerous elongated, thin folds *(curved arrows)* (original magnification approximately ×9000). (From Dvorak AM, Monahan RA, Osage JE, Dickersin GR. Crohn's disease: transmission electron microscopical studies. *Hum Pathol*. 1980;11:606–619, with permission from Ann M. Dvorak.)

KEY CONCEPTS

Mast Cell Origin and Development

- MCs originate from the bone marrow and belong to a myeloid lineage.
- MCps migrate to different tissues where they complete differentiation/maturation and may reside for long periods of time.
- Stem cell factor and c-KIT signaling are central to MC survival and function as well as development of mastocytosis.
- MC progenitors reside in a CD34/CD117/CD13 positive progenitor pool.

Mast Cell Homing

Tissue-specific homing of MCps is a critical portion of MC development as it enables maturation and differentiation in unique microenvironments. Integrins and cell adhesion mole-

cules help mediate the process by which MCps localize to different tissues. Human MCps express α4β1 integrin and this regulates adhesion to activated endothelial cells, which in turn helps facilitate migration of MCps from circulation to tissue. Integrin expression also varies by target tissue. For example, mucosal MCs require β7 to mediate homing to the gut; the integrin α4β7 mediates homing to the gut by binding MadCAM1 (mucosal address in cell adhesion molecule 1) and VCAM-1 (vascular cell adhesion molecule 1). MCps express a number of different chemokine receptors including CCR1, CXCR2, CCR3, CXCR4, and CCR5.[8] In humans, expression of CCR1 and CCR5 has been linked to MCp retention in the bone marrow as compared to circulating precursors.[8] CXCR2 is involved in the homing of MC precursors to the intestine.

Mast Cell Heterogeneity

There is substantial heterogeneity among MCs in regard to protease content, receptor expression, mediator content, and responsiveness to immunologic and nonimmunologic stimuli. MC heterogeneity can occur between different tissues and within the same tissue. Historically, MCs have been classified based on their protease content with two major subtypes: MCt (tryptase only) and MCtc (tryptase and chymase). Classically, MCts are found predominantly at mucosal surfaces while MCtcs are found in connective tissue. Recent studies suggest that MCs are far more heterogeneous than their protease content would suggest. Transcriptional analysis of MCs isolated from different tissue sites has revealed significant variance in genes encoding serine proteases, metalloproteases, adhesion molecules, mediator production, and cytokine production.[9,10] Differential transcriptional expression has also been noted to occur in MCA with different stimuli such as in immunoglobulin E (IgE) cross-linking versus interleukin (IL)-33 stimulation.[11] Different disease states are also associated with specific changes in MCs. For example, in eczema, MCs in the skin display increased substance P and neurokinin receptor expression as compared to MCs in the gastrointestinal (GI) tract where chymase expression is increased in response to parasitic infection. In murine models of food allergy, activation of connective tissue MCs was linked to systemic anaphylaxis and activation mucosal mast cell protease-1 (MMCP-1) MCs was linked to GI-predominant reactions.[12] Collectively, these findings support the concept that MCs express different gene profiles in response to environmental stimuli resulting in inter- and intra-tissue–specific variability that contribute to both MC-mediated disease and homeostatic function.

KEY CONCEPTS

Mast Cell Homing and Heterogeneity

- MC homing to tissue is mediated by adhesion molecules such as integrins and also by chemokines. For homing to the gut, α4β7 binding to MadCAM1 is critical.
- There is substantial heterogeneity among MC populations with variances in protease content, receptor expression, mediator content, and responsiveness to stimuli.
- MC heterogeneity and phenotype play important roles in both normal MC function and disease pathogenesis.

MAST CELL ACTIVATION

A characteristic feature of MCs is their ability to be activated by multiple stimuli in complex environments. This functionality is central to their role in homeostasis and a number of disease states. Table 44.1 summarizes key ligands and receptors in MCA. The most extensively studied MC receptor is FcεR1 (high affinity IgE receptor). FcεR1 is a tetrameric receptor (composed of α, β, and 2 γ subunits) lacking intrinsic kinase activity, where α subunit binds IgE, and γ subunit mediates signal transduction. Crosslinking of FcεR1 via IgE-antigen complexes results in receptor aggregation that initiates a tyrosine phosphorylation cascade stimulated by lyn/syk, fyn, and hck kinases leading to a multiphasic cell response.[4] This response is characterized by MC degranulation with immediate release of preformed mediators (such as histamine), followed by rapid production of lipid mediators (prostaglandins such as PGD_2 and cysteinyl leukotrienes [LT]), and cytokine production.[4]

MCA response coincides with the biphasic early and late phase symptoms of allergic diseases such as allergic rhinitis. Activation of FcεR1 by IgE also has effects independent of degranulation including promotion of MC survival and increasing cell expression of FcεR1.[4] Whether or not IgE *independent* of antigen is able to activate MCs has been an area of controversy. Monomeric IgE bound to FcεR1 without antigen has been shown to trigger MC degranulation, amplify migratory response, and enhance MC survival.[13–16] Histamine-releasing-factor (HRF) has also been implicated in antigen-independent IgE activation of MCs by facilitating crosslinking of IgE bound to FcεR1.[17] These findings support a role of MCA in allergic disease independent of sensitization status and also highlights the importance of IgE-FcεR1 binding interaction as a therapeutic target for MC-mediated processes.

Non–IgE-mediated MCA is an area of active investigation as it provides insight into disease processes where symptoms mediated by MC degranulation are present, but there is no clear IgE-mediated sensitization. MCA by a variety of different triggers has been described including IgG (through FcγR1, RIIa, and RIII), complement receptors, PAMPs (pathogen associated molecular patterns) via toll-like receptors (TLRs), MRGPRX2 (Mas-related G-protein-coupled receptor-X2) through various ligands, alarmins, chemokines, and cytokines.[4,18] IgG-mediated MCA occurs via IgG binding to FcγR and provides a mechanism by which MCA may occur in autoimmune diseases as many autoimmune conditions are characterized by IgG immune complex formation. Likewise, complement activation in autoimmune and infectious disease may contribute to MCA via C3a and C5a receptors. MRGPRX2 is a G protein coupled receptor thought to be involved in several MC functions including innate immune responses, wound repair, and induction of MC degranulation in a manner independent of IgE.[19] *In vitro* experimentation using small molecule drugs such as fluoroquinolone antibiotics and neuromuscular blocking agents have demonstrated that such medications are capable of eliciting MC degranulation via MRGPRX2.[20] This would suggest that a subset of immediate drug hypersensitivity reactions may be non–IgE-mediated and potentially explains how a patient without prior exposure or sensitization to a drug could display an allergic reaction to that drug on first exposure. Recently, there has been an increased focus on MC neural interactions given the close proximity of MCs to nerve fibers in peripheral tissue. Many neuropeptides, such as substance P (also acting through MRGPRX2 receptor), have been demonstrated to cause MC degranulation, and neurologic symptoms such as itch and sometimes pain are often present in MC mediated processes.[18]

TABLE 44.1 Mast Cell Activation

Ligand/Receptor	Biologic Effect	Clinical Significance
SCF/c-kit	Enhances MC survival, migration, and adhesion as well as stimulation of IL-6 production.	Mutations in c-kit are associated with mastocytosis, and kit inhibitors are emerging as therapy for MC disease.
IgE/FcεR1	Causes MC degranulation, lipid mediator, and cytokine production.	Mechanism of IgE-mediated allergic disease.
IgG/FcγR1, RIIa and RIII	Causes MC degranulation, proinflammatory cytokine production, and lipid mediator production.	Contribution of MC to immune response to pathogens; possible role of MC activation in autoimmune disease.
C3a, C5a/C3aR, C5aR	Enhances IgG/FcγR-induced MC degranulation; promotes MC migration, adhesion, and mediator production.	Contribution of MC to immune response to pathogens; possible role of MC activation in autoimmune disease.
PAMPs/TLR-1, 2, 4, 5, 6	Increases cytokine production from MCs; depending on context can inhibit IgE/FcεR1 activation.	Contribution of MC to immune response to pathogens; possible role in regulation of IgE allergic responses.
Neuropeptides and drugs/ MRGPRX2	Cause MC degranulation in response to different stimuli including substance P and small molecule drugs	Possible mechanism of pseudo-allergic reactions and of allergic reactions upon first exposure to a drug.
IL-4/IL-4R	Enhances FcεR1 expression, increases release of MC mediators, and enhances MC production of other Th2 cytokines.	Possible mechanism of anti-IL4 therapies in allergic disease.
IL-33/ST2	Promotes MC activation, survival, maturation, adhesion, and cytokine production.	IL-33 is an alarmin and is released in response to cellular damage or inflammation, thus highlighting the role of MCs in immune responses to tissue damage and inflammatory diseases.
TSLP/TSLP-R	Promotes MC survival, but with little MC activation or stimulation of cytokine production.	TSLP is released by barrier/epithelial cells in response to a number of stimuli and is one mechanism by which MCs interact with their local environment.

IL, Interleukin; *MCs*, mast cells; *MRGPRX*, Mas related G protein coupled receptor X; *PAMP*, pathogen associated molecular pattern; *SCF*, stem cell factor; *TSLP*, Thymic stromal lymphopoietin.

MCA is regulated by inhibitory receptors containing tyrosine-based inhibitory motifs (ITIMs) intracellularly. Siglec 6 and 8, CD300a, and FcγRIIb are examples of such inhibitory receptors in humans.

MAST CELL MEDIATORS

MC and basophils express a myriad of potent biologically active mediators. These mediators are either stored preformed in cytoplasmic secretory granules or synthesized *de novo* upon cellular activation and help to orchestrate various functions in inflammation and host defense (Table 44.2).

Preformed Mediators

Preformed mediators, such as histamine, proteases (*e.g.*, tryptase, chymase, and carboxypeptidase A), and proteoglycans (*e.g.*, heparin) are stored in cytoplasmic granules and rapidly released within seconds to minutes upon cell activation. The specific profile of preformed mediators is influenced by MC subtype and the surrounding microenvironment. Degranulation can occur by several mechanisms, generally involving the fusion of membrane granules with the plasma membrane, resulting in the release of granular contents into the extracellular environment. Similar to lysosomes, MC granules are stabilized by a low pH and contain lysosomal enzymes. One such enzyme is β-hexosaminidase, found in all MC subtypes. Quantification of released β-hexosaminidase activity has been used as a measure of MC degranulation *in vitro*.[2] Uniquely, MCs are able to re-granulate and therefore remain functional following activation and subsequent degranulation.[2] Histamine, tryptase and proteoglycans are discussed more detail below due to their clinical significance.

Histamine

Histamine is derived by the decarboxylation of the amino acid histidine. The main source of histamine in humans are MCs and, to a lesser extent, basophils.[21] However, histamine is also produced by neurons, enterochromaffin-like cells in the stomach, and bacteria including those in the gut.[21] Circulating histamine interacts with target cells via four types of receptors (H1-4) and ultimately contributes to the increased vascular permeability, vasodilation, smooth muscle contraction, bronchoconstriction, mucus production, increased heart rate and cardiac output, gastric acid secretion, and pruritus associated with allergic and inflammatory reactions.[21] In clinical practice, H1 receptor antagonists (*e.g.*, cetirizine, levocetirizine, fexofenadine, loratadine, desloratadine) are frequently used in the management of allergic diseases such as rhinitis and chronic urticaria.

Mast Cell Proteases

MC proteases (*e.g.*, tryptases, chymases, and carboxypeptidase A), like other preformed mediators, are stored as active enzymes within secretory granules. A wide range of non–mast-cell–specific proteases can also be found within MC secretory granules including granzyme B, lysosomal cathepsins, active caspase 3, kallikreins, matrix metalloproteinase-9, and renin.

Tryptase

Tryptase isoforms. Tryptase is the most abundant preformed granular enzyme found in all human MCs and is frequently used as a marker for MC activation. Of the four isoforms of tryptase (α, β, γ, δ), β-tryptase, and to a lesser extent α-tryptase, are considered the most biologically significant.[21] The γ and δ forms of tryptase are believed to have little, if any, catalytic activity. Tryptase-encoding genes are clustered together on chromosome 16 (TPSAB1, TPSB2, TPSG1, TPSD1). Notably, α-tryptase and β1-tryptase alleles are co-allelic at the *TPSAB1* gene locus,[22] while *TPSB2* only encodes β-tryptase. In human MCs, β-tryptase undergoes conformational change from an inactive monomer to its biologically active tetrameric form comprised of identical subunits. However, the stability of the β-tryptase tetramer relies upon the presence of heparin and

TABLE 44.2 Selected Mast Cell Mediators

Mediator Ligand/Receptor	Pathophysiologic Effects	Commonly Used Targeted Therapy in Allergic Disease
Preformed		
Histamine/H1, H2, H3, H4R	Increased vascular permeability, vasodilation, smooth muscle contraction, bronchoconstriction, mucus production, increased heart rate and cardiac output, gastric acid secretion, pruritus	H1-receptor antagonists (cetirizine, levocetirizine, fexofenadine, loratadine, rupatadine); H2-receptor antagonists (ranitidine, famotidine)
Proteases		
Tryptase (α, β, γ, δ)/PAR (protease activated receptors)	Allergy, inflammation, tissue remodeling, immune response; defective in hereditary α-tryptasemia (HAT)	
Chymase	Tissue remodeling and injury; potent enzymatic conversion of angiotensin I to angiotensin II	
Carboxypeptidase A (CPA3)	Vasoprotective neutralization of endotoxin I; neutralizes sarafoxtoxin (snake venom)	
Proteoglycans		
Heparin	B-tryptase tetramer stabilizer; vascular permeability, smooth muscle contraction;	
Newly formed		
Lipid mediators		
PGD2/DP1, CRTH2 (DP2); TXA2	vasodilation (DP1), bronchoconstriction (CRTH2/DP2); basophil, eosinophil, Th2 and ILC2 cell chemoattractant (PGD2); aspirin-exacerbated respiratory disease (AERD)	Cyclooxygenase (COX) inhibitors, other nonsteroidal antiinflammatory drugs (NSAIDs), aspirin
CysLTs (LTC4, LD4, LTE4), LTB4/CysLT1R, CysLT2R, GPR99	Bacterial defense; neutrophil chemoattractant (LTB_4); bronchoconstriction (LTD4); smooth muscle contraction; increased microvascular permeability; amplified Th2 response	CysLT1R-antagonists (montelukast, zafirlukast); 5-LO inhibitor (zileuton)
PAF/PAF-R	Platelet aggregation, vasodilation, potent bronchoconstriction	Platelet-activating factor receptor (PAFR) antagonist (rupatadine); epinephrine
Cytokines		
IL-1β, IL-2, IL-3, IL-4, IL-5, IL-6, IL-8, IL-9, IL-10, IL-11, IL-13, IL-15, IL-16, IL-17A, IL-18, IL-22, IL-24, IL-25, IL-33, TNF-α, SF, NGF, TGF-β, FGF-2, VEGF, IFN-α, GM-CSF, TSLP	Cell signaling, growth, proliferation and migration	
Chemokines		
CCL5, CXCL8	Immune cell recruitment to sites of infection	Anti-IL-4, IL-13/IL-4Rα antagonist (dupilumab); anti-IL-5 (mepolizumab, reslizumab); IL-5R antagonist (benralizumab)

only occurs in pH values under 6.5. Therefore it follows that either heparin depolymerization or increasing intragranular pH may result in decreased tryptase enzyme activity. In contrast to β-tryptase stored in granules, α-tryptase is a protryptase lacking enzyme activity that is secreted constitutively outside of the cell; as such, its levels may correlate with total body MC burden. This, in turn, is useful in assessing MC burden and monitoring response to cytoreductive therapy in mastocytosis.

Pathophysiology and diagnostic function. The pathologic and physiologic effects of tryptase after MC degranulation may play a role in the development of allergy, inflammation, tissue remodeling, and the immune response. Baseline serum levels (median around 5 ng/mL, normally ranging from 1 to 11 ng/mL) reflect the constitutive secretion of immature monomeric tryptases (mainly α tryptase) by resting MCs, and are therefore reflective of total body MC burden.[21] The heparin-bound, highly active tetrameric form of β-tryptase is released by MCs only upon activation and represents the increase in serum tryptase during an allergic response.[21] In contrast, an increase in serum level of α-tryptase is not seen following allergen-induced MC degranulation.[21] In clinical practice, the measurement of serum tryptase is used to diagnose and monitor mastocytosis and disorders of MCA.[21] Currently, commercially available tryptase assays measure only total tryptase (*i.e.*, α and β). A transient and rapid increase of total serum tryptase (>2 + 1.2*baseline level) within 4 hours of an anaphylactic event is generally considered laboratory evidence of MCA.[21] In the evaluation of systemic mastocytosis, a baseline serum tryptase level of greater than 20 ng/mL is considered a minor diagnostic criterion.[21] Tryptase levels may also be useful in monitoring response to cytoreductive therapy in mastocytosis.

Hereditary α-tryptasemia. Hereditary α-tryptasemia (HAT) is a recently described autosomal dominant condition caused by increased monoallelic α-tryptase copy number at TPSAB1,[22,23] where serum tryptase levels (generally above 8 ng/mL) correlate with the copy number of the α-tryptase allele. It is estimated to be present in 6% of general population. Whether HAT is associated with a distinct clinical phenotype is a matter of active investigation. Patients with HAT appear to be more prone to severe hymenoptera allergy reactions if they also have an IgE-mediated sensitization. While it is tempting to suggest that persistent symptoms of otherwise undetermined etiology may be due to excess α-tryptase, many patients with this condition are asymptomatic and there remains no conclusive evidence associating HAT with a common set of specific symptoms or other disease states.

Proteoglycans and Heparin

Chondroitin sulfate and heparin are proteoglycans stored in preformed MC granules and render the metachromatic staining properties of the cells. In addition to its anticoagulant effect, heparin contributes to vascular permeability and smooth muscle contraction.[21] In maculopapular lesions of cutaneous mastocytosis, heparin is associated with clusters of MCs within the lesions.[24]

Newly Formed Mediators

Lipid mediators (*e.g.*, prostaglandins, LT, platelet activating factor [PAF]), cytokines, chemokines, and growth factors are produced *de novo* upon MC activation.[2]

Lipid Mediators

Lipid mediator synthesis occurs shortly after degranulation and begins with the production of arachidonic acid and lysophosphatidylcholine from membrane phosphatidylcholine by phospholipase A_2.[25] The eicosanoids, prostaglandins (PG), and LT are generated by the metabolism of arachidonic acid through the cyclooxygenase (COX) and lipoxygenase pathways, respectively.[2] PAF is synthesized from lysophosphatidylcholine.[2,25] These lipid mediators go on to participate in the regulation of vascular permeability, smooth muscle contraction, and the recruitment of immune effector cells.[2,25]

Prostaglandins

Prostaglandin synthesis via the COX pathway begins with conversion of arachidonic acid to PGH_2, the bioactive precursor of all PGs, by the COX enzyme.[2] PGD_2 synthase subsequently converts PGH_2 to PGD_2, arguably the most important PG in MCs.[2] Liberated PGD_2 can elicit vasodilation and bronchoconstriction when bound to its receptors, DP_1 and CRTH2 (DP_2), respectively.[21] A strong chemoattractant for basophils, eosinophils, Th2 cells, and group 2 innate lymphoid cells (ILC2), all of which express CRTH2 (DP_2) receptors, PGD_2 can thereby amplify type 2 inflammation by inducing cytokine generation (IL-4, IL-5, IL-13).[21] In contrast to the large quantities generated by MCs, PGD_2 is not produced by basophils.[21] PGD_2 is, however, made by other immune cells including eosinophils, Th2 cells, DCs, and by nonhematopoietic tissues such as brain, heart, lungs, and kidney.[21] Clinically, the COX pathway plays a role in aspirin-exacerbated respiratory disease (AERD), which is discussed in more detail later in this chapter.

Leukotrienes

In activated MCs and basophils, the lipoxygenase (LO) pathway of LT synthesis begins with oxidation of arachidonic acid by 5-LO and 5-LO-activating protein (FLAP) into unstable metabolites, 5-HpETE and LTA_4.[2,26] LTA_4 hydrolase converts LTA_4 to LTB_4, which undergoes conjugation by LTC_4 synthase to form LTC_4, the precursor for all cysteinyl-leukotrienes (CysLT).[2,26] LTC_4 is released by an energy-dependent mechanism involving multi-drug resistance protein (MRP-1), converted extracellularly to LTD4, a potent bronchoconstrictor, and finally to the stable metabolite LTE_4.[2,26] In addition to MCs and basophils, myeloid dendritic cells (DC), eosinophils, and macrophages also release CysLTs (LTC4, LTD4, LTE4).[21] Generally, LTs contribute to host defense against bacterial infections by influencing local vascular endothelium to promote rolling and recruitment of neutrophils and eosinophils.[2] Specifically, the small quantity of LTB_4 produced acts as a potent neutrophil chemoattractant.[2] However, despite their beneficial effects in bacterial defense, CysLTs play a significant role in allergic inflammation by inducing smooth muscle contraction and airway constriction, and increasing microvascular permeability.[21] Mouse models suggest CysLT amplifies Th2 responses through three distinct receptors ($CysLT_1R$, $CysLT_2R$, GPR99), with effects achieved at concentrations up to 1000-fold lower than histamine.[21] In clinical practice, $CysLT_1R$ antagonists (*e.g.*, montelukast, zafirlukast) and 5-LO inhibitors (*e.g.*, zileuton) are used in the management of asthma.[4] Notably, a recent study demonstrated montelukast-sensitive MC activation by LTE4 through $CysLT_1R$, inducing production of PGD_2 and thromboxane A_2 (TXA_2).[26]

Platelet-Activating Factor

PAF is synthesized *de novo* from membrane lysophosphatidylcholine by several cell types including MCs, eosinophils, platelets, neutrophils, monocytes, basophils, epithelial and endothelial cells in response to stress and other stimuli.[25] Effects attributed to PAF include platelet aggregation, vasodilation, and potent bronchoconstriction.[25] PAF is rapidly hydrolyzed and degraded to its inactive metabolite, lysoPAF, by intracellular and plasma types of the PAF acetylhydrolase (PAF-AH) enzyme.[21,25] Many of the cells that produce PAF are also targets for its bioactivity through G protein-linked receptors (PAF-r). PAF receptor binding leads to mobilization of intracellular calcium and activation of kinases, resulting in release of arachidonic acid.[25] In human lung MCs and peripheral blood-derived MCs, PAF induces histamine and PGD2 release following receptor binding.[25] In human airways, PAF is also a potent mediator of eosinophil chemotaxis.[25] Notably, PAF has no effect on skin MCs which lack PAF-r expression.[25] However, increased serum PAF levels and decreased PAF-AH have been linked to the development of allergic diseases such as rhinitis, asthma, and chronic urticaria. PAF is produced during the early and late phases of the allergic response, contributes to nasal congestion and rhinorrhea in allergic rhinitis, and is among many proinflammatory mediators involved in asthma pathogenesis.[25] PAF also plays a role in IgE-dependent and IgE-independent anaphylaxis; higher PAF levels may correlate with more severe reactions.[25] PAF has been implicated as a mediator in sepsis, atherosclerosis, and malignancy.[25]

Cytokines and Chemokines

Cytokines are small secreted proteins essential for cell signaling, growth, proliferation, and migration. Via autocrine secretion, MCs themselves can be affected by a number of cytokines including SCF, IL-4, IL-6, IL-10, transforming growth factor (TGF)-β, IL-33, and thymic stromal lymphopoietin (TSLP). The unique responses elicited influence the heterogeneity and phenotype variation seen among MC populations. The central role of SCF in MC growth and differentiation was discussed above. IL-6 is critical for human MC maturation.[4] TGF-β, classically immunosuppressive, enhances early MCp differentiation, but reduces survival of late-stage precursors. TGF-β also elicits MC chemotaxis, reduces FcεR1 expression IgE-induced cytokine production, and inhibits IL-33 production.[4] IL-33 is produced by barrier cells in response cell damage or inflammation and can activate MCs, as well as promote their survival, maturation, adhesion, and cytokine production.[4] TSLP is also expressed by barrier cells and promotes MC proliferation.[4] Much research has focused on MC-derived IL-4 as a major contributor to symptom development in allergy and asthma.[4] IL-4 elicits Th2 differentiation of naïve T cells, and amplifies type 2 inflammatory responses via a positive feedback loop between MCs, Th2, and ILC2 cells.[4] The effect of MC cytokine production in food allergy has also been suggested by a recent study demonstrating MCps in the intestinal mucosa as an important source of IL-9 and IL-13.[12] Conversely, MCs can express antiinflammatory cytokines TGF-β and IL-10.[4] As a part of the innate immune response to microbial infection, MCs generate TNF-α, IL-6, and IL-1β.[2] In addition to an array of cytokines, MCs produce chemokines, notably CCL5 and CXCL8, which help recruit immune cells to sites of infection.[5] Chemokines were previously discussed in the context of MC homing.

NORMAL MAST CELL FUNCTION

As previously discussed, MC development, migration, and survival are influenced by growth factors, most notably SCF.[4] In healthy individuals, tissue-resident MCs are maintained in constant numbers at baseline and their numbers are upregulated after allergic or inflammatory triggers. MCs are involved in host

immune defenses, immune tolerance, tissue repair, neutralization of venoms, and homeostasis.[2]

Immune Response to Infection

MCs contribute to both innate and adaptive immune responses. Although found in most tissues, the strategic distribution of MCs at surfaces with environmental exposure, such as skin and mucosal surfaces,[2] supports a role in pathogen recognition and host defense.

Innate Immunity

MCs contribute to the innate immune response through toll-like receptors (TLRs), NOD-like receptor (NLR), RIG-like receptor (RIG), and complement receptor recognition of various antigens, toxins, and pathogens,[4] resulting in the release of inflammatory mediators help to contain and clear infection and aid in the recruitment of effector cells (*e.g.*, neutrophils and macrophages). TLR-2 recognizes peptidoglycans from gram-positive bacteria, gram-negative bacteria, and mycobacteria, resulting in cytokine release (TNF-α, IL-1β, IL-4, IL-5, IL-6, IL-13), production of LTC4, and MC degranulation.[2,4] Similarly, TLR-4 binds lipopolysaccharides (LPSs) from gram-negative bacteria, lipid A, fibrinogen, and *Mycobacterium tuberculosis*, to induce cytokine release (TNF-α, IL-1β, IL-6, IL-13) without degranulation.[2,4] MCs are able to directly eliminate microbes through intracellular and extracellular antimicrobial mechanisms. MCs release peptides with direct bactericidal effects (*e.g.*, cathelicidins, defensins, and psidins), and are also capable of phagocytosis and production of reactive oxygen species to aid in bacterial killing.[2] To help ensnare and kill extracellular bacteria that may have evaded phagocytosis, activated MCs can produce aptly named MC extracellular traps (MCETs).[2] MC proteases can also degrade endogenous peptides and venoms, thereby limiting their toxicity.[2] In early parasitic infections, MCs and basophils release IL-4 and IL-13 to promote IgE production, immune cell recruitment, and regulation of gastrointestinal permeability which, in addition to growth factors IL-3, SCF, and IL-9 effects, promote the expulsion of the parasite and containment of a chronic infection.[2] The role of MCs in viral infections is less well defined. In response to viral products, activated MCs express IL-1β, IL-6, CCL3, CCL4, CCL5, and CCL8.[2] MC viral recognition is also thought to incite cellular responses, specifically, the recruitment of CD8 T lymphocytes to the site of infection and the production of IFN-1 to promote viral clearance.[2]

Adaptive Immunity

In addition to their interaction with adaptive immune systems through involvement in allergic reactions through high affinity receptor for IgE, studies on the role of MCs in adaptive immunity have shown that MCs, like DC, are capable of antigen processing and presenting via MHCI and MHCII complexes.[2] MC mediators promote DC function, phenotypical maturation, and migration to lymph nodes.[2] MCs can also directly activate T lymphocytes through the release of TNF.[2] In contrast to their significant role in the propagation of adaptive immunity, MCs can also temper the duration and magnitude of the immune response via expression of antiinflammatory cytokines such as IL-10 and TGF-β.[2]

Immune Tolerance

Although long considered mediators of inflammation and immediate hypersensitivity, MCs can also act as modulators of the immune response. MC-derived IL-10 and TGF-β, for example, can downregulate the expression of the IgE-receptor FcεRI to limit IgE-mediated MC degranulation.[2,5]

Mast Cell Homeostasis

Several studies have demonstrated the importance of proinflammatory and antiinflammatory effects of IL-4, IL-10, and TGF-β on pathways of MC homeostasis.[4] In an effort to balance cell growth, MCs have pathways for both apoptosis and autophagy.[4] While MCs are susceptible to Fas- or TNF-related apoptosis-inducing ligand (TRAIL)-mediated apoptosis via caspase activation, SCF can inactivate pro-apoptotic proteins and increase expression of pro-survival proteins.[4] A separate process, autophagy, involves the degradation and reuse of intracellular organelles and proteins. Dysregulation of MC autophagy has been implicated in chronic rhinosinusitis, asthma, and systemic sclerosis.[4]

KEY CONCEPTS

Normal Mast Cell Function

- Mast cells have roles in both innate and adaptive immune responses, and are strategically distributed at environmental interfaces to optimize pathogen recognition and host defense.
- Mast cells mediate inflammation and immediate hypersensitivity, but can also act as response modulators to promote immune tolerance.
- Pathways for both apoptosis and autophagy help maintain mast cell homeostasis.
- Mast cells also produce growth factors to promote tissue repair.

MAST CELL RELATED DISEASE

Role of Mast Cells in Allergic Disease

MCs are perhaps best known for their pathogenic role in allergic diseases such as asthma and allergic rhinitis. In most cases, clinical signs of allergic response (*e.g.*, increased vascular permeability, bronchospasm, erythema, smooth muscle contraction, augmented mucus production) result from localized effects of MC mediators released in response to allergen recognition and activation of the IgE-bound FcεR1 receptor in sensitized individuals.[2,4,25]

Anaphylaxis

Anaphylaxis is an acute, potentially life-threatening, multisystem clinical reaction caused by immunological or nonimmunological activation of MCs and basophils leading to precipitous release of mediators into the circulation.[25] In the clinical setting, triggers of anaphylaxis can include food allergies, drug allergies, venom, exercise, or idiopathic etiology. Systemic mediator release can lead to rapid development of urticaria, gastrointestinal upset, respiratory distress, and cardiovascular collapse, the hallmarks of anaphylaxis. Although histamine and β-tryptase are released together, histamine is a less reliable clinical biomarker of MC degranulation due to its short half-life and other non-MC sources.[21] An acute rise in serum β-tryptase levels therefore remains the best marker of systemic MC activation in anaphylaxis.[21] The classic mechanism of anaphylaxis can be considered a severe form of IgE-mediated type I hypersensitivity, although anaphylaxis can also be independent of IgE-mediated activation.[25] Recent studies have also suggested PAF may amplify the physiologic manifestations of both IgE- and IgG-mediated anaphylaxis.[25,27] PAF response is partially dependent on calcium influx, and therefore downregulation of signal-

ing through PAF-r occurs with epinephrine and other vasoactive agents that activate adenylate cyclase, thereby increasing intracellular calcium.[25,27] In drug-induced anaphylaxis, both immunological and nonimmunological mechanisms have been proposed, including immune-complex–mediated and immune-complex–independent complement activation.[19,25] Opiates and vancomycin have been shown to directly activate MCs, and some drugs (*e.g.*, muscle relaxants) are thought to mediate anaphylaxis through activation of MRGPRX2 or other related surface receptors.[19,20]

Asthma

Asthma is a disease of chronic airway inflammation defined by the presence of variable obstruction and airway hyperresponsiveness (Chapter 43). MCs release numerous proinflammatory mediators that contribute to bronchoconstriction, impaired gas exchange, increased mucus secretion, inflammatory cell infiltration, and airway remodeling seen in asthma. Abnormal localization of MCs to airway smooth muscle bundles may be seen in bronchial biopsies of asthmatic subjects, and studies demonstrate a strong correlation between smooth muscle MC density and airway hyper-responsiveness.[28] A substantially higher proportion of (Mast cell tryptase and chymase) MC tissue culture cells was also noted on biopsy samples from subjects with severe asthma, as compared to those with mild asthma.[27,28] In clinical practice, the efficacy of anti-IgE monoclonal antibodies (*e.g.*, omalizumab), LT antagonists, and MC stabilizers (*e.g.*, cromolyn), as well as improvements with allergen immunotherapy, emphasizes MC involvement in early- and late-phase asthma responses.[17,21]

Aspirin-Exacerbated Respiratory Disease

MCs play a significant role in AERD, a triad disorder of asthma, severe rhinosinusitis, and nasal polyposis associated with distinct respiratory reactions to aspirin and other COX-1 inhibitors.[29] Notably, AERD is associated with elevated baseline levels of MCA products, including tryptase, PGD2 and LTE_4, which further increase with aspirin or COX-1 inhibitor challenge.[29] COX-1 inhibitors are thought to cause reactions by depleting homeostatic PGE_2 from surrounding tissue.[29] The role of MCs in AERD is also implied by the attenuating effects of MC stabilizers (*e.g.*, cromolyn and nedocromil) on bronchoconstriction.[30] In clinical practice, high-dose daily aspirin therapy, following successful desensitization, has been used in an effort to suppress AERD-associated nasal polyp growth. The 5-LO inhibitor zileuton may reduce baseline respiratory symptoms in AERD.[29] The novel anti-IL-4 and IL-13 monoclonal antibody dupilumab has been shown to reduce urinary PGD_2 metabolites and LTE_4[29] and is now approved for the treatment of chronic rhinosinusitis with nasal polyposis, asthma, and atopic dermatitis.

Allergic Rhinitis

Allergic rhinitis is characterized by the inflammatory response of nasal mucosa to seasonal and perennial airborne allergens (Chapter 88). Classic symptoms include sneezing, rhinorrhea, nasal and palatal itch, and nasal congestion. Increased numbers of activated MCs can be found in the nasal epithelium of patients with both seasonal and perennial allergic rhinitis.[27] Currently, intranasal corticosteroids, oral antihistamines, and oral LT inhibitors remain mainstays of therapy for allergic rhinitis. Allergen immunotherapy is beneficial in selected patients as a second-line therapy.

Food Allergy

Type 1 hypersensitivity reactions are the most widely studied mechanism of food allergy. Food allergens cross-link IgE bound to the FcεR1 receptors of MCs, with subsequent MCA and release of mediators, most notably histamine. Peanut-induced anaphylaxis in mice and humans is thought to be influenced by MC release of histamine and PAF, specifically[12] FcεR1 receptor expression is positively regulated by increased IgE concentration.[2] Symptoms of allergic reaction can involve a single or multiple organ system, generally manifesting within 2 hours of food allergen ingestion. As previously discussed, tissue microenvironment can influence differentiation of MC subpopulations. MC_T dominate in the normal lung and intestinal mucosa, while (Mast cell tryptase and chymase) MC_{tC} dominate within deep connective tissue. A recent study by Benedé and Berin suggested MC heterogeneity as an important contributor to the range of food allergy symptoms.[12] Murine model of food allergy implied that the connective tissue MC subset, specifically, is key in the development of severe systemic anaphylaxis.[12]

Role of Mast Cells in Nonallergic Disease

MCs may contribute to the pathogenesis of other nonallergic diseases, and have been implicated in cardiovascular disease, several autoimmune diseases, and even cancer. In coronary artery disease, cardiac MCs may contribute to the development of coronary inflammation and cardiac ischemia, while increased numbers of MCs can be found in atherosclerotic plaques.[2] MC–derived tryptase and chymase induce proteolytic changes in HDL particles and thereby are thought to promote formation of foam cells.[2] Notably, chymase is also a potent angiotensin converting enzyme and main producer of the coronary constrictor angiotensin II in humans.[2] MCs have been implicated in multiple autoimmune diseases including Crohn disease, rheumatoid arthritis, psoriasis, multiple sclerosis (MS), rheumatoid arthritis, Type 1 diabetes mellitus, Guillain-Barre syndrome, allergic encephalitis, bullous pemphigoid, and Sjögren syndrome.[4] However, the precise role of MCs in autoimmune pathophysiology remains mostly unclear and remains the focus of continued debate and research.

KEY CONCEPTS

Mast Cells in Allergic and Nonallergic Disease

- Mast cells significantly contribute to the pathogenesis of allergic diseases like asthma and allergic rhinitis, as well as anaphylaxis.
- An acute rise in serum β-tryptase levels therefore remains the best marker of systemic MC activation in anaphylaxis.
- Aspirin-exacerbated respiratory disease (AERD) is a triad disorder of asthma, severe rhinosinusitis with nasal polyposis, and distinct respiratory reactions to aspirin and other COX-1 inhibitors, in which mast cell activation can be demonstrated at baseline and after induction with COX-1 inhibitors.
- Mast cells have also been implicated in cardiovascular disease, several autoimmune diseases, and even cancer.

MAST CELL DISORDERS

Disorders in which MCs are involved (MCDs) can be globally classified as primary, secondary, or idiopathic. In a primary MCD, there is an inherent defect within the MC or its progenitor resulting in pathology. Primary MCDs are typically clonal disorders as is the case in mastocytosis.[5,31] In secondary MCDs,

there is a primary disease process, such as an IgE-mediated allergy that results in nonclonal MCA. In idiopathic MCDs, MCA occurs in the absence of a clearly identifiable MC abnormality or systemic disease triggering secondary MCA (examples of this include idiopathic anaphylaxis and idiopathic mast cell activation syndrome or iMCAS).[23]

Mastocytosis

Mastocytosis is characterized by abnormal clonal MC expansion driven by D816V KIT mutation and neoplastic MC accumulation in one or more organ systems.[5,31] Mastocytosis can be broadly categorized as cutaneous (CM) or systemic (SM) and then further subcategorized based on additional features.[31] CM is most commonly diagnosed in children while SM is the predominant category in adults (Fig. 44.2).

WHO classification and diagnostic criteria for mastocytosis are listed in Table 44.3. Monoclonal mast cell activation syndrome (MCAS) is diagnosed when the patient does not meet full criteria for SM but only D816V KIT mutation or $CD25^+$ MCs are detectable. Appropriate classification of mastocytosis is critical as treatment and prognosis varies by type. For example, patients with cutaneous mastocytosis who do not meet criteria

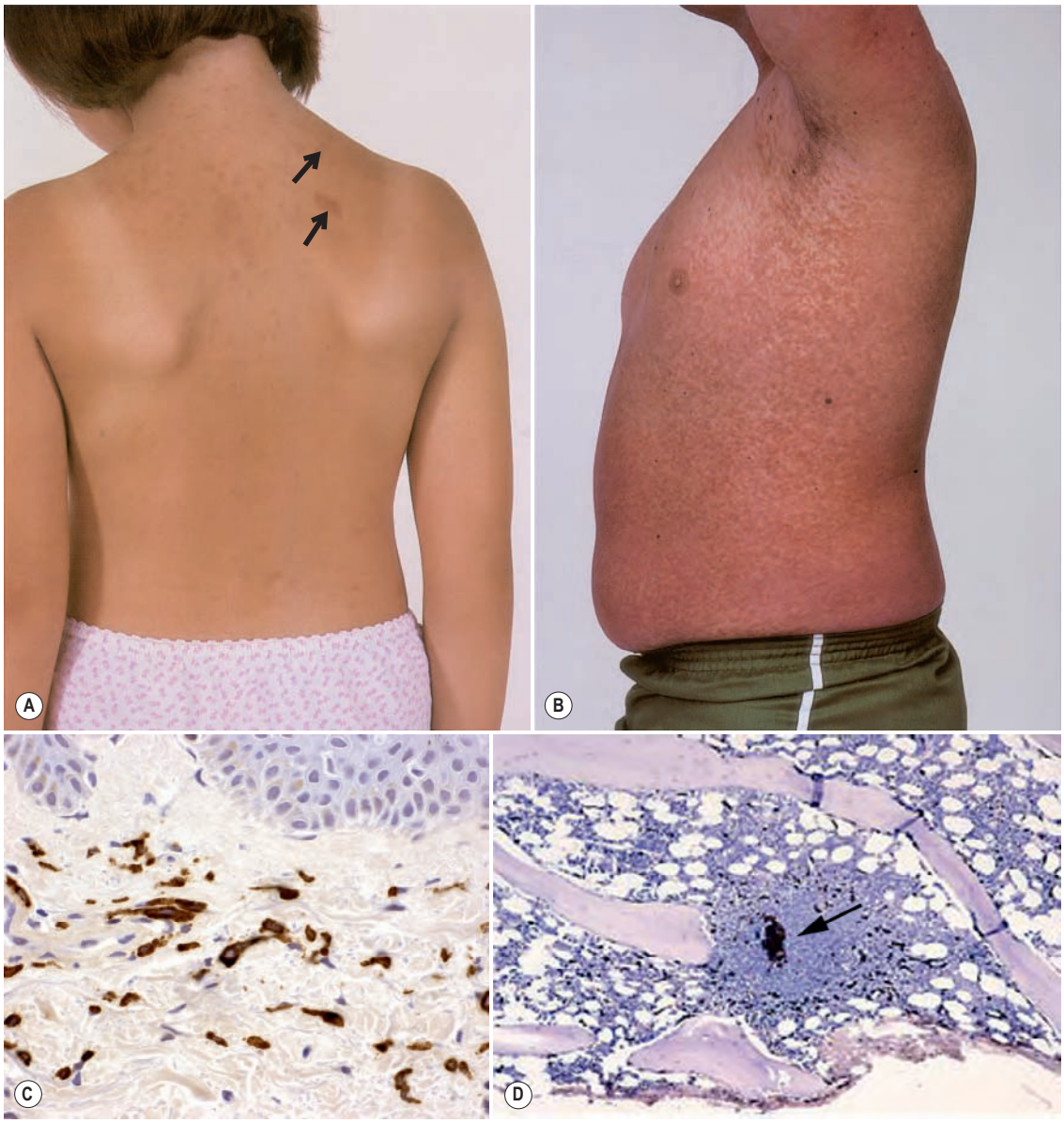


FIG. 44.2 Cutaneous Mastocytosis and Histopathology. Urticaria pigmentosa is the most common form of cutaneous mastocytosis. In childhood, lesions are disseminated and consist of well-demarcated hyperpigmented macules (*e.g.*, *arrows*) (A). In adults, lesions may be numerous with less well-demarcated brownish-red macules and papules (B). Histopathology of cutaneous mastocytosis shows many mast cells containing abundant tryptase immunoreactive cytoplasmic granules in the papillary dermis (antitryptase, AA1 clone, Dako; original magnification approximately ×400) (C). Bone marrow pathology of indolent mastocytosis is characterized by paratrabecular lymphoid nodule containing small, well-differentiated lymphoid cells around substantial numbers of fusiform cells with prominent granules in the tryptase stain *(arrow)* (antitryptase, AA1 clone, Dako; original magnification approximately ×250) (D). (Courtesy of Cem Akin.)

TABLE 44.3 Mastocytosis

Type	Diagnostic Criteria	Subtypes/Variants	Clinical Pearls
Cutaneous mastocytosis	Typical skin findings on exam. Typical infiltrates of MCs in a multifocal or diffuse pattern on skin biopsy. Does not meet diagnostic criteria for SM.	Urticaria pigmentosa (UP)/Maculopapular cutaneous mastocytosis (MPCM) Diffuse cutaneous mastocytosis (DCM) Mastocytoma of the skin	UP/MPCM is the most common skin manifestation of mastocytosis. CM is the most common form of mastocytosis in children and typically resolves by puberty. Adults w/skin findings have a higher likelihood of having SM and thus should be evaluated for SM. Patients with CM w/o evidence of SM have the best prognosis.
Systemic mastocytosis	Major criterion – multifocal, dense infiltrates of MCs (≥15 in aggregates) detected in biopsy (bone marrow and/or an extra cutaneous organ). Minor criteria • 25% MCs in infiltrate (on biopsy) are spindle shaped, have atypical morphology, or are immature. • Detection of an activating point mutation (codon 816/D816V) in biopsy or blood. • MCs in biopsy or blood express CD25 (w/ or w/o CD2). • Serum total tryptase greater than 20 ng/mL (unless there is an associated myeloid neoplasm). Diagnosis is made if one major and one minor criterion are met or if three minor criteria are met.	Indolent systemic mastocytosis (ISM) Smoldering systemic mastocytosis (SSM) = SM + "B findings"[a] Systemic mastocytosis associated with a hematologic neoplasm (SM-AHN) Aggressive systemic mastocytosis (ASM) = SM + "C findings"[b] Mast cell leukemia (MCL)	Patients w/elevated tryptase and a hx of anaphylaxis and/or negative HAT testing should be evaluated for SM. On biopsy, it is important to assess for abnormalities in other hematopoietic cell lines. After CM, ISM and SSM have the best prognosis. Skin lesions are present in most cases of ISM & frequently absent in ASM & MCL.
Mast cell sarcoma	Unifocal mast cell tumor with a destructive growth pattern. Does not meet criteria for SM.	N/A	Poor prognosis. May transform to MCL.

*B findings—30% MC infiltration on bone marrow biopsy, dysplasia, and/or proliferation of other marrow cell lines, but not enough to meet criteria for another diagnosis, and hepatosplenomegaly and/or lymphadenopathy without signs of dysfunction.
**C findings—Bone marrow dysfunction caused by MC infiltration, hepatosplenomegaly with dysfunction, osteolytic lesions (with or without fracture), and malabsorption and/or weight loss (presumably due to gastrointestinal MC infiltration).

for systemic mastocytosis have the best disease prognosis and typically require minimal therapy.[5,31] Patients with indolent systemic mastocytosis also have good prognosis with life expectancy similar to that of the general population and typically can be managed by symptomatic therapy.[5] The signs and symptoms of mastocytosis are due to MC mediator release and include pruritus, flushing, nausea, vomiting, diarrhea, abdominal pain, and hypotension. In addition, signs of tissue dysfunction (such as cytopenias, liver failure, malabsorption, pathologic fractures) due to MC burden and co-existing hematologic disorders are observed in advanced cases.

For this reason, therapy is often directed toward either mast cell mediators (avoidance of triggers to reduce mediator release as well as direct anti-mediator therapy) or reducing mast cell burden in advanced cases (cytoreductive therapy) (Table 44.4). Approximately 80% of patients with SM have a somatic gain of function mutation in KIT with the most common being in codon 816 (D816V).[5] This mutation causes KIT to be continuously active and leads to uncontrolled MC differentiation and survival that results in pathology. Given these findings, KIT inhibitors have become an area of active investigation as treatment for mastocytosis.[31] Other forms of mastocytosis such as aggressive systemic mastocytosis (ASM), mast cell leukemia (MCL), and systemic mastocytosis with associated hematologic neoplasm (SM-AHN) often have more rapid and complex disease courses that require multidisciplinary care including a referral to hematology as well as antineoplastic therapies.[31] SM-AHN is frequently associated with myeloid neoplasms which often negatively affects prognosis. SM and AHN components are treated separately. Conversion between different forms of mastocytosis is rare, but does occur; thus these patients require regular follow-up for monitoring of disease progression and complications.[5,31]

Mast Cell Activation Syndrome

MCAS is heterogeneous disorder characterized by (1) severe, multisystem, and episodic symptoms of MC mediator release; (2) improvement or resolution of symptoms with anti-mediator therapy; and (3) laboratory evidence of MCA.[23] Diagnostic testing for MCAS includes serum tryptase and urinary metabolites of MC mediators.[23] As a number of different processes can mimic MCA symptoms, the differential diagnosis can be overwhelming; thus it is essential that all three consensus criteria be fulfilled before the diagnosis is established. It has been proposed that idiopathic anaphylaxis be viewed as the quintessential presentation of MCAS, thus establishing the paradigm that MCAS is a severe systemic reaction resulting from mast cell mediator release.[23] As such, therapy for MCAS should be focused on reducing MCA (avoiding triggers, use of MC stabilizers, *etc.*) and pharmacologically counteracting MC mediators. As our

TABLE 44.4 Treatment of Mastocytosis

Target	Drug/Intervention	Clinical Pearls
MC mediator symptoms	H1 blockers	Most patients with mastocytosis require at least an H1 blocker.
	H2 blockers	H2 blockers are often used an adjunct to H1 blockers for cutaneous symptoms and also to target GI symptoms.
	Leukotriene antagonists	
	Mast cell stabilizers	Cromolyn (mast cell stabilizer) can be used as an adjunct to H1 blockers and may help with GI symptoms.
	Epinephrine	
	Corticosteroids (oral and topical)	Patients with diarrhea tend to respond well to oral steroids. An alternative to consider are anticholinergics.
	Anticholinergics	
	Light therapy	Topical steroids can improve lesions and symptoms in patients with UP or DCM.
		8-Methoxypsoralen in conjunction with ultraviolet phototherapy can be used for adults with refractory cutaneous disease and symptoms
Osteoporosis	Calcium supplementation	All patients with SM should be screened for osteoporosis.
	Estrogen replacement	
	Bisphosphonates	Radiotherapy can be used in a palliative setting for management of isolated bone lesions in aggressive disease.
	Denosumab	
	Radiotherapy	
Cytoreductive therapies	Interferon alpha2b (IFN-α2b)2-chloro-2-deoxyadenosine (cladribine, 2-CdA)	Typically used in advance disease.
	Tyrosine kinase inhibitors (TKI)	IFN-α2b is historically first line therapy for ASM, but has poor tolerance and relapse is common after discontinuation.
	• Imatinib (Gleevec)	Cladribine/2-CdA is also historically a first-line therapy for ASM, but is associated with many toxicities (myelosuppression, opportunistic infection, *etc.*).
	• Midostaurin	
	• Investigational TKIs (including avapritinib (D816V selective KIT inhibitor and ripretinib (switch pocket inhibitor)Hematologic stem cell transplant	Efficacy of imatinib is limited by its lack of activity against D816V KIT mutation, the most common mutation in mastocytosis. Imatinib is effective for patients with juxtamembrane or transmembrane *KIT* mutations, wild-type *KIT*, or *FIP1L1-PDGFRA* fusion oncogene (all rare findings in systemic mastocytosis).
		Midostaurin is a multikinase inhibitor with activity against wild type and D816V mutated KIT, and avapritinib is a D816V selective tyrosine kinase inhibitor, both of which are FDA approved is FDA approved for treatment of ASM, SMAHN, and MCL.
		Stem-cell transplant may be considered for select patients (ASM, SMAHN, and MCL); survival benefit seems to be greatest in SMAHN.

understanding of mast cell biology evolves we may be better able to identify aberrancies that lead to MCAS as well as future targets for therapy.

BASOPHILS AND RELATED DISORDERS

Basophils are granulocytic cells that share many morphologic and biochemical features with MCs. In contrast to MCs, basophils complete maturation in the bone marrow and are released into peripheral circulation as mature end-stage cells. Basophil development and maturation is regulated by a number of cytokines. One of the most critical cytokines for human basophil growth is IL-3. Granulocyte/macrophage colony stimulating factor (GM-CSF), IL-5, TGF-β, and TSLP are also thought to play a role in basophil development.

Basophils share many of the same mediators (*e.g.*, histamine and CysLTs) as well as mechanisms of activation (*e.g.*, IgE/FcεR1) with MCs.[21] For this reason, basophils may also contribute to pathology involving IgE-mediated MC activation and degranulation. Basophils are an important source of IL-4 and IL-13 after antigen exposure,[32] suggesting that they play an important role in Th2-mediated allergic disease. Several studies have shown that basophils in allergic individuals undergo a process known as *priming* whereby phenotypic and functional changes increase basophil propensity for stimulation by allergen exposure.[33] Basophils also migrate to sites of inflammation in allergic disease such as the nose as well as the lung and play an important role in the generation of the late phase allergic response.[34,35] While basophils and MCs share many similarities, basophils do not produce significant amounts of PGD2 and have a limited cytokine profile after activation (predominantly Th2). Collectively, these findings suggest that basophils are a critical component of Th2- and IgE-mediated disease processes.

Given the role basophils play in allergic disease, flow cytometry–based tests of basophil activation have been developed for the assessment of immediate hypersensitivity. Two of the most common markers used in the basophil activation test (BAT) are CD63 and CD203c which are induced on the basophil surface after an allergen-IgE complex binds to FcεR1.[21] The BAT has the potential to be used both for the detection of immediate hypersensitivity as well as monitoring for the induction of a hyposensitized state (Fig. 44.3).

Basophils are found to be decreased in blood and have alterations in their IgE-receptor–mediated activation in patients with chronic spontaneous urticaria (CSU). This is thought to reflect their migration to tissue from blood and their possible involvement in symptomatology of CSU. In addition to diseases in which basophils are co-activated with MCs, basophils can be increased in hematologic neoplasias. Basophilic leukemia (defined by >40% circulating basophils) can be classified as *acute* (>20% blasts) or *chronic* (<20% blasts), *primary* (where there is an inherent basophil issue), or *secondary* (where another hematopoietic disorder such as chronic myeloid leukemia results in a secondary basophilia). The treatment of basophil-mediated symptomatology is similar to that of MC disease and dependent upon presence of cell activation and total cell burden. In the case of activation, treatment is focused on avoidance of triggers, and counteracting released mediators. In the case of increased basophil cell burden, antineoplastic or cytoreductive therapy may be sought.

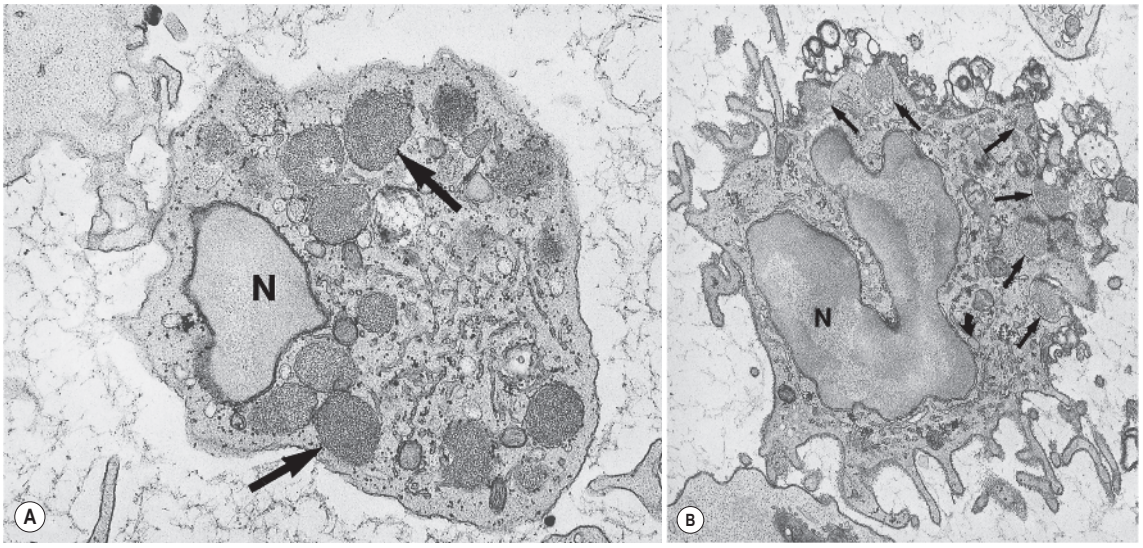


FIG. 44.3 (A) Transmission electron micrograph of a human basophil in a preparation of peripheral blood leukocytes obtained by separation over Ficoll-Hypaque. All of the cytoplasmic granules (some indicated by *solid arrows*) contain particulate electron-dense material. *N*, Nucleus. (Original magnification approximately ×19,800.) (B) A human basophil 2 minutes after exposure to antigen *in vitro*. The cell exhibits extrusion of granules from six separate sites on the plasma membrane *(small arrows)*. At this time after cell stimulation, particle-filled granules retain their shape and characteristic structure even after exposure to extracellular milieu. Cationized ferritin coats the cell surface and enters culs-de-sac that contain exteriorized granules. The cell exhibits no fully intracytoplasmic typical basophilic granules, but one of the smaller kind of granules *(curved arrow)* can be observed in the perinuclear region. *N*, Nucleus. (Original magnification approximately ×19,200.) (From Dvorak AM, Newball HH, Dvorak HF, Lichtenstein LM. Antigen-induced IgE-mediated degranulation of human basophils. *Lab Invest*. 1980;43:126–139, with permission from Nature Publishing Group Ltd.)

KEY CONCEPTS

Basophils

- Basophils originate from the bone marrow and are of myeloid lineage, but, unlike MCs, complete their maturation in the marrow before being released into the peripheral circulation as mature cells.
- Basophil development occurs in various stages and is regulated by multiple factors with one of the most important being IL-3.
- The basophil activation test (BAT) is a flow cytometry–based test that looks for expression of CD63 and CD203c on the basophil surface in response to allergen-IgE complex stimulation to assess immediate hypersensitivity.
- Basophil-related disease is due to either basophil activation, increased basophil number, or a combination of both. Treatment is similar to that of mast cell disorders and includes foiling cell activation and reducing cell number.

REFERENCES

1. Dahlin JS, Hallgren J. Mast cell progenitors: origin, development and migration to tissues. *Mol Immunol*. 2015;63(1):9–17.
2. da Silva EZ, Jamur MC, Oliver C. Mast cell function: a new vision of an old cell. *J Histochem Cytochem*. 2014;62(10):698–738.
3. Kirshenbaum AS, Goff JP, Semere T, et al. Demonstration that human mast cells arise from a progenitor cell population that is CD34(+), c-kit(+), and expresses aminopeptidase N (CD13). *Blood*. 1999;94(7): 2333–2342.
4. Caslin HL, Kiwanuka KN, Haque TT, et al. Controlling mast cell activation and homeostasis: work influenced by Bill Paul that continues today. *Front Immunol*. 2018;9:868.
5. Theoharides TC, Valent P, Akin C. Mast cells, mastocytosis, and related disorders. *N Engl J Med*. 2015;373(2):163–172.
6. Dahlin JS, Ekoff M, Grootens J, et al. KIT signaling is dispensable for human mast cell progenitor development. *Blood*. 2017;130(16): 1785–1794.
7. Chen CC, Grimbaldeston MA, Tsai M, et al. Identification of mast cell progenitors in adult mice. *Proc Natl Acad Sci U S A*. 2005;102(32):11408–11413.
8. Salomonsson M, Dahlin JS, Ungerstedt J, et al. Localization-specific expression of CCR1 and CCR5 by mast cell progenitors. *Front Immunol*. 2020;11:321.
9. Dwyer DF, Barrett NA, Austen KF. Immunological Genome Project C. Expression profiling of constitutive mast cells reveals a unique identity within the immune system. *Nat Immunol*. 2016;17(7):878–887.
10. Akula S, Paivandy A, Fu Z, et al. Quantitative in-depth analysis of the mouse mast cell transcriptome reveals organ-specific mast cell heterogeneity. *Cells*. 2020;9(1):211.
11. Chhiba KD, Hsu CL, Berdnikovs S, et al. Transcriptional heterogeneity of mast cells and basophils upon activation. *J Immunol*. 2017;198(12):4868–4878.
12. Benede S, Berin MC. Mast cell heterogeneity underlies different manifestations of food allergy in mice. *PLoS One*. 2018;13(1):e0190453.
13. Bax HJ, Bowen H, Dodev TS, et al. Mechanism of the antigen-independent cytokinergic SPE-7 IgE activation of human mast cells in vitro. *Sci Rep*. 2015;5:9538.
14. Asai K, Kitaura J, Kawakami Y, et al. Regulation of mast cell survival by IgE. *Immunity*. 2001;14(6):791–800.
15. Kalesnikoff J, Huber M, Lam V, et al. Monomeric IgE stimulates signaling pathways in mast cells that lead to cytokine production and cell survival. *Immunity*. 2001;14(6):801–811.
16. Slodka A, Wiktorska M, Brzezinska-Blaszczyk E. IgE by itself affects mature rat mast cell preformed and de novo-synthesized mediator release and amplifies mast cell migratory response. *PLoS One*. 2013;8(10):e79286.
17. Kawakami T, Blank U. From IgE to Omalizumab. *J Immunol*. 2016;197(11):4187–4192.
18. Yu Y, Blokhuis BR, Garssen J, et al. Non-IgE mediated mast cell activation. *Eur J Pharmacol*. 2016;778:33–43.
19. Lyons DO, Pullen NA. Beyond IgE: alternative mast cell activation across different disease states. *Int J Mol Sci*. 2020;21(4):1498.
20. Porebski G, Kwiecien K, Pawica M, et al. Mas-Related G Protein-Coupled Receptor-X2 (MRGPRX2) in drug hypersensitivity reactions. *Front Immunol*. 2018;9:3027.

21. Kabashima K, Nakashima C, Nonomura Y, et al. Biomarkers for evaluation of mast cell and basophil activation. *Immunol Rev*. 2018;282(1):114–120.
22. Lyons JJ, Yu X, Hughes JD, et al. Elevated basal serum tryptase identifies a multisystem disorder associated with increased TPSAB1 copy number. *Nat Genet*. 2016;48(12):1564–1569.
23. Khokhar D, Akin C. Mast cell activation: when the whole is greater than the sum of its parts. *Med Clin North Am*. 2020;104(1):177–187.
24. Metcalfe DD, Pawankar R, Ackerman SJ, et al. Biomarkers of the involvement of mast cells, basophils and eosinophils in asthma and allergic diseases. *World Allergy Organ J*. 2016;9:7.
25. Gill P, Jindal NL, Jagdis A, et al. Platelets in the immune response: revisiting platelet-activating factor in anaphylaxis. *J Allergy Clin Immunol*. 2015;135(6):1424–1432.
26. Bankova LG, Boyce JA. A new spin on mast cells and cysteinyl leukotrienes: leukotriene E4 activates mast cells in vivo. *J Allergy Clin Immunol*. 2018;142(4):1056–1057.
27. Kajiwara N, Sasaki T, Bradding P, et al. Activation of human mast cells through the platelet-activating factor receptor. *J Allergy Clin Immunol*. 2010;125(5):1137–1145. e6.
28. Siddiqui S, Mistry V, Doe C, et al. Airway hyperresponsiveness is dissociated from airway wall structural remodeling. *J Allergy Clin Immunol*. 2008;122(2):335–341. 41 e1–3.
29. Cahill KN, Boyce JA. Aspirin-exacerbated respiratory disease: mediators and mechanisms of a clinical disease. *J Allergy Clin Immunol*. 2017;139(3):764–766.
30. Boyce JA. Aspirin sensitivity: lessons in the regulation (and dysregulation) of mast cell function. *J Allergy Clin Immunol*. 2019;144(4):875–881.
31. Valent P, Akin C, Metcalfe DD. Mastocytosis: 2016 updated WHO classification and novel emerging treatment concepts. *Blood*. 2017;129(11):1420–1427.
32. MacGlashan Jr. D, White JM, Huang SK, et al. Secretion of IL-4 from human basophils. The relationship between IL-4 mRNA and protein in resting and stimulated basophils. *J Immunol*. 1994;152(6):3006–3016.
33. Schroeder JT, Chichester KL, Bieneman AP. Human basophils secrete IL-3: evidence of autocrine priming for phenotypic and functional responses in allergic disease. *J Immunol*. 2009;182(4):2432–2438.
34. Schroeder JT, MacGlashan Jr. DW, Lichtenstein LM. Human basophils: mediator release and cytokine production. *Adv Immunol*. 2001;77:93–122.
35. Nouri-Aria KT, Irani AM, Jacobson MR, et al. Basophil recruitment and IL-4 production during human allergen-induced late asthma. *J Allergy Clin Immunol*. 2001;108(2):205–211.

45

Eosinophils and Eosinophilic Disorders

Amy D. Klion and Paneez Khoury

First described in the 1800s by Paul Ehrlich, eosinophils are rare cells of the myeloid lineage that are characterized by a bilobed nucleus and abundant secondary granules that bind the acidic red dye, eosin. Best known for their role as innate effector cells in allergic inflammation and helminth infection, eosinophils or eosinophil-like cells have been found in all vertebrates examined to date, including fish and even invertebrate cockroaches.[1] Conservation of eosinophils across species over millions of years combined with the lack of evidence for congenital eosinophil deficiency in humans[2] suggest that eosinophils play important biologic roles. Despite the viability of genetically engineered "eosinophil-less" murine models and the lack of safety signals in clinical trials of eosinophil-depleting therapies, recent data from murine and human studies support this hypothesis and are beginning to delineate roles for eosinophils in a wide variety of homeostatic processes. This is of particular interest as the number of therapeutic agents targeting eosinophils increases. After a brief summary of the basic biology of eosinophils and eosinophil activation, this chapter focuses primarily on the role of eosinophils in health and disease, including novel therapeutic approaches and their contribution to our understanding of the role that eosinophils play in homeostasis and pathogenesis.

EOSINOPHIL BIOLOGY

Eosinophil Life Cycle

The relative numbers of eosinophils in the bone marrow, blood, and tissue at any one time depend on the balance between eosinophilopoiesis, migration to and from the bloodstream, and apoptosis. Eosinophils develop from pluripotent $CD34^+$ hematopoietic stem cells in the bone marrow through the complex interplay of transcription factors, cytokines, and growth factors (reviewed in Fulkerson and Rothenberg[3]). A key step in this process is the generation of lineage-committed eosinophil precursors (EoPs) from a common eosinophil/mast cell progenitor that expresses the transcription factor GATA-1. *Gata1* gene expression is essential for eosinophil development in mice and humans and occurs via an eosinophil-lineage-specific enhancer in the gene itself. This double palindromic enhancer is also present in genes encoding other characteristic eosinophil proteins, such as eosinophil peroxidase *(EPX)*, interleukin-5 receptor α *(IL-5RA)*, and the eotaxin receptor gene *CCR3*; targeted deletion of the enhancer led to the creation of one of the first mouse strains completely lacking eosinophils.[4] Additional events in early eosinophil lineage differentiation include downregulation of friend of GATA-1 *(FOG1)* in the context of expression of CCAAT/enhancer binding protein (C/EBP)α C/EBPε, interferon regulatory factor (IRF8), PU.1, and tribbles homolog 1 (TRIB1).

EoPs are blastic cells that have fine intracytoplasmic granules and express both CD34 and IL-5RA on their surface.[5] As they differentiate into mature eosinophils, they lose expression of CD34 and acquire the characteristic morphologic features of mature eosinophils. Although EoPs are programmed to become eosinophils even in the absence of IL5, increased IL5 levels, as seen in helminth infection and allergic reactions, stimulate their proliferation and terminal differentiation. Other cytokines and chemokines, including IL33, IL3, granulocyte–macrophage colony-stimulating factor (GM-CSF), and members of the eotaxin family, can induce eosinophil differentiation in vitro in the absence of IL5, adding to the complexity of the process.

Mature eosinophils (and small numbers of EoPs) exit the bone marrow into the peripheral circulation, where, under homeostatic conditions, they remain for approximately 1 day before migrating to the tissue.[6] As is true for other circulating leukocytes, eosinophil egress from the bloodstream involves multiple steps and surface expression of various selectins and integrins important in endothelial interactions, including tethering, rolling, and transendothelial migration (reviewed in Klion et al.[7]; Fig. 45.1). Whereas eotaxin-1 (CCL11) appears to be the major driver of eosinophil trafficking to the thymus, lymph nodes, spleen, stomach, and intestine in healthy individuals, local production of additional mediators, including eotaxin-3 (CCL26) and prostaglandin D_2, in response to allergen stimulation, tissue damage, or other factors, can attract eosinophils to almost any tissue in the setting of eosinophilic disease. This has clinical implications since therapeutic agents that affect expression or block function of these molecules, such as natalizumab (anti-CD49d), vedolizumab (anti-α4β7), and dupilumab (anti-IL4 receptor α), can cause impaired eosinophil tracking to affected tissues resulting in increased blood eosinophilia.

The life span of eosinophils in the tissue under homeostatic conditions is generally believed to be less than a week. In the setting of inflammation and local secretion of anti-apoptotic cytokines, including IL5, eosinophils likely persist longer. Moreover, in situ eosinophilopoiesis from EoPs in response to local production of IL5 and other mediators may contribute to tissue eosinophilia in some cases.[8]

Eosinophil Structure

Mature resting eosinophils are characterized by a bilobed nucleus and a cytoplasm containing abundant specific granules (Fig. 45.2). These granules are composed of an electron-dense core of major basic protein (MBP) and a matrix, containing the highly cationic eosinophil granule proteins (MBP, EPX,

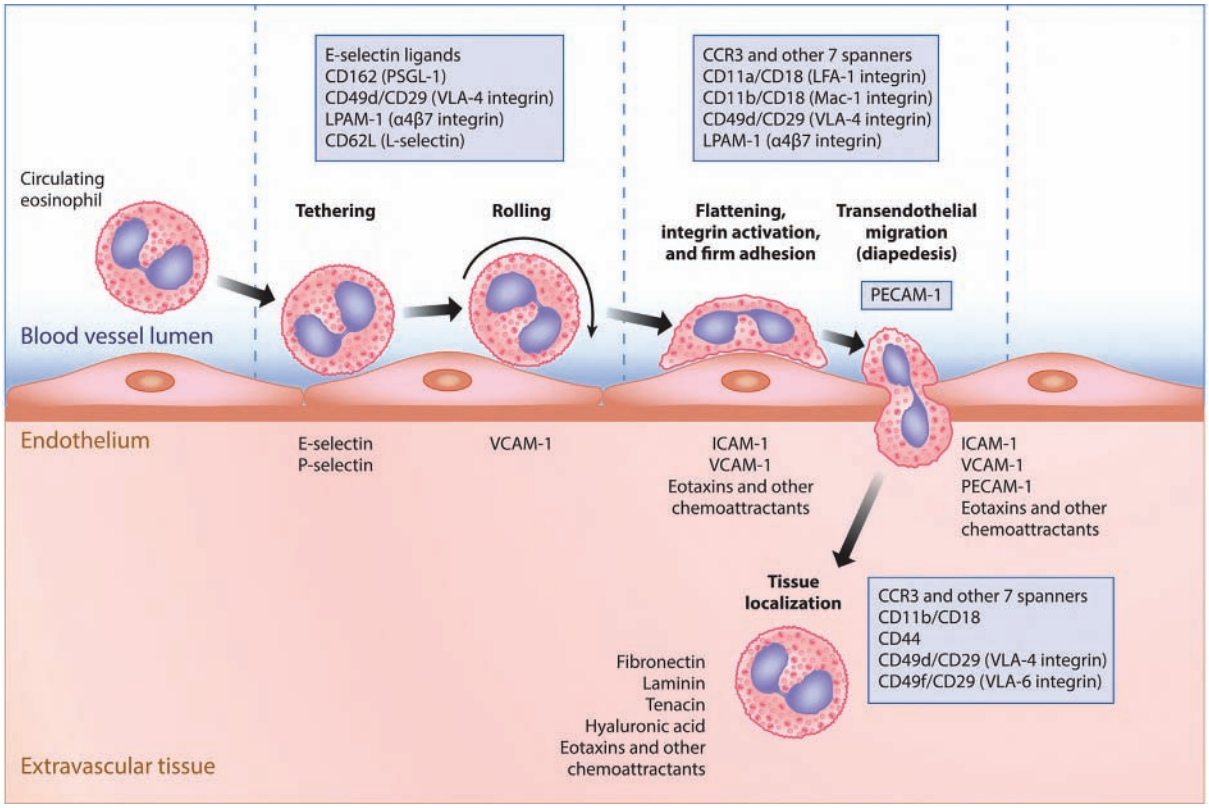


FIG. 45.1 **Eosinophil life cycle** (from Klion, Ackerman and Bochner (artist: Jacqueline Schaffer). Annu Rev Pathol; PMID 31977298).

eosinophil-derived neurotoxin [EDN], and eosinophil cationic protein [ECP]), as well as a multitude of preformed cytokines, chemokines, growth factors, and receptors.[9] In addition to the cellular machinery typical of any blood cell, the cytoplasm of eosinophils contains a complex vesiculotubular network and galectin-10 (Charcot-Leyden crystal protein). Galectin-10, one of the most abundant eosinophil proteins, appears to play an essential role in specific granule formation.[10] Although not present in resting eosinophils, lipid bodies, cytoplasmic organelles rich in arachidonic acid that synthesize inflammatory lipid mediators, including leukotriene C4 and prostaglandin D_2, are rapidly formed following eosinophil activation.[11]

Eosinophil Activation and Degranulation

Eosinophil activation can be intrinsic (i.e., due to activating mutations in the myeloid lineage) or in response to external stimuli, including cytokines, chemokines, and alarmins. Depending on the stimulus, activation can lead to degranulation by one of several mechanisms: (1) exocytosis (fusion of one or multiple granules creating a channel for extracellular release of mediators), (2) piecemeal degranulation (shuttling of selected granule contents to the cell surface through the tubulovesicular network via specific receptors), and (3) cytolysis (granule release during cell death with or without the release of nuclear DNA in the form of eosinophil extracellular traps [EETs]).[12] Activation-induced rapid release of mitochondrial DNA nets containing granule proteins has also been described in the absence of cytolysis.

Another prominent feature of eosinophil activation is altered surface receptor expression. In addition to upregulation of adhesion molecules involved in egress from the bloodstream, changes in expression levels of chemokine receptors, immunoglobulin receptors, vesicle-associated membrane proteins (e.g., CD63), and activation markers (e.g., CD69, CD25) have been described.[7] Other events accompanying eosinophil activation include the secretion of lipid mediators and reactive oxygen species and the formation and release of Charcot-Leyden crystals, characteristic birefringent crystals found at sites of eosinophilic inflammation in tissues and in association with EETs.

Irrespective of the mechanism of activation, deposition of eosinophil granule protein contents and release of inflammatory mediators contribute to the effector functions of eosinophils but can also cause tissue damage, hypercoagulability, and fibrosis. Although the factors that determine the outcome of eosinophilic inflammation in a given tissue or individual are poorly understood and remain a topic of investigation, the generation of murine models lacking eosinophils and the availability of therapies specifically targeting eosinophils in humans have led to recent advances in our understanding of eosinophil function in health and disease.

ROLE OF EOSINOPHILS IN HOMEOSTASIS

Whereas eosinophil-deficient mice have no apparent developmental defects and normal life spans, an expanding number of abnormalities in basic homeostatic functions have been reported, including reduced body fat mass, glucose intolerance, impaired vaccine recall responses, and alterations in immune responses and microbiome composition in the gut.[12] Most of these functions have been attributed to eosinophil secretion of mediators that modulate the development and function of other

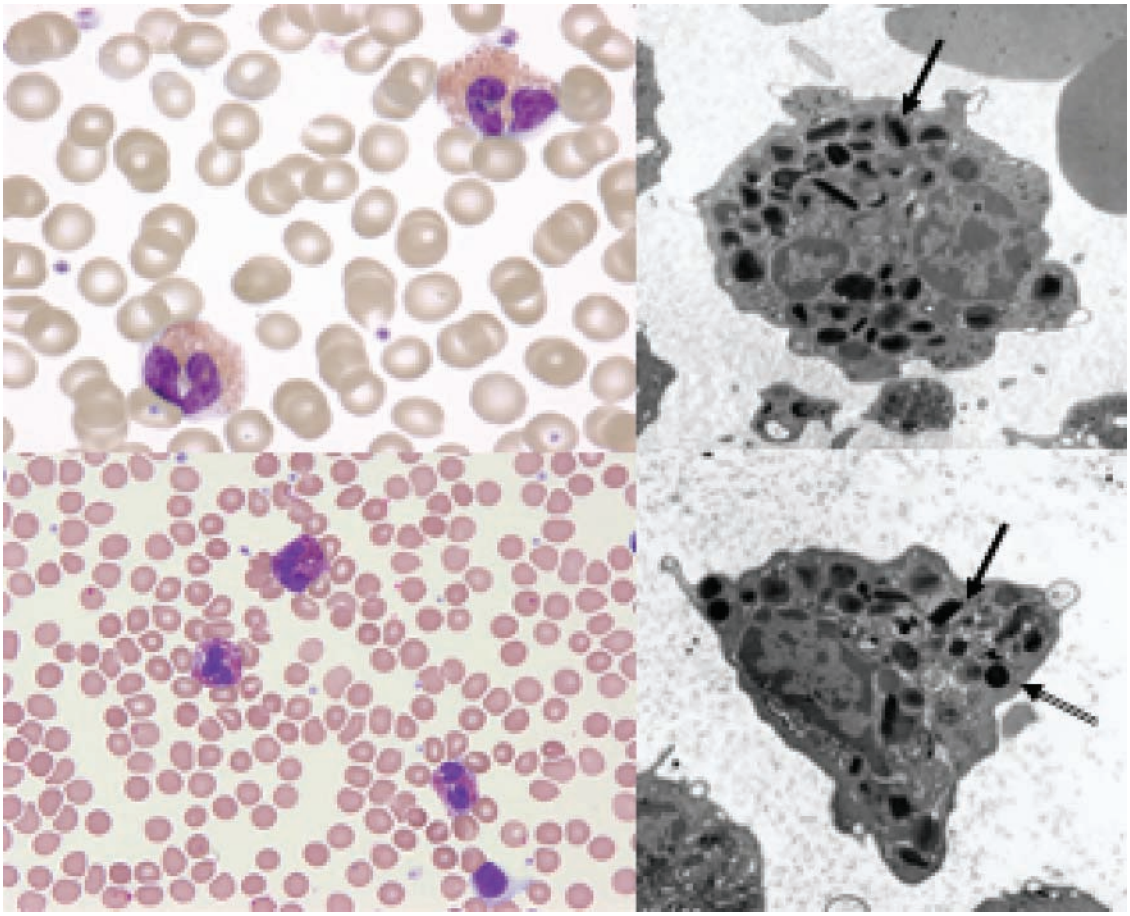


FIG. 45.2 Eosinophil Structure. Resting *(upper panels)* and activated *(lower panels)* eosinophils are shown by light and electron microscopy. Specific granules are indicated by *solid arrows* and lipid bodies by an *open arrow*. (Adapted from Klion et al. Blood 2004 PMID 14988154.)

immune cells, although eosinophils likely play a more direct role in some processes, including tumor surveillance.

The applicability of these findings to humans has been more difficult to unravel. Many of the relationships between numbers of eosinophils and other cell types have been confirmed in humans; however, the functional consequences of these associations remain unclear. Congenital absence of eosinophils has not been reported to date, and although rare cases of acquired eosinophil deficiency have been described, they do not appear to be associated with consistent abnormalities related to eosinophil deficiency. More recently, the availability of therapeutics that target eosinophils has begun to directly address the consequences of eosinophil depletion in humans. Although studies to date have failed to confirm many of the murine findings (e.g., impaired responses to immunization, glucose intolerance, predisposition to malignancy), additional studies over longer time frames are needed. Moreover, subclinical effects of eosinophil reduction may be potentiated when the function of other cell types is compromised, either as a result of the underlying disease process or therapeutic intervention.

Recent data from murine models have convincingly demonstrated the existence of a distinct subset of lung-resident eosinophils (rEOs) that play a regulatory role in the absence of inflammation.[13] These rEOs are functionally IL-5-independent (but express IL-5Rα) and have a distinct surface immunophenotype and RNA expression profile compared to eosinophils that enter the tissue following allergen challenge and steady-state eosinophils in the small intestine in mice. Whereas human eosinophils exhibit phenotypic heterogeneity in different microenvironments, the existence of a distinct terminally differentiated regulatory eosinophil subset in humans remains controversial.

EOSINOPHILIA AND EOSINOPHILIC DISORDERS

Definition of Eosinophilia

Eosinophilia is defined as an absolute eosinophil count (AEC) of approximately greater than 500/mm^3 and hypereosinophilia (HE) as an AEC at least 1500/mm^3 in the peripheral blood, or markedly increased tissue eosinophilia. Whereas the prevalence of eosinophilia ranges from approximately 5% in healthy military recruits to more than 30% in selected immigrant groups, HE is much less common and more likely to be associated with clinical manifestations. Hypereosinophilic syndromes (HES) are a rare group of heterogeneous conditions characterized by HE and confirmed or presumed eosinophil-driven clinical manifestations. Whereas the level of blood eosinophilia can be helpful in monitoring disease activity, especially in an individual patient with HES, this does not always reflect the extent and severity of organ or tissue damage involvement. The use of targeted therapeutics with differing effects on blood and tissue

eosinophils has further complicated the interpretation of blood eosinophil levels.

Diagnostic Approach to the Patient With Eosinophilia

The causes of eosinophilia and eosinophil-associated disease are myriad (Table 45.1) and range from disorders with relatively benign manifestations and only minimal peripheral eosinophilia, such as allergic rhinitis and chronic rhinosinusitis, to more severe conditions with marked eosinophilia and severe end-organ manifestations, including eosinophilic granulomatosis with polyangiitis (EGPA) and eosinophilic myeloid malignancies. Although helminth infections likely still account for the highest percentage of mild to moderate eosinophilia worldwide, their prevalence is decreasing. In contrast, the prevalence of atopic conditions associated with eosinophilia, including asthma, chronic rhinosinusitis, atopic dermatitis, and drug-related eosinophilia, continues to rise.

The urgency of evaluation and need for treatment depend on the degree of eosinophilia, clinical context, and potential underlying causes. In this regard, a comprehensive history, including onset of the eosinophilia, travel and potential exposures, current and recently discontinued medications, prior medical history, and current symptoms, is essential. A complete physical examination with assessment of lymphadenopathy, splenomegaly, and skin findings, complete blood count with differential and routine biochemical testing should be performed in all patients with additional laboratory and diagnostic studies dictated by the preliminary clinical findings. Of note, intact eosinophils may be absent in biopsies despite immunohistochemical evidence of eosinophil degranulation. Although an in-depth discussion of all of the potential causes of eosinophilia is beyond the scope of this chapter, the following section provides a system-based discussion of some of the more common eosinophil-associated disorders. Clinically distinct syndromes and hypereosinophilic disorders restricted to a single organ system (*i.e.*, overlap HES) are included in the system-based section. Multisystem eosinophilic disorders, including rare clinical subtypes of HE/HES, are discussed separately.

TABLE 45.1 Eosinophil-Associated Disorders and Clinical Syndromes

Broad Categories	Examples (not exhaustive)
Allergic and atopic skin and soft tissue	Atopic dermatitis, eosinophilic asthma, chronic rhinosinusitis
	Atopic dermatitis — allergic and in the setting of immune dysregulatory disorders
	Bullous diseases (pemphigoid and pemphigus)
	Eosinophilic folliculitis
	Eosinophilic cellulitis (Wells syndrome)
	Eosinophilic fasciitis
	Episodic angioedema with eosinophilia
	Cutaneous manifestations of parasitic infection, such as cutaneous larva migrans and onchocercal dermatitis
	Drug hypersensitivity reactions
	Cutaneous T-cell lymphoma
Pulmonary	Acute and chronic eosinophilic pneumonia
	Allergic bronchopulmonary aspergillosis
	Helminth infection, including Loeffler syndrome (transpulmonary migration of helminth parasites) and tropical pulmonary eosinophilia
	Fungal infection, including coccidiomycosis, cryptococcosis, and histoplasmosis
	Drug hypersensitivity pneumonitis
	EGPA
Gastrointestinal	Eosinophilic gastrointestinal disorders (EGIDs)
	Eosinophilic hepatitis — idiopathic, autoimmune, drug, or helminth-associated
	Eosinophilic cholecystitis
	Inflammatory bowel disease
Renal or urinary	Acute interstitial nephritis, often medication-induced
	Eosinophilic vasculitis
	Eosinophilic cystitis, including primary and secondary (*e.g.*, to urinary schistosomiasis)
Cardiac	Endomyocardial fibrosis, eosinophilic myocarditis, or eosinophilic pericarditis, typically in the setting of EGPA or rare HES variants
Neuromuscular	Mononeuritis multiplex, typically in the setting of EGPA or rare HES variants
	Eosinophilic meningitis due to medication, fungal or parasite infection, and rarely lymphoma
	Eosinophilic myositis, including calpain-3 associated, autoimmune, and infectious (*e.g.*, trichinosis and sarcocystosis)
Drug and toxin-related	Drug hypersensitivity reactions, including DRESS
	Drug effects, such as those due to IL-2, GM-CSF infusion
	Eosinophilia-myalgia syndrome
Infectious	Helminth infection, ectoparasites (scabies, myiasis), various fungal infections, human immunodeficiency virus
immune	Rheumatologic disorders, including EGPA, IgG4-related disease and rheumatoid arthritis
	Sarcoidosis
	Primary immunodeficiency disorders, such as DOCK8 deficiency, and Omenn syndrome
	Transplant rejection
Hematologic and neoplastic	Myeloid HESs, including those associated with abnormalities in *PDGFRA, PDGFRB, JAK2*, and *FGFR1* and chronic eosinophilic leukemia-not otherwise specified
	Systemic mastocytosis
	Lymphocytic leukemias, especially T-cell leukemia and pre-B-cell ALL, and varied lymphomas
	Solid tumors, especially adenocarcinoma
Other	Rare HESs, hypoadrenalism, cholesterol embolism, serosal surface irritation, familial hypereosinophilia

ALL, Acute lymphocytic leukemia; *GM-CSF*, granulocyte–macrophage colony-stimulating factor; *DOCK8*, dedicator of cytokinesis 8; DRESS, drug rash with eosinophilia and systemic symptoms; *EGPA*, eosinophilic granulomatosis with polyangiitis; *GM-CSF*, granulocyte–macrophage colony-stimulating factor;*HES*, hypereosinophilic syndrome; *IgG4*, immunoglobulin G4; *IL-2*, interleukin-2.

Eosinophilia by Organ System

Dermatologic and Soft Tissue Disorders

Eosinophilic skin infiltration is extremely common and can occur in the context of a wide variety of diagnoses with or without concomitant peripheral eosinophilia (reviewed in Leiferman and Peters[14]). The clinical manifestations of skin and soft tissue eosinophilia are protean and include pruritus, urticaria, angioedema, maculopapular eruptions, bullae, ulcers, and fibrosis. Although clinical findings are sometimes pathognomonic (e.g., the "groove sign" in eosinophilic fasciitis or the serpiginous pruritic track of cutaneous larva migrans), this is the exception rather than the rule. Common conditions, such as atopic dermatitis, can present with moderate to marked blood eosinophilia and can be particularly difficult to differentiate from the initial presentation of a systemic HES. In addition, eosinophilic dermatologic conditions can sometimes be a clue to an underlying non-eosinophilic diagnosis, as in the case of eosinophilic pustular folliculitis and human immunodeficiency virus (HIV) infection.

Parasitic infections, infestations (e.g., scabies), and environmental or drug hypersensitivity reactions, which can include drug reaction with eosinophilia and systemic symptoms (DRESS), are among the most frequent secondary causes of eosinophilia, and dermatologic involvement should be considered in the setting of an appropriate exposure history. Autoimmune blistering disorders (e.g., pemphigoid and pemphigus), immunoglobulin G4 (IgG4)-related diseases, vasculitic disorders (not limited to EGPA), and various lymphoproliferative and myeloproliferative disorders, including T-cell lymphoma and systemic mastocytosis, commonly present with eosinophilic infiltration of the skin and mild to marked peripheral eosinophilia. Extensive skin involvement and significant eosinophilia in a young child or in the context of recurrent or unusual infections should prompt consideration of a primary immunodeficiency, such as dedicator of cytokinesis 8 (DOCK8) deficiency.

Rare idiopathic eosinophilic conditions that predominantly involve the skin and soft tissues include eosinophilic cellulitis (Wells syndrome) and eosinophilic fasciitis. Although varied triggers have been reported, the pathogenesis of eosinophilic cellulitis remains unknown. It can occur in the absence of other clinical manifestations or as part of a systemic eosinophilic disorder. Dermal involvement is common, and typical histopathologic findings include superficial and deep dermal inflammation with histiocytes, lymphocytes, and eosinophils. Eosinophil granule deposition in the dermis can result in characteristic "flame figures" on histopathology. Eosinophilic fasciitis is primarily a disease of the fascia and perimysium, although skin involvement can be present. Antecedent dermal edema followed by skin tightening, induration, and eventual sclerosis is typical. Fascial involvement and fibrosis may result in the characteristic "groove sign," skin tightening, muscle atrophy, and potentially irreversible contractures. Both eosinophilic cellulitis and eosinophilic fasciitis can occur as paraneoplastic syndromes.

Pulmonary Disorders

Eosinophilic conditions affecting the upper and lower airways include various allergic and immunologic disorders, of which eosinophilic asthma and chronic rhinosinusitis with and without nasal polyposis are the most common. Although each of these conditions can occur in isolation, they can also be the presenting symptom of another allergic or immunologic diagnosis, such as aspirin-exacerbated respiratory disease, allergic bronchopulmonary aspergillosis/mycosis, hypersensitivity pneumonitis, or EGPA. Features suggestive (but not diagnostic) of the latter include HE and failure to respond to standard or low-dose topical therapies. Although the role of eosinophils in the pathogenesis of chronic obstructive pulmonary disease is controversial, sputum and blood eosinophilia occur and biologics that target eosinophils have shown efficacy in some clinical trials.

The differential diagnosis of eosinophilia and pulmonary parenchymal disease is very broad and includes allergic, infectious, rheumatologic, neoplastic, and idiopathic disorders. A careful history and routine laboratory testing are useful in narrowing the differential, and computed tomography (CT) imaging of the chest can provide valuable information, especially when characteristic features are present. Nevertheless, bronchoalveolar lavage (BAL) and/or lung biopsy are often necessary for definitive diagnosis. Acute eosinophilic pneumonia (AEP) is a diagnosis of exclusion in a previously healthy patient presenting with fever, cough, and dyspnea of less than 7 days' duration and radiologic evidence of pulmonary infiltrates. There is frequently a history of recent initiation of cigarette smoking or significant smoke, sand, or dust exposure. BAL reveals greater than 25% eosinophils and no evidence of an underlying cause. Although hypoxemia and need for intubation are common, most patients recover, and recurrent episodes are extremely rare. In contrast, chronic eosinophilic pneumonia is a chronic illness most commonly diagnosed in women with a history of asthma. Symptoms are subacute and include cough, fever, weight loss, and dyspnea. Bilateral pleural-based or peripheral infiltrates are characteristic radiologic findings. Moderate blood eosinophilia is frequent, and BAL shows greater than 25% eosinophils.

Pleural fluid eosinophilia is defined by the presence of greater than 10% eosinophils and is a normal response to air or blood in the pleural space in the context of both accidental or iatrogenic trauma (e.g., following thoracotomy or repeated thoracentesis). Other causes include malignancy, especially lung cancer, drug hypersensitivity, pulmonary embolism, and a wide range of infectious pathogens, including helminths, fungi, viruses, and mycobacteria and bacteria (reviewed in Kalomenidis and Light[15]). Rarely, eosinophilic pleural effusions occur in the setting of immunologic disorders, including rheumatoid arthritis, EPGA, and sarcoidosis. The cause of pleural eosinophilia is not identified in up to 25% of cases.

Gastrointestinal Disorders

Gastrointestinal (GI) disorders associated with blood and/or tissue eosinophilia include those where eosinophils play a key role in pathogenesis (e.g., eosinophilic GI disorders [EGIDs], eosinophilic hepatitis, eosinophilic cholecystitis, eosinophilic pancreatitis) and those where eosinophilia is common but of unknown clinical significance (e.g., inflammatory bowel disease [IBD], celiac disease, autoimmune hepatitis). Moreover, eosinophilic infiltration of the GI tract can occur in isolation or as part of a multisystem disorder, including EGPA, *PDGFRA*-associated myeloid neoplasm, and other rare HES variants. Secondary causes of GI eosinophilia are many and include helminth infection, drug hypersensitivity reactions, and immune dysregulatory disorders.

EGIDs are rare disorders with characteristic symptoms and histopathologic evidence of eosinophils in the GI tract in the absence of a secondary etiology.[16] Although eosinophils are believed to play a key role in the pathogenesis of EGIDs, additional cell types, including mast cells, T cells, and fibroblasts, have also been implicated. The nomenclature of EGIDs continues to evolve as diagnostic criteria are developed for EGID involving different segments of the GI tract and/or layers of tissue involvement (e.g., mucosal vs. serosal).

KEY CONCEPTS

- Eosinophilic GI disorders (EGIDs) are rare, chronic/relapsing atopic conditions of the gastrointestinal tract that are diagnosed using a combination of clinical and pathologic features.
- Pediatric EGIDs may be more complex in presentation owing to nutritional deficiencies, growth delay, and need for feeding tubes for enteral nutrition. Developmental delay and behavioral disturbances have been described.
- Uncommon in isolated eosinophilic esophagitis (EoE), marked peripheral eosinophilia frequently accompanies EGID involving the stomach, small bowel, and/or colon, blurring the lines between EGID and systemic hypereosinophilic syndrome (HES) with GI involvement.
- Histologic features of EoE include eosinophilic inflammation with disruption of the epithelial architecture, basal zone hyperplasia, dilated intracellular spaces, and lamina propria fibrosis.
- Although the pathogenesis of EGIDs is incompletely understood, non-IgE-mediated reactions to food antigens play a role in the pathogenesis of EoE and EG in many patients.
- Treatment is generally indicated in EoE irrespective of symptom severity as chronic inflammation in the esophagus can lead to fibrosis, stricture formation, and food impactions. Choice of treatment of non-EoE EGIDs is generally guided by the severity of symptoms and presence of complications.
- Current therapeutic options for EGIDs include dietary interventions and swallowed and systemic glucocorticoids. Several promising biologic therapies are currently in clinical trials.

Eosinophilic esophagitis (EoE) is a chronic, immune-mediated disorder defined by the presence of greater than 15 eosinophils per high power field (hpf) in the mucosa of the esophagus and food-related symptoms. The current incidence of EoE is estimated at 5 to 10 cases per 100,000 individuals with a prevalence of 0.5 to 1 case per 1000 individuals in the United States and a male predominance. The prevalence of EoE is increased in family members of affected patients, and monogenic causes have been described (e.g., Loeys-Dietz syndrome [LDS] and DOCK8). Children commonly present with food refusal, failure to thrive, abdominal pain, and vomiting or regurgitation, whereas adults are more likely to complain of nausea, vomiting, dysphagia, food impaction, and chest pain. IgE-mediated food allergy and atopic disease are frequent concomitant diagnoses and peripheral eosinophilia, when present, is generally mild unless other GI segments are also involved. Endoscopy may reveal normal-appearing mucosa, edema, erythema, longitudinal furrows, plaques, and exudates, or a "crepe paper" appearance (Fig. 45.3). Trachealization and strictures can be seen in more advanced stages of the disease. Histologic assessment with adequate sampling is recommended even when the mucosa appears grossly normal. Tissue eosinophilia and other histologic features (EoEHSS) are useful for diagnosis.[17] Esophageal eosinophilia due to gastroesophageal reflux, achalasia, Crohn disease, connective tissue disorders, helminth

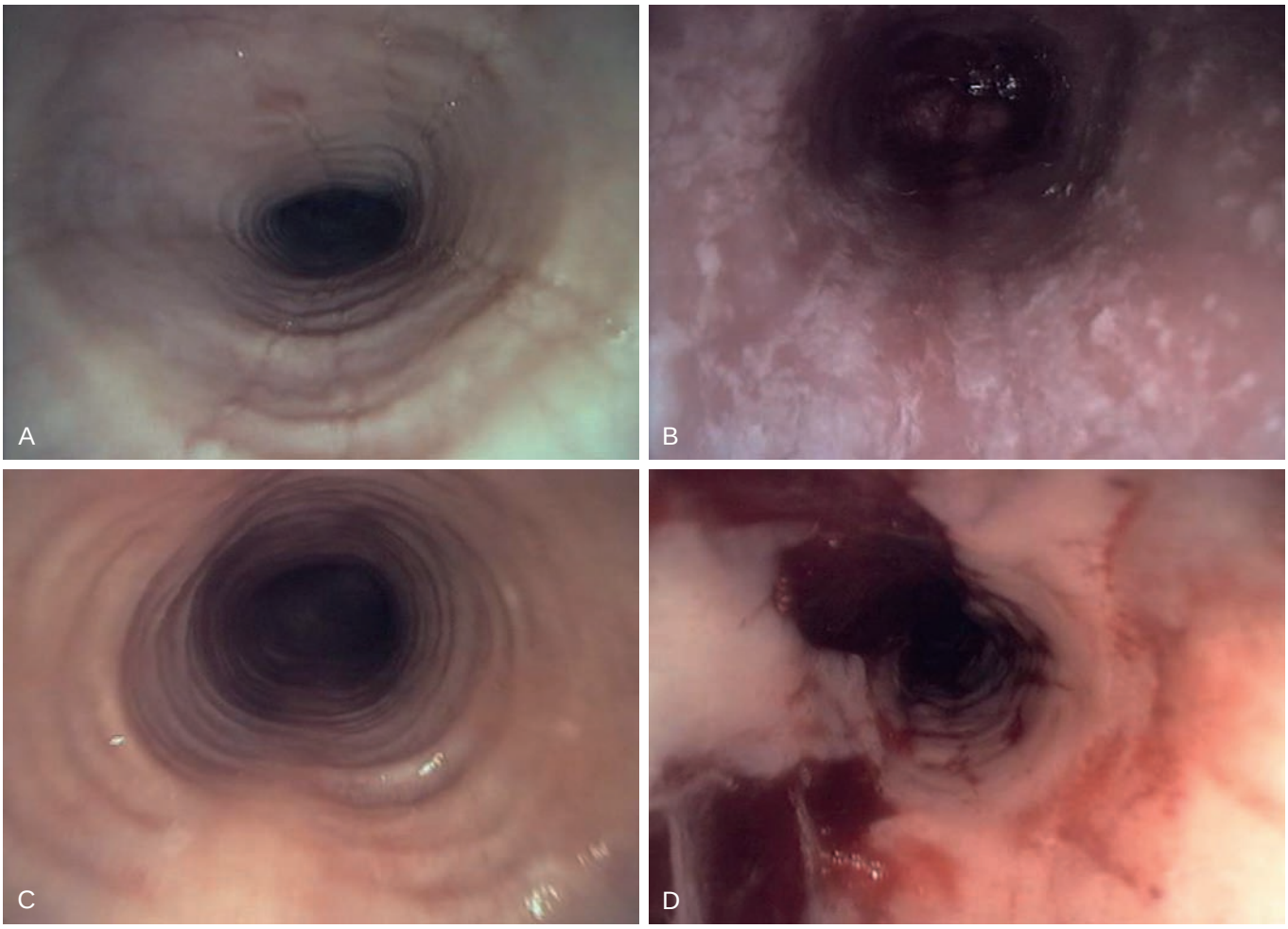


FIG. 45.3 Endoscopic Findings in Eosinophilic Esophagitis. Representative endoscopic pictures of active eosinophilic esophagitis showing loss of vascularity and red furrows (A), whitish exudates (B), corrugated rings (C), and a long-distance stricture with a deep laceration after dilatation (D).

infections, and drug hypersensitivity reactions must be excluded to make a diagnosis of EoE.

Non-EoE EGIDs (*i.e.*, eosinophilic gastritis, enteritis, and colitis) have been less well studied. Estimated prevalences range from 3.3 to 8.4 per 100,000 individuals but are confounded by the lack of consensus diagnostic criteria. Tissue eosinophilia (30 or more eosinophils/hpf in 5 high powered fields in the stomach, greater than 30/hpf in the small bowel, and greater than 60/hpf in the colon) in the absence of IBD or a known secondary cause supports the diagnosis but may be absent in patients with disease restricted to the deep submucosal, muscularis, or serosal layers. Males and females appear to be equally affected. Symptoms reflect the location and extent of intestinal involvement and can include early satiety, nausea, vomiting, abdominal pain, bloating, and diarrhea. Complications include iron deficiency anemia, protein-losing enteropathy, malabsorption, ascites (with serosal involvement), ulceration, and, rarely, perforation. Extraintestinal eosinophilic involvement of the liver, gallbladder, or pancreas has also been reported.

Renal and Genitourinary Disorders

The most common cause of eosinophilic infiltration of the kidney is acute interstitial nephritis (AIN) as a manifestation of drug hypersensitivity or a systemic eosinophilic disorder, such as EGPA or IgG4-related disease. Patients with AIN typically present with renal dysfunction accompanied by eosinophilia, rash, and/or fever. The presence of eosinophiluria was previously considered pathognomonic for this syndrome; however, eosinophils and their granule proteins can also be present in the urine of patients with systemic eosinophilic disease with and without renal abnormalities and in patients with eosinophilic cystitis. Renal vein thrombosis should be considered in hypereosinophilic patients with reduced creatinine clearance in the absence of an active urinary sediment.

Idiopathic eosinophilic cystitis is a rare disorder of unknown etiology that affects all ages, with approximately 20% of cases in children. Patients typically present with frequency, urgency, suprapubic pain, urinary incontinence, and/or retention. Peripheral eosinophilia and elevated IgE are common. This and an association with EGID in some cases suggest that idiopathic eosinophilic cystitis may be allergen-driven. Secondary causes of eosinophilic cystitis include drug hypersensitivity reactions, especially to medications delivered intravesically, urinary schistosomiasis, and bladder cancer. Rarely, eosinophilic cystitis occurs as a manifestation of idiopathic HES.

Cardiac Disorders

Eosinophilic cardiac involvement[18] is most commonly a manifestation of HES, where it occurs in up to 20% of patients and is an important cause of mortality. Etiologies of particular concern, because of the therapeutic implications, are eosinophilic myeloid neoplasms, especially those associated with mutations in *PDGFRA* and *JAK2*, helminth infection, EGPA, and DRESS. Although endomyocardial fibrosis (EMF) with valvular involvement, apical thrombus formation, and a high risk of thromboembolic events is characteristic (Fig. 45.4), additional manifestations include eosinophilic myocarditis, eosinophilic pericarditis, and eosinophilic coronary arteritis (in EGPA). Myocardial disease can be patchy, and tissue biopsies may not demonstrate intact eosinophils, especially in EMF. Immunohistochemical staining for granule proteins can be helpful in this regard. Eosinophilic infiltration of cardiac tissues has been described in other conditions, including viral myocarditis, sarcoid, and amyloid, but the role of eosinophils in the pathogenesis of these conditions remains controversial.

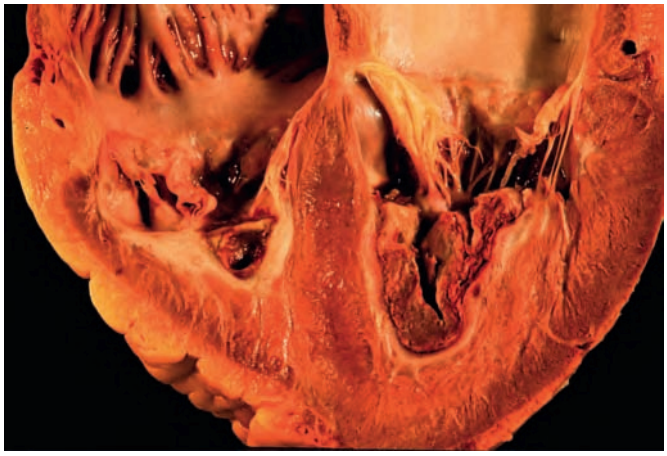


FIG. 45.4 Eosinophilic Endomyocardial Disease.

Neuromuscular Disorders

Eosinophilic involvement of the nervous system can manifest as peripheral neuropathy, meningitis, or central nervous complications. Eosinophil-related peripheral neuropathy is most frequently seen in the context of HES, including EGPA. Clinical manifestations range from mild sensory abnormalities to severe sensorimotor impairment involving multiple nerves (*i.e.*, mononeuritis multiplex). Eosinophilic meningitis may be due to drug hypersensitivity, infection (e.g., angiostrongyliasis, schistosomiasis, and cryptococcosis), or hematologic malignancy. Diagnosis requires the presence of greater than 10 eosinophils/μL (or 10% or more eosinophils) in the cerebrospinal fluid in a patient with clinical signs of meningitis. The most common central nervous complications of eosinophilia are thromboembolic, although rare cases of encephalitis have been reported.

The differential diagnosis of inflammatory eosinophilic myositis is broad and includes drug-induced myopathy, infectious (typically parasitic) or toxin-associated myositis, autoimmune (e.g., dermatomyositis or polymyositis), and genetic causes (e.g., calpain-3-associated limb-girdle myopathies). When no cause is identified, primary eosinophilic myositis should be considered. There are several clinically defined phenotypes of primary eosinophilic myositis that can be differentiated using a combination of clinical, histopathologic, and radiologic findings (reviewed in Selva-O'Callaghan et al.[19]); however, classical eosinophilic myositis presents with pain, overlying erythema, weakness, and tenderness with muscle biopsies demonstrating myonecrosis and endomysial inflammation. Eosinophilic fasciitis may be associated with myopathy and perimyositis with elevated aldolase but does not typically cause true myositis or muscle atrophy (except as a result of disuse in the setting of contractures).

Drug and Toxin-Related Eosinophilia

Drug-related eosinophilia is one of the most common causes of peripheral eosinophilia and can occur in response to any

prescription or over-the-counter medication, nutritional supplement, or herbal remedy. Manifestations of drug hypersensitivity range in severity from asymptomatic eosinophilia to the potentially life-threatening systemic involvement seen in DRESS, a nonimmediate hypersensitivity that typically presents with fever greater than 38°C, skin rash, lymphadenopathy, organ involvement, atypical lymphocytosis, and eosinophilia at least 3 weeks after drug initiation. Whereas many drugs can cause DRESS, the most common offending agents include anti-epileptics, reverse transcriptase inhibitors, dapsone, minocycline, and allopurinol. Skin, liver, and kidneys are most often affected, but lung and heart abnormalities are reported. Drug-specific T effector cells, viral reactivation, especially human herpes virus-6 (HHV-6), and decreased regulatory T cells have all been implicated in the pathogenesis of DRESS. Long-term-sequelae can include autoimmune disorders, such as thyroiditis, lupus, and diabetes.

CLINICAL PEARLS

Drug Hypersensitivity

- Eosinophilia and eosinophil-related complications can occur in response to any drug (prescription or over-the-counter) or dietary supplement.
- The clinical manifestations of drug hypersensitivity range from asymptomatic eosinophilia to severe life-threatening complications.
- Some clinical presentations may suggest a particular class of offending agent (e.g., DRESS and anti-epileptic medications, eosinophilia-myalgia syndrome, and tryptophan), whereas others (e.g., rash, pulmonary infiltrates) are less specific.
- Eosinophilia may not resolve for many months after the causative drug is discontinued and glucocorticoids or other immunosuppressive therapy may be needed to treat eosinophil-associated clinical manifestations.

Some drugs are associated with characteristic clinical syndromes (e.g., tetracyclines and eosinophilic hepatitis, L-tryptophan, and eosinophilia-myalgia syndrome); whereas others cause pleiomorphic clinical manifestations (*e.g.*, checkpoint inhibitors, cephalosporins). Yet others drive eosinophilia as a consequence of their mechanism of action, including some cytokine therapies (IL-2, GM-CSF) and monoclonal antibodies (e.g., dupilumab and vedolizumab).

Infectious and Immune Causes of Eosinophilia

A wide range of infectious agents, including parasites, fungi, viruses, and bacteria, can cause peripheral and/or tissue eosinophilia (reviewed in O'Connell and Nutman[20]). Since many of these are geographically or environmentally restricted, an appropriate exposure history and assessment of endemicity and travel is essential. The clinical picture is sometimes strongly suggestive of a particular diagnosis (e.g., periorbital edema and myositis in a patient with trichinosis or a subconjunctival worm in an African patient with loiasis). More commonly, clinical manifestations are nonspecific but can help narrow the differential diagnosis. The pattern of organ involvement and duration and degree of eosinophilia are particularly helpful in this regard.

Helminths are the most common cause of infection-related eosinophilia worldwide and can be broadly categorized into those that cause acute eosinophilia and those that cause persistent or chronic eosinophilia. Acute eosinophilia is a common early feature of infection with parasites that migrate through the lungs, including *Ascaris*, hookworm, and schistosomes. Acute marked eosinophilia can also be seen with serosal membrane involvement (*i.e.*, echinococcal cyst rupture, eosinophilic meningitis due to *Angiostrongylus* infection) and in hyperresponsive syndromes, such as tropical pulmonary eosinophilia (a rare presentation of lymphatic filariasis) and eosinophilic enteritis due to *Anisakis*. Among the multitude of helminths that cause chronic eosinophilia, *Strongyloides* is of particular concern due to its worldwide distribution, ability to persist for decades in an infected host, and propensity to disseminate with glucocorticoid administration. Hypersensitivity reactions to ectoparasite infestations, such as scabies and myiasis, can present with varying degrees of peripheral eosinophilia and should be considered in patients with consistent skin findings. Conversely, protozoan infections are almost never the cause of eosinophilia with the exception of acute watery diarrhea due to *Cystoisospora belli* infection and eosinophilic myositis due to *Sarcocystis* infection (restricted to southeast Asia). Whereas mild to moderate peripheral and/or tissue eosinophilia is frequently seen in a variety of cutaneous and systemic fungal infections, marked eosinophilia is characteristic of only a few, of which coccidioidomycosis is most common. Viral infections cause eosinopenia with rare exceptions, of which the most notable is HIV infection.

Primary disorders of immune dysregulation or immunodeficiency can present with peripheral eosinophilia that may or may not be associated with eosinophil-related clinical manifestations (reviewed in Williams et al.[21]). Onset at birth or in early childhood, characteristic features, and/or recurrent or atypical infections should prompt consideration of an underlying genetic cause. Some immunodeficiencies typically present very early in life with peripheral eosinophilia and eosinophilic manifestations, such as the exfoliative dermatitis seen in Ommen syndrome; others, such as autosomal dominant hyper-IgE syndromes (HIES) and DOCK8 deficiency, are more often diagnosed later in childhood in the setting of recurrent infections, elevated IgE levels, and HE with eosinophilic skin and GI involvement. Other disorders characterized by peripheral eosinophilia and eosinophilic tissue involvement include phosphoglucomutase 3 deficiency (PGM3); adenosine deaminase-severe combined immunodeficiency (ADA-SCID); immunodysregulation, polyendocrinopathy, enteropathy, X-linked syndrome (IPEX), and IPEX-like disorders; LDS; severe dermatitis, multiple allergies and metabolic wasting (SAM); and Wiskott-Aldrich syndrome (WAS). Asymptomatic peripheral eosinophilia has also been reported in up to 10% of patients with autoimmune lymphoproliferative syndrome (ALPS) and is associated with increased mortality.

Eosinophilia is common following immune reconstitution after hematopoietic cell or solid organ transplantation. Causes include drug reactions, infections, acute severe graft-versus-host disease, and in the case of hematopoietic cell transplantation, immune reconstitution itself.

Finally, eosinophilia can be associated with a wide range of rheumatologic and idiopathic inflammatory conditions. In some cases, eosinophils are implicated in the pathogenesis of the clinical manifestations, such as eosinophil-associated myopathies, eosinophilic synovitis, eosinophilic panniculitis, EGPA, and IgG4-associated disease. In others, mild to moderate peripheral eosinophilia of uncertain clinical significance has been reported with some frequency (e.g., sarcoidosis, rheumatoid arthritis, dermatomyositis, and Sjögren syndrome).

GENETIC IMMUNE DISORDERS ASSOCIATED WITH HYPEREOSINOPHILIA

- A wide variety of inherited immune disorders are associated with blood and/or tissue eosinophilia
- Presentation in infancy, recurrent or atypical infections, and syndromic features are clues to an underlying inherited immune cause of eosinophilia
- Disorders commonly associated with eosinophilia include
  - IL6 pathway disorders: autosomal dominant hyper-IgE syndromes (HIES)-STAT3 deficiency
  - Actin dysregulatory disorders: DOCK8 deficiency
  - CID with T-cell receptor abnormalities or associated syndromic features: ARPC1B deficiency (small series), Omenn syndrome, ADA-SCID, ZAP-70 deficiency, TCR α or γ deficiency
  - Conditions of immune dysregulation: STAT5b gain of function
- Disorders in which eosinophilia has been reported in a subset of affected individuals include
  - Autoinflammatory disorders: neonatal-onset multisystem inflammatory disease (NOMID), chronic infantile neurological cutaneous and articular (CINCA), Blau syndrome
  - IL6 pathway disorders: IL6 receptor mutations, autosomal recessive HIES (e.g., ZNF341-deficiency)
  - Actin dysregulatory disorders: Wiskott-Aldrich syndrome, CARMIL2 deficiency syndrome (small series)
  - CID with T-cell receptor abnormalities or associated syndromic features: immunoskeletal dysplasia with neurodevelopmental abnormalities (EXTL3; small series), Comel-Netherton syndrome (*SPINK5* deficiency) and EPHKE syndrome (small series or case reports), ataxia-telangiectasia (rare), PGM3 deficiency (up to 50%)
  - Disorders associated with CARD11 variants
  - Loeys-Dietz syndrome
  - SAM syndrome
  - Combined immunodeficiencies and disorders of T-cell receptor generation: CD40/CD40L deficiency (case reports), CARD9 deficiency
  - Conditions of immune dysregulation: autoimmune lymphoproliferative syndrome (ALPS), JAK1 gain of function, IPEX and IPEX-like syndromes (seen in many affected individuals, counts frequently not reported)

ADA-SCID, Adenosine deaminase-severe combined immunodeficiency; *CADINS*, CID with early-onset asthma, eczema and food allergies, autoimmunity ID with atopic dermatitis; *CARD*, caspase recruitment domain family member; *CARMIL2*, capping protein, Arp2/3, myosin-I linker 2; *DOCK8*, dedicator of cytokinesis 8; *IPEX*, Immunodysregulation, polyendocrinopathy, enteropathy, X-linked syndrome; *PGM3*, phosphoglucomutase 3; *SAM*, severe dermatitis, multiple allergies, and metabolic wasting; *STAT3*, signal transducer and activator of transcription 3; *SPINK5*, serine peptidase inhibitor Kazal type 5; *TCR*, T-cell receptor; *ZAP-70*, zeta-chain-associated protein kinase 70.[21,33]

EGPA, previously known as Churg-Strauss syndrome, is a systemic vasculitic disorder defined in part by blood and/or tissue eosinophilia in the setting of preceding asthma and/or sinus disease.[22] Clinical complications are related primarily to eosinophilic infiltration and vasculitis and include myocarditis and/or pericarditis with resultant cardiomyopathy, peripheral neuropathy and thromboembolic events, migratory pulmonary infiltrates, skin involvement including purpuric and necrotic lesions, renal involvement, and, rarely, tracheal stenosis or optic neuritis. Anti-neutrophil cytoplasmic antibody (ANCA) positivity is often considered a hallmark of the disease; however, there are likely multiple phenotypes of EGPA and greater than 50% of patients with profound eosinophilia and characteristic clinical manifestations are ANCA-negative. Although eosinophil-rich inflammatory infiltrates and granulomatous inflammation in small to medium-sized vessels with vasculitis are diagnostic, these features may be absent in patients receiving glucocorticoid or other systemic therapies.

Hematologic and Neoplastic Causes of Eosinophilia

Eosinophilia can be the presenting manifestation of many acute and chronic lymphocytic neoplasms, including Hodgkin lymphoma, T-cell leukemia/lymphoma, and pre-B-cell acute lymphoblastic leukemia, and is a frequent feature of myeloid neoplasms and myeloproliferative disorders, including acute and chronic eosinophilic leukemia, chronic myelomonocytic leukemia, and systemic mastocytosis. The 2016 WHO classification of myeloid/lymphoid neoplasms with eosinophilia defines the following molecular abnormalities as the most common culprits: *PDGFRA, PDGFRB, FGFR1*, and *PCM1-JAK2*; however, other fusion partners (e.g., *ETV6)* and molecular abnormalities associated with a myeloid picture with eosinophilia (e.g., *JAK2* V16F, and *JAK2* exon 13 mutations, particularly V617F) have also been described.[23] Whereas peripheral eosinophilia is relatively uncommon in solid tumors, eosinophils are frequently present in the tumor microenvironment where they may play a role in the antitumor response.[24] Solid tumors, especially adenocarcinomas, are a rare cause of HES.

Hypereosinophilic Syndromes

HES are a heterogeneous group of disorders defined by AEC greater than or equal to 1500/mm^3 and evidence of eosinophil-mediated clinical manifestations. Although any organ can be affected, dermatologic, GI, and pulmonary manifestations are most common. Life-threatening cardiac and neurologic complications develop in up to 20% of patients over time. Following the landmark paper by Chusid and colleagues in 1984, describing 14 patients with "idiopathic" HES, a number of classification schemes have been proposed in an attempt to incorporate advances in our understanding of the drivers and pathogenesis of eosinophilic disease. The following categories are based on the consensus classification developed by the International Cooperative Working Group on Eosinophil Disorders (ICOG-EO) in 2012 and have proven useful in predicting treatment responses to a number of agents, including glucocorticoids, imatinib, and targeted biologics.[25]

Myeloid Hypereosinophilic Syndromes

Myeloid HES (MHES) is defined as HES due to definite or presumed clonal eosinophilia. Clinical features suggestive of MHES include markedly increased dysplastic eosinophils and EoP on peripheral smear, anemia or erythrocytosis, thrombocytopenia, increased serum vitamin B_{12} and/or tryptase levels, splenomegaly, and characteristic findings on bone marrow examination. The most common etiology of MHES, accounting for greater than 80% of cases in most series, is an interstitial deletion in chromosome 4 leading to the imatinib-sensitive fusion gene *FIP1L1-PDGFRA*. This mutation is almost exclusively seen in males and likely accounts for the male predominance and poor prognosis of "idiopathic" HES in early series. Clonal abnormalities involving *PDGFRB, JAK2, FGFR1*, and *KIT*, and chromosomal deletions and translocations account for most of the remaining cases of MHES with an identified cause.

Lymphoid Hypereosinophilic Syndromes

Lymphoid or "lymphocytic" HES (LHES) is defined by the presence of a phenotypically aberrant or clonal T-cell population

producing Type 2 cytokines that drive reactive HE. The most commonly identified aberrant T-cell population is $CD3^-$ $CD4^+$, followed by $CD3^+$ $CD4^+$ $CD8^+$ and $CD3^+$ $CD4^-$ $CD8^-$. Clonality can frequently, but not always, be demonstrated by assessment of the γ chain of the T-cell receptor by polymerase chain reaction (PCR) and/or Vβ chain assessment by flow cytometry. Although patients with LHES can present with a wide range of clinical manifestations, skin involvement (e.g., erythroderma, pruritic dermatitis, poikiloderma), lymphadenopathy, and elevated IgE are characteristic. Exclusion of lymphoma is important at presentation and periodically thereafter since progression to lymphoma has been reported in 5% to 30% of patients, sometimes decades after the initial diagnosis of LHES. Aberrant $CD3^-$ $CD4^+$ T-cell populations have been described in patients with asymptomatic hypereosinophilia (HEus) and episodic angioedema with eosinophilia (Gleich syndrome), although their contribution to the underlying disease process is unknown.

Overlap Hypereosinophilic Syndromes

By far the most confusing clinical category is overlap HES, which includes single-organ hypereosinophilic disorders (e.g., eosinophilic pneumonia, EGID, eosinophilic fasciitis) and defined syndromes associated with HE and eosinophilic manifestations (e.g., EGPA). These disorders overlap in presentation with idiopathic HES but are considered separately due to differences in treatment approaches (see below).

Associated Hypereosinophilic Syndromes

Associated HES is the term used for HES secondary to a known cause, such as drug hypersensitivity, infection, neoplasia, or immunodeficiency, where definitive treatment is directed at the underlying cause.

Familial Hypereosinophilic Syndromes. Hereditary cases of HES are extremely uncommon. Dominant mutations of STAT3 underlie the rare multisystem disorder characterized by recurrent boils, cyst-forming pneumonias, eosinophilia, and extreme elevations of IgE. Autosomal dominant transmission of HES has been mapped to chromosome 5q31-33 in one large kindred with dysregulation of IL5 expression. Familial clustering has been described in EoE; however, with the exception of EoE as part of a Mendelian syndrome (e.g., Loeys-Dietz, SAM), the underlying mechanism of inheritance appears complex. Rare cases of EGPA occurring in two members of the same family have also been reported.

Idiopathic Hypereosinophilic Syndromes. Idiopathic HES is the term used to describe patients with HES who do not fit into any of the categories above. In some cases, patients present with persistent asymptomatic HE without evidence of end-organ manifestations or tissue involvement in the absence of therapy (HEus). Although rare, HEus presents a unique problem as the risk of progression to HES is unknown.

THERAPY OF EOSINOPHIL-ASSOCIATED DISORDERS

General Treatment Approach

The approach to treatment of eosinophilic disorders depends on a variety of factors, including the severity of clinical manifestations, most likely diagnosis, and role of eosinophils in disease pathogenesis.[26] As such, the initial approach to the patient with mild to moderate eosinophilia does not differ from the approach to the patient with HE. In a patient with presumed eosinophil-related manifestations requiring urgent intervention (e.g., active cardiac involvement or thromboembolic events), glucocorticoid therapy is considered first-line therapy unless there is clear evidence of a hematologic malignancy or another disorder that requires rapid therapy directed at the underlying cause. Ivermectin therapy should be administered concurrently if there is a potential history of exposure to *Strongyloides* to prevent glucocorticoid-induced hyperinfection syndrome. Early initiation of other agents (e.g., cytotoxic therapy in EGPA or imatinib in *PDGFR*-associated myeloid neoplasms) depends on the level of suspicion for a particular underlying etiology.

Once the patient is stabilized, and in cases where treatment is not urgently needed, evaluation should focus on assessment of end-organ manifestations and the most likely etiology of the eosinophilia (Table 45.2). If a secondary cause, such as a helminth infection, drug hypersensitivity, or primary immunodeficiency, is identified, this should be treated, and resolution of the eosinophilia confirmed (if appropriate). In some cases, concomitant therapy directed at the eosinophilia may be necessary to manage clinical manifestations prior to definitive treatment (for example, glucocorticoid therapy to treat severe eczema prior to stem cell transplant in a patient with DOCK8 deficiency or pulmonary manifestations in a patient with tropical pulmonary eosinophilia prior to definitive therapy with diethylcarbamazine). At the opposite end of the spectrum, mild to moderate eosinophilia in an asymptomatic patient without evidence of end-organ manifestations can be observed without treatment. The decision to withhold therapy in the setting of HE is more complex and requires additional evaluation.

Whereas the optimal treatment for symptomatic patients with mild to moderate eosinophilia has not been systematically studied (necessitating an individualized approach), the treatment approach for patients with blood and/or tissue HE (*i.e.*, HES) has evolved considerably in recent years due in large part to the availability of new diagnostic tests and novel targeted agents. The most dramatic example of this is certainly the discovery of *FIP1L1-PDGFRA* as a driver of myeloid HES and the dramatic response of these patients to imatinib.

Historically, the most common agents used to treat all forms of HES were, in order of frequency, glucocorticoids, hydroxyurea, and interferon-α. Although these agents have been used for decades, and in the case of glucocorticoids, are highly effective in the short term, many patients become refractory over time and long-term toxicity is significant.[27] Consequently, a wide variety of cytotoxic, immunomodulatory, and immunosuppressive drugs have been tried with varying success. The development of imatinib and its ultimate approval in 2006 for the treatment of HES marked a major shift in approach by providing a targeted therapy for a defined subset of HES patients. More recently, regulatory approval of biologics that target eosinophils for the treatment of asthma, atopic dermatitis, EGPA, and, in the case of mepolizumab (anti-IL5 antibody), HES, is beginning to alter conventional approaches to therapy of many eosinophilic disorders.

Approach to Therapy of Hypereosinophilic Syndromes

There are considerable data to support a role for clinical subtype in determining therapeutic responses in HES as outlined below. Additional factors to consider include medication side effects, comorbid conditions, concomitant medications, and cost and convenience of therapy.

TABLE 45.2 Evaluation of the Patient With Hypereosinophilia

Test	Comment
All Patients with HES	
Complete blood count[a]	
Routine chemistries, including liver function tests[a]	
Quantitative serum immunoglobulin levels, including IgE	
Serum troponin,[a] echocardiogram	If abnormal, cardiac MRI should be considered as this may show characteristic features of eosinophilic involvement; tissue involvement may be patchy limiting the utility of biopsy
Pulmonary function tests[a]	
Chest/abdomen/pelvis CT[a]	To assess for splenomegaly, lymphadenopathy, and occult neoplasms
Bone marrow biopsy, including cytogenetics[a]	Recommended in all patients with AEC >5.0 × 10^9/L and features of M-HES or L-HES. Should be considered in other patients
Biopsies of affected tissues (if possible)[a]	
Other testing as indicated by history, signs, and symptoms	Including parasitic serologies, anti-neutrophil cytoplasmic antibodies, and HIV
Serum tryptase and vitamin B_{12} levels	
FIP1L1/PDGFRA analysis by FISH or RT-PCR	Testing of peripheral blood is sufficient
T- and B-cell receptor rearrangement studies	
Lymphocyte phenotyping by flow cytometry[a]	At a minimum CD3, CD4, and CD8 and CD19 or 20 staining should be performed to assess for aberrant $CD3^-CD4^+$, $CD3^+$ $CD4^+$ $CD8^+$, and $CD3^+CD4^-CD8^-$ populations and B-cell lymphoproliferative disorders
Patients with Features of M-HES	
Additional testing for *BCR-ABL1, PDGFRB, JAK2, FGFR1*, and *KIT* mutations by PCR, FISH, or other methods, as appropriate	Testing should be guided by bone marrow findings
Patients with Evidence of L-HES	
Consider PET scan,[a] lymph node biopsy[a]	
EBV viral load	

AEC, Absolute eosinophil count; *CT*, computed tomography; *EBV*, Epstein–Barr virus; *FISH*, fluorescence in situ hybridization; *HES*, hypereosinophilic syndromes; *HIV*, human immunodeficiency virus *L-HES*, lymphoid HES; *M-HES*, myeloid HES; *MRI*, magnetic resonance imaging: *PET*, positron emission tomography; *RT-PCR*, reverse transcription polymerase chain reaction.

[a]Substantially affected by corticosteroid therapy.

From Klion AD. How i treat hypereosinophilic syndromes. *Blood* 2015;126(9):1089–1077.

Myeloid Hypereosinophilic Syndromes

Prior to the recognition of *FIP1L1-PDGFRA* as a major cause of glucocorticoid-resistant HES, 5-year mortality in patients presenting with HE and myeloid features exceeded 30%. With the availability of imatinib, outcomes have dramatically improved in this group of patients due, in large part, to the near-universal imatinib responsiveness in patients with molecular abnormalities involving *PDGFRA* or *PDGFRB*. Imatinib responses in the absence of a detectable *PDGFR* abnormality vary depending on the series but features suggestive of a primary myeloid disorder (*i.e.*, MHES) strongly predicted response in one series. Side effects of imatinib include transient cytopenias, elevated transaminases, swelling, diarrhea, and muscle cramps. Pre-treatment with glucocorticoids is recommended for patients with evidence of eosinophilic cardiac involvement due to rare reports of cardiac necrosis in this population. Second-generation tyrosine kinase inhibitors, such as nilotinib, have shown efficacy in the setting of imatinib intolerance. Although there are relatively little data addressing treatment duration, cures have been reported and a large retrospective study of *PDGFRA*-positive patients found that length of therapy-induced remission prior to drug discontinuation was a major predictor of relapse-free survival.[28]

Whereas therapies targeting other molecular abnormalities, including mutations and translocations in *JAK2*, *FGFR1*, and *KIT*, are currently available or in development, data in patients presenting with HES are limited. Other therapies that have shown some efficacy in MHES include hydroxyurea, interferon-α, cytarabine, and stem cell transplantation.

Idiopathic Hypereosinophilic Syndromes

Systemic glucocorticoids remain first-line treatment for IHES, although rare patients do not respond, and many patients require moderate to high doses and ultimately require steroid-sparing therapies. Until recently, hydroxyurea and interferon-α were the preferred second-line steroid-sparing agents, with efficacy rates in the 30% to 40% range, followed by a wide range of cytotoxic and immunosuppressive agents that have been tried with varied success and considerable toxicity. The recent approval of mepolizumab for HES treatment and promising data from a phase 2 trial of benralizumab have already begun to change this paradigm. Although a number of questions remain, including the predictors of response and relapse, available data suggest that biologics targeting IL-5 or its receptor are safe, well-tolerated, and effective in a majority of patients with IHES.

Lymphoid Hypereosinophilic Syndromes

The general approach to LHES is similar to that for IHES, although patients with LHES are more likely to require higher doses of glucocorticoids. Second-line therapies are typically directed at suppression of T lymphocytes and/or IL-5 with interferon-α at the top of the list due to its multifaceted effects on clonal T-cell numbers and cytokine production. Other agents that have shown some efficacy in case reports and small series include cyclosporine, mycophenolate mofetil, alemtuzumab, and *JAK* inhibitors. Although mepolizumab is effective in controlling symptoms in LHES patients, it was significantly less effective in suppressing AEC to the normal range in LHES patients in the phase 2 study and does not have a direct effect on the clonal population.

OVERLAP HYPEREOSINOPHILIC SYNDROMES

A general guide to the treatment of overlap HES is precluded by the number and diversity of conditions included in this category, each of which requires an individualized approach. Consequently, for the purpose of this chapter, discussion is limited to two of the more common diagnoses encountered by subspecialists in allergy and immunology: namely. EGID and EGPA.

Eosinophilic Gastrointestinal Disorders

Current treatment modalities for EoE include dietary approaches, pharmacologic therapy, and endoscopic management (*i.e.*, food disimpaction and dilation of strictures).[29] Consequently, an interdisciplinary approach involving allergists, gastroenterologists, registered dieticians, and in some cases psychologists with experience in food-related disorders, is most effective. Since clinical and histologic improvement may be discordant in EoE and persistent inflammation can lead to long-term sequelae, including fibrosis and stricture formation, endoscopy remains the gold standard for evaluation of treatment response. Less-invasive assessment tools, such as the Esophageal String Test™, have been used in research settings and are in clinical development.

Dietary elimination using a hypoallergenic formula has been shown to induce histopathologic remission and symptomatic improvement in both children and adults. Consequently, dietary interventions are frequently initiated as first-line treatment. Empiric approaches include six-food elimination diets (SFED; milk, gluten, egg, soy, peanut, tree-nut, fish, and shellfish) and single- or step-up elimination approaches starting with the most common culprits (2-FED; milk and gluten) with extension to four foods (4-FED; milk, gluten, egg, and legumes) if the 2-FED diet is unsuccessful.[30] Data directly comparing the different dietary approaches are not currently available, and targeted testing with specific IgE is not helpful in identifying culprit foods in the absence of IgE-mediated food allergy in this patient population. Since the feasibility of an initial dietary approach and continuation as maintenance therapy requires significant resources and effort, shared decision making with the patient and family should be implemented when appropriate.

Pharmacologic interventions include acid suppression with proton pump inhibitors, topical (swallowed) steroids, and novel biologics. Acid suppression with proton-pump inhibitors leads to clinical improvement in some patients but frequently fails to induce histologic remission in EoE. In contrast, fluticasone and off-label modification of preparations of standard delayed release or liquid preparations of glucocorticoids used in asthma or IBD have shown efficacy in reducing dysphagia scores and peak eosinophil counts. Inhaled fluticasone can be swallowed or delayed-release capsules (e.g., Entecort) can be opened and enteric-coated granules crushed, made into a slurry or liquid preparation in thickeners of various sorts (e.g., sucralose or honey), and swallowed on an empty stomach. New delivery methods for oral suspensions and ready-to-use dispersible tablets for direct delivery to the esophagus are under investigation.

Although biologic therapies that target eosinophil-active cytokines, their receptors, or eosinophils themselves are in development for EoE, none are currently Food and Drug Administration (FDA)-approved for this indication.

Clinical trials using anti-IL-5 antibodies (mepolizumab and reslizumab) have been generally disappointing despite reduction in tissue eosinophil levels. The reasons for this are likely multifactorial and include study design, incomplete tissue depletion, and irreversible structural changes. A number of other biologics, including benralizumab, an afucosylated antibody to IL-5 receptor that efficiently depletes eosinophils in the blood and tissue through antibody-dependent cell cytotoxicity; dupilumab, an antibody to IL-4 receptor α that reduces eosinophil tracking to the tissue; and lirentelimab (AK002), an afucosylated antibody to Siglec-8 that both depletes eosinophils and inhibits mast cell degranulation, are currently in phase 3 trials after promising results in smaller studies.

There are currently no consensus guidelines for treatment of non-EoE EGIDs. Solitary gastric involvement is frequently treated in the same way as EoE. Patients with severe disease or more extensive GI involvement often require treatment with systemic glucocorticoids or other immunosuppressive agents. Phase 3 clinical trials of benralizumab and lirentelimab for eosinophilic gastritis and duodenitis are in progress. Treatment options for eosinophilic colitis remain largely unexplored.

Eosinophilic Granulomatosis with Polyangiitis

Whereas EGPA treatment guidelines have been established by the EGPA Consensus Task Force,[31] the importance of suppressing eosinophil counts to normal ranges and the most appropriate monitoring parameters for EGPA remain controversial. Glucocorticoids are recommended as first-line treatment and can be used as monotherapy to induce remission in patients with mild disease. When life-threatening or severe clinical manifestations are present, additional immunosuppressive agents (e.g., cyclophosphamide) are recommended to induce remission. This has typically been followed by maintenance therapy with azathioprine or methotrexate, although this is likely to change with the recent regulatory approval of mepolizumab (300 mg subcutaneously) for EGPA treatment.[32] Phase 3 trials are ongoing with rituximab (anti-CD20 antibody) and benralizumab, both of which have shown efficacy in preventing relapse in case reports and small open-label studies in EGPA.

Familial Eosinophilia

Familial EGID, EGPA, or symptomatic HES should be treated no differently than sporadic cases. The role for preemptive treatment in asymptomatic patients with familial eosinophilia has not been established. These patients should, however, be monitored for development of clinical manifestations.

HEus

The most appropriate approach to the asymptomatic patient with HE without clinical manifestations remains unclear, as predictors of progression have not been identified. At a minimum, such patients should undergo evaluation, including bone marrow examination, to exclude an occult myeloid neoplasm (MHES) requiring treatment. A number of factors, including the level of the AEC, potential side effects of treatment, and comfort level of the patient and physician should be considered in deciding whether to treat empirically.

ON THE HORIZON

- A better understanding of the underlying mechanisms that drive eosinophilia in varied eosinophilic disorders is needed for more accurate diagnosis and treatment of eosinophil-associated disorders.
- With the increasing availability of therapeutic agents that target eosinophilia, biomarkers and patient-reported outcome tools that correlate with eosinophilic disease activity are paramount in assessing the comparative efficacy of different treatments.
- Elucidating the role of eosinophils in homeostatic processes is a key factor in predicting the long-term effects of eosinophil-depleting therapies in humans.

REFERENCES

1. Lee JJ, Jacobsen EA, McGarry MP, et al. Eosinophils in health and disease: the LIAR hypothesis. *Clin Exp Allergy*. 2010;40(4):563–575.
2. Gleich GJ, Klion AD, Lee JJ, Weller PF. The consequences of not having eosinophils. *Allergy*. 2013;68(7):829–835.
3. Fulkerson PC, Rothenberg ME. Eosinophil development, disease involvement, and therapeutic suppression. *Adv Immunol*. 2018;138:1–34.
4. Lee JJ, Dimina D, Macias MP, et al. Defining a link with asthma in mice congenitally deficient in eosinophils. *Science*. 2004;305(5691):1773–1776. 17.
5. Mori Y, Iwasaki H, Kohno K, et al. Identification of the human eosinophil lineage-committed progenitor: revision of phenotypic definition of the human common myeloid progenitor. *J Exp Med*. 2009;206(1):183–193.
6. Farahi N, Loutsios C, Simmonds RP, et al. Measurement of eosinophil kinetics in healthy volunteers. *Methods Mol Biol*. 2014;1178:165–176.
7. Klion AD, Ackerman SJ, Bochner BS. Contributions of eosinophils to human health and disease. *Annu Rev Pathol*. 2020;15:179–209.
8. Salter BMA, Smith SG, Mukherjee M, et al. Human bronchial epithelial cell-derived factors from severe asthmatic subjects stimulate eosinophil differentiation. *Am J Respir Cell Mol Biol*. 2018;58(1):99–106.
9. Melo RCN, Weller PF. Contemporary understanding of the secretory granules in human eosinophils. *J Leukoc Biol*. 2018;104(1):85–93.
10. Grozdanovic MM, Doyle CB, Liu L, et al. Charcot-Leyden crystal protein/galectin-10 interacts with cationic ribonucleases and is required for eosinophil granulogenesis. *J Allergy Clin Immunol*. 2020;146(2):377–389. e10.
11. Melo RCN, Weller PF. Unraveling the complexity of lipid body organelles in human eosinophils. *J Leukoc Biol*. 2014;96(5):703–712.
12. Weller PF, Spencer LA. Functions of tissue-resident eosinophils. *Nat Rev Immunol*. 2017;17(12):746–760.
13. Mesnil C, Raulier S, Paulissen G, et al. Lung-resident eosinophils represent a distinct regulatory eosinophil subset. *J Clin Invest*. 2016;126(9):3279–3295. 1.
14. Leiferman KM, Peters MS. Eosinophil-related disease and the skin. *J Allergy Clin Immunol Pract*. 2018;6(5):1462–1482. e6.
15. Kalomenidis I, Light RW. Eosinophilic pleural effusions. *Curr Opin Pulm Med*. 2003;9(4):254–260.
16. Pesek RD, Reed CC, Muir AB, et al. Increasing rates of diagnosis, substantial co-occurrence, and variable treatment patterns of eosinophilic gastritis, gastroenteritis, and colitis based on 10-year data across a multicenter consortium. *Am J Gastroenterol*. 2019;114(6):984–994.
17. Collins MH, Martin LJ, Alexander ES, et al. Newly developed and validated eosinophilic esophagitis histology scoring system and evidence that it outperforms peak eosinophil count for disease diagnosis and monitoring. *Dis Esophagus*. 2017;30(3):1–8.
18. Ogbogu PU, Rosing DR, Horne MK. Cardiovascular manifestations of hypereosinophilic syndromes. *Immunol Allergy Clin North Am*. 2007;27(3):457–475.
19. Selva-O'Callaghan A, Trallero-Araguás E, Grau JM. Eosinophilic myositis: an updated review. *Autoimmun Rev*. 2014;13(4–5):375–378.
20. O'Connell EM, Nutman TB. Eosinophilia in infectious diseases. *Immunol Allergy Clin North Am*. 2015;35(3):493–522.
21. Williams KW, Milner JD, Freeman AF. Eosinophilia associated with disorders of immune deficiency or immune dysregulation. *Immunol Allergy Clin North Am*. 2015;35(3):523–544.
22. Khoury P, Grayson PC, Klion AD. Eosinophils in vasculitis: characteristics and roles in pathogenesis. *Nat Rev Rheumatol*. 2014;10(8):474–483.
23. Shomali W, Gotlib J. World Health Organization-defined eosinophilic disorders: 2019 update on diagnosis, risk stratification, and management. *Am J Hematol*. 2019;94(10):1149–1167.
24. Grisaru-Tal S, Itan M, Klion A, Munitz A. A new dawn for eosinophils in the tumour microenvironment. *Nat Rev Cancer*. 2020;20(10):594–607.
25. Valent P, Klion AD, Horny H-P, et al. Contemporary consensus proposal on criteria and classification of eosinophilic disorders and related syndromes. *J Allergy Clin Immunol*. 2012;130(3):607–612. e9.
26. Klion AD. How I treat hypereosinophilic syndromes. *Blood*. 2015;126(9):1069–1077.
27. Ogbogu PU, Bochner BS, Butterfield JH, et al. Hypereosinophilic syndrome: a multicenter, retrospective analysis of clinical characteristics and response to therapy. *J Allergy Clin Immunol*. 2009;124(6):1319–1325. e3.
28. Rohmer J, Couteau-Chardon A, Trichereau J, et al. Epidemiology, clinical picture and long-term outcomes of FIP1L1-PDGFRA-positive myeloid neoplasm with eosinophilia: data from 151 patients. *Am J Hematol*. 2020;95(11):1314–1323.
29. Hirano I, Chan ES, Rank MA, et al. AGA institute and the joint task force on allergy-immunology practice parameters clinical guidelines for the management of eosinophilic esophagitis. *Gastroenterology*. 2020;158(6):1776–1786.
30. Molina-Infante J, Lucendo AJ. Dietary therapy for eosinophilic esophagitis. *J Allergy Clin Immunol*. 2018;142(1):41–47.
31. Groh M, Pagnoux C, Baldini C, et al. Eosinophilic granulomatosis with polyangiitis (Churg-Strauss) (EGPA) Consensus Task Force recommendations for evaluation and management. *Eur J Intern Med*. 2015;26(7):545–553.
32. Wechsler ME, Akuthota P, Jayne D, et al. Mepolizumab or placebo for eosinophilic granulomatosis with polyangiitis. *N Engl J Med*. 2017;376(20):1921–1932.
33. Navabi B, Upton JEM. Primary immunodeficiencies associated with eosinophilia. *Allergy Asthma Clin Immunol*. 2016;12:27.

46

Urticaria, Angioedema, and Anaphylaxis

Elena Borzova and Clive E.H. Grattan

Urticaria is a common skin disease that may be associated with a substantial disease burden.[1] Recent advances in our knowledge and understanding of its pathophysiology offer new diagnostic and treatment approaches for many patients.

DEFINITION

Urticaria is characterized by itchy wheals, angioedema, or both.[2] Urticarial lesions resulting from localized edema of the upper dermis are called *wheals* (Fig. 46.1), whereas acute reversible swelling of the lower dermis, subcutaneous and submucosal tissues is known as *angioedema* (Fig. 46.2).

KEY CONCEPTS

Definition of Urticaria

- Urticaria is an illness characterized by wheals, angioedema, or both.
- Wheals are superficial swellings of the skin: they are pale and itchy with surrounding redness when they appear and then become pink before fading.
- Angioedema is a deeper swelling below the skin or mucosa that usually lasts longer than wheals and may be painful rather than itchy.
- Wheals may occur with or without angioedema in urticaria.
- Acute urticaria lasts less than 6 weeks.[2] It is common and often precipitated by viral infections. Urticaria caused by drugs and foods falls in this category and here the cause is usually clear from the history.
- Chronic urticaria lasts 6 or more weeks with continuous disease activity.[2] A cause may not be found.
- When angioedema occurs without wheals, C1 esterase inhibitor deficiency should be excluded.
- Neonatal urticarial rash raises the possibility of a hereditary autoinflammatory disease or other rare genetic syndromes.

EPIDEMIOLOGY

Urticaria is common, affecting up to 20% of the general population at least once over their lifetime. The lifetime prevalence of chronic urticaria (CU) of all types is estimated at 1.4% and point prevalence at 0.7%.[3]

Acute urticaria mainly affects young adults, patients with atopic diseases, and has an obvious female preponderance, whereas CU is more common in adults, affecting mainly middle-aged women.[2,3] By definition, acute urticaria resolves within 6 weeks, whereas CU may last for years, rarely for more than 10 years. In adults, CU resolves spontaneously in 35% to 57% of patients within 1 to 1.2 years from the onset of the symptoms. In children, annual resolution rate of CU is approximately 10%.[4]

GENETICS

Although there is no indication of Mendelian inheritance, the frequency of chronic spontaneous urticaria (CSU) is higher in first-degree relatives of patients. CSU was linked to several HLA risk alleles including HLA class I (e.g., Bw4, B14, B44) and HLA class II (e.g., DRB1*01, DRB1*04, DRB1*0901, DRB1*12, DQ1, DQB1*0302).[5] Genetic variation in histamine-related genes, including *FCERI* and *HNMT*, and leukotriene-related genes, such as *ALOX5*, *LTC4S* and the PGE2 receptor gene *PTGER4*, were reported to be linked to CU. Additionally, variation in genes associated with autoimmunity such as *PTPN22* appeared to be relevant to CSU.[6]

Rarely, cold-induced urticarial lesions can be a cutaneous manifestation of several monogenic disorders such as *PLCG2*-associated antibody deficiency and immune dysregulation (PLAID) syndrome[7] and cryopyrin-associated periodic syndromes (CAPS),[8] including newly described FXII-associated cold autoinflammatory syndrome (FACAS).[9]

CLINICAL PATTERNS

Urticaria is classified by the duration of continuous activity into acute (less than 6 weeks) or chronic (6 weeks or more).[2] It may be spontaneous and/or inducible. Inducible urticarias may be triggered by mechanical, thermal, or other stimuli.

ETIOPATHOPHYSIOLOGY

Although many aspects of the pathophysiology of urticaria remain unclear, our understanding has advanced considerably over the last two decades (Table 46.1).

Mast Cell–Dependent Mechanisms

Skin mast cells (MCs) are key players in the pathogenesis of urticaria. They are predominantly located around the small blood vessels and lymphatic vessels, as well as around peripheral nerves. The MC density is greatest at distal body areas, including the face, hands, and feet. A threefold increase in the density of cutaneous MCs in the lesional skin of CSU patients has been reported.[10]

Human skin MCs contain preformed mediators in their granules, including histamine, proteases (tryptase, chymase), and heparin. They express many membrane receptors, including high-affinity immunoglobulin E (IgE), low-affinity IgG receptors (FcγRIII), and the inhibitory receptor Siglec-8. Unlike MCs elsewhere (e.g., lung and gastrointestinal [GI] tract), skin

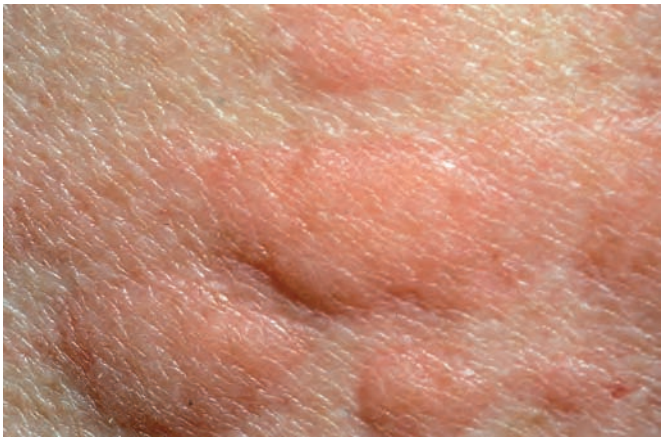


FIG. 46.1 Spontaneous Wheals in Severe Spontaneous Urticaria Showing Superficial Pink Swellings with Pale Edematous Centers.

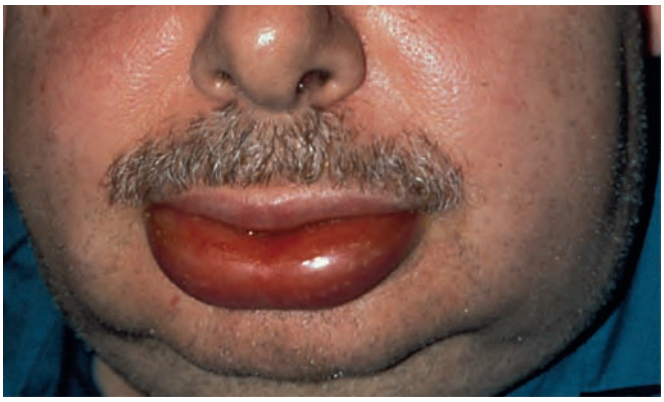


FIG. 46.2 Angioedema of the Mouth in Acquired C1 Esterase Inhibitor Deficiency.

TABLE 46.1 Etiopathogenesis of Urticaria

Acute
Idiopathic
Infection-related
Allergic (mediated by specific IgE)
Chronic
Idiopathic
Autoimmune (IgG against IgE or FcεRI)
Infection-related
Drug-induced
Diet-related

MCs have high expression of complement C5a receptors and the neuropeptide receptors MRGPRX2.[11]

Skin MC activation is central to the pathophysiology of CU. MCs can be activated by a variety of immunological and nonimmunological triggers (Fig. 46.3). Immunological and nonimmunological pathways of MC activation are characterized by distinct patterns of mediator release. In CSU, immunological activation of MCs is thought to occur by cross-linking of high-affinity IgE receptors (FcεRI) by autoantigens binding specific IgE[12] or by MC activating autoantibodies (anti-FcεRIα or anti-IgE antibodies). Histamine release (HR) peaks at 5 to 10 minutes, followed by de novo synthesis of lipid-derived mediators (leukotriene C_4 [LTC_4] and prostaglandin D_2 [PGD_2]) and cytokines/chemokines. In contrast, nonimmunological stimulation of MCs by neuropeptides, opiates, or C5a leads to rapid HR within 15 to 20 seconds, without generation of eicosanoids or cytokines.[12]

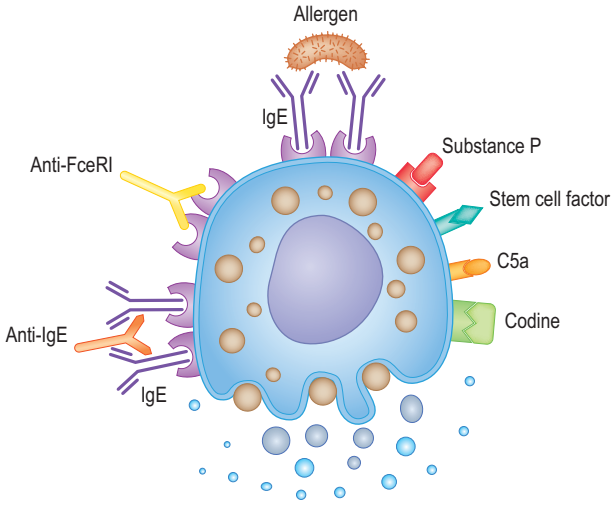


FIG. 46.3 Schematic Representation of a Mast Cell or Basophil. Activation of the immunoglobulin E (*IgE*) receptor by cross-linking with immunological stimuli (allergen/specific IgE binding, anti-IgE, or anti-FcεRI autoantibodies) or independent activation by nonimmunological stimuli (substance P, stem cell factor, codeine, or C5a), leading to degranulation.

Allergic Urticaria

The classic example of immunological MC activation via FcεRI is IgE-mediated urticaria (often termed *allergic urticaria*). Cross-linking of receptor-bound IgE leads to the release of preformed mediators and newly synthesized lipid mediators and cytokines, resulting in the early- and late-phase IgE-mediated allergic inflammatory responses.

IgE-mediated MC activation can present as acute urticaria in those individuals previously sensitized to exogenous allergens such as foods, drugs, and latex. Allergic urticaria to inhaled allergens (e.g., latex, animal epithelia) is often accompanied by respiratory symptoms. Generalized allergic urticaria may progress to anaphylaxis. Allergic urticaria resolves rapidly on withdrawal of allergen exposure and recurs with each re-exposure to the allergen or cross-reactive agents.

Autoimmune Urticaria

Two types of autoimmunity (type I and type IIb) are proposed in CSU.[13] IgE-mediated type I autoimmunity is referred to as autoallergy.[13] Autoallergens are thought to activate MCs by cross-linking of receptor-bound IgE autoantibodies. In CSU patients, IgE autoantibodies have been demonstrated to thyroperoxidase and thyroglobulin, double-stranded DNA, interleukin-24 (IL-24) and tissue factor.[12,13] There is currently no direct evidence for this theory, which remains controversial, but type I autoimmunity may play a role in CSU based on (1) IgE sensitization to a wide range of autoantigens,[12] (2) a correlation of total IgE with CSU severity, and (3) the therapeutic efficacy of anti-IgE monoclonal antibody omalizumab in CSU.

Type IIb autoimmunity in CSU is mediated by functional autoantibodies directed against the extracellular α-chain of FcεRI on dermal MCs and basophils or, less frequently, against receptor-bound IgE, which can lead to degranulation in vitro. As determined by Western blotting or immunoenzymometric assays, the rates of IgG-anti-FcεRI in CSU ranged from 4% to 64% and that of IgG-anti-IgE, from 0% to 69%.[13] It is currently accepted that only functional autoantibodies (e.g., with histamine-releasing activity) are clinically relevant in CSU.

Although direct evidence for type II autoimmunity in animal models of CSU is lacking, the indirect evidence is compelling: (1) the detection of histamine-releasing IgG1/IgG3 anti-FcεRIα autoantibodies in CSU but not in other cutaneous diseases such as pemphigus vulgaris, dermatomyositis, systemic lupus erythematosus, bullous pemphigoid; (2) the functional relevance of anti-FcεRIα/anti-IgE autoantibodies demonstrated by in vitro activation of basophils and cutaneous MCs by IgG purified from CSU sera and by neutralization of serum-induced basophil HR by soluble recombinant α-chains of the high-affinity IgE receptors in some patients; (3) passive transfer of the autologous skin test response by serum or IgG serum fraction from CSU patients to healthy recipients;[14] (4) correlation with CSU severity; and (5) a decline of serum HR activity in parallel with CSU clinical improvement following plasmapheresis or cyclosporine.[15]

Immune Complex–Mediated Urticarial Rash

MC activation can result from binding of circulating immune complexes to FcγRIII, expressed on MCs. In addition, circulating immune complexes can activate complement, leading to C3a and C5a anaphylatoxin formation. Urticarial rash caused by immune complexes can occur acutely in serum sickness-like reactions, transfusion reactions, some drug-induced urticarias, and urticaria associated with infections or autoimmune diseases. Immune complex–mediated urticarial rashes usually develop 1 to 3 weeks after initial exposure to the antigen and disappear several weeks after antigen discontinuation.

Chronic whealing mediated by immune-complex damage and associated with histological evidence of leukocytoclastic vasculitis is known as *urticarial vasculitis* (see below). In this condition, urticaria can be associated with systemic symptoms, such as fever, arthritis, or nephritis, mainly with the uncommon hypocomplementemic variant (hypocomplementemic urticarial vasculitis syndrome [HUVS]). Damage to postcapillary venules results from the deposition of immune complexes or complement in the vessel wall. Immune complexes are formed on exposure to external (drug or infections) or internal (collagen-like region of C1q) antigens. Complement activation via the classical pathway leads to neutrophil chemotaxis through cytokine expression and adhesion molecule activation (Fig. 46.4). Proteolytic enzymes released from neutrophils damage vessel walls, leading to wheal formation and red blood cell (RBC) extravasation.

Mast-Cell Activation via MRGPRX2 Receptors

In humans, a novel G protein-coupled receptor, Mas-related GPRX2 was identified in 2015.[16] Numerous agents, including neuropeptides (substance P, neuropeptide Y, vasoactive intestinal peptide, or somatostatin), major basic protein, eosinophil peroxidase, basic secretagogues (e.g., synthetic compound 48/80), antimicrobial peptides, and many drugs (e.g., cationic peptidergic drugs), can activate cells via MRGPRX2.[16] The MRGPRX2 expression in cutaneous MCs was reported to be increased in severe CSU patients compared with healthy controls.[17]

Mast Cell and Basophil Releasability in Urticaria

In patients with CSU, dermal MCs show a decreased activation threshold to anti-IgE. Although CSU basophils are often

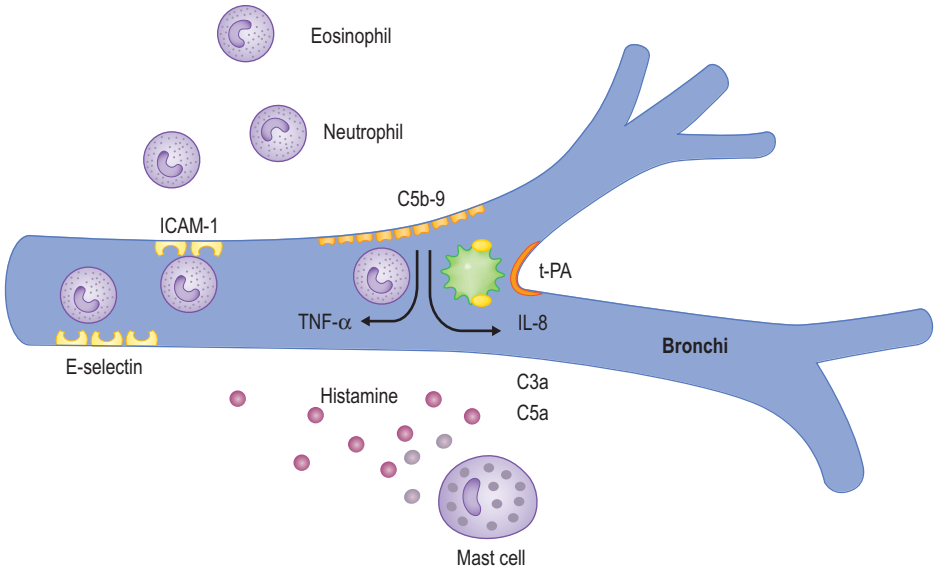


FIG. 46.4 Immune Complex–Mediated Urticaria Initiated by Lodging of Antigenic Complexes in Small Blood Vessels Followed by C3a and C5a Generation. This results in mast-cell degranulation and cytokine upregulation of adhesion molecules (intercellular adhesion molecule-1 (*ICAM-1*), E-selectin) by tumor necrosis factor-α (*TNF-α*) and interleukin-8 (*IL-8*), which leads to tissue recruitment of neutrophils and eosinophils and activation of tissue plasminogen activator (*t-PA*).

hyporesponsive to anti-IgE, they are paradoxically hyperresponsive to an as-yet unidentified factor in normal human serum.

Skin Response to Mast Cell Activation in Chronic Urticaria

Despite the existence of different pathways for MC activation, the end result is degranulation with release or secretion of mediators. Histamine is crucial for the development of cutaneous manifestations of urticaria and is found in high concentrations in tissue fluid of wheals. It causes localized redness resulting from vasodilatation, wheal formation from increased vascular permeability leading to plasma leakage, and a surrounding axon reflex-neuropeptide mediated flare-up. Histamine is also the main mediator of itch in urticaria. However, histamine-induced wheals are short-lived in contrast to urticarial lesions that may persist up to 24 hours, implying that other proinflammatory mediators and/or cellular infiltrates contribute to CSU pathogenesis, especially the delayed phase.

Mast Cell–Independent Mechanisms of Urticaria

Pseudoallergy, or nonallergic hypersensitivity, mimics immediate-type allergic reactions clinically without evidence of underlying immunological mechanisms. The most common triggers of pseudoallergic reactions are aspirin and other nonsteroidal anti-inflammatory drugs (NSAIDs), as well as some food ingredients and additives, such as salicylates, benzoates, and tartrazine. These reactions do not involve IgE sensitization and can, therefore, occur on first exposure. Pseudoallergic reactions are dose-dependent and usually occur with chemically unrelated substances. Diagnosis of nonallergic hypersensitivity is based on a distinctive clinical pattern, time course, clinical signs, and response to elimination of the cause. In the appropriate clinical context, pseudoallergy can be confirmed with oral challenge tests.

Nonsteroidal Anti-Inflammatory Drugs

Aspirin and other NSAIDs inhibit constitutive cyclooxygenase (COX-1) and inducible cyclooxygenase (COX-2), thereby diverting arachidonic acid metabolism toward the 5-lipoxygenase pathway in some cell types, notably eosinophils. This modulation of arachidonic acid metabolism results in overproduction of cysteinyl-leukotrienes LTC_4, $-D_4$, and $-E_4$, leading to vasodilatation and edema. Furthermore, reduction of PGE_2 formation by COX inhibition has two further effects that promote urticaria: first, by reducing inhibition of cysteinyl leukotriene production and second, by reducing an inhibitory effect on immunologically mediated mast-cell degranulation (Fig. 46.5). Cross-sensitivity occurs with other nonselective NSAIDs in susceptible individuals, depending on their pharmacological potency for COX inhibition but not their chemical structure.

Aspirin and NSAIDs can trigger acute urticaria as well as causing flare-ups of pre-existing chronic spontaneous (but not physical) urticaria. Aspirin-induced exacerbations have been reported in up to 30% of patients with CSU although the frequency of NSAID intolerance may fluctuate depending on the activity of the underlying CSU. NSAID intolerance may even resolve in the remission of CSU in up to one-third of patients. In some cases, aspirin can act as a cofactor in food- or exercise-induced anaphylaxis.

Urticaria can develop from a few minutes to 24 hours after aspirin ingestion but usually within 1 to 2 hours. Angioedema of the lips and tongue, impaired swallowing, and laryngeal edema develop occasionally. Aspirin-induced skin symptoms usually subside 24 to 48 hours after discontinuing the drug. However, severe exacerbations of CSU caused by NSAIDs can last from several days to several weeks.

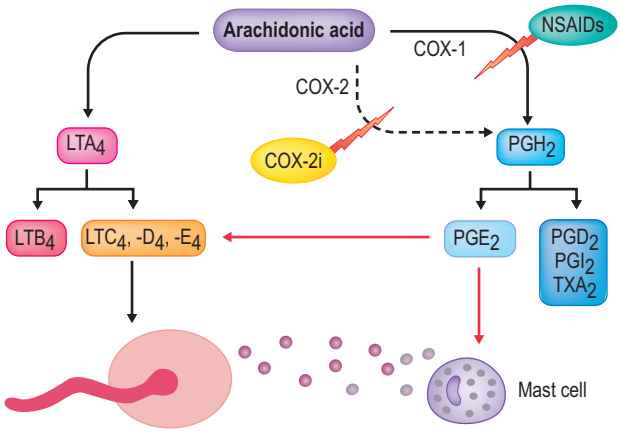


FIG. 46.5 Arachidonic Acid Pathways. Potential Diversion from Prostaglandin Synthesis to Cysteinyl leukotrienes (***LTC$_4$, -D$_4$, -E$_4$***) by Blocking Cyclooxygenase (***COX***), Leading to Increased Vasopermeability. Reduction of prostaglandin E_2 (PGE_2) also reduces its direct inhibitory effect on leukotriene production and on immunological mast-cell degranulation.

The diagnosis can be established by challenge tests. Patients should avoid aspirin and other NSAIDs, but COX-2 inhibitors (coxibs) are usually well tolerated and can probably be used safely. Long-term low-dose aspirin for thromboprophylaxis is often well tolerated and should not be discontinued routinely.

Food-Induced Pseudoallergic Reactions in Chronic Spontaneous Urticaria

Pseudoallergic food reactions appear to be important in some patients with CSU. Pseudoallergic food triggers include natural salicylates in fruit and vegetables and artificial food additives in processed foods, such as benzoates and tartrazine. Low-molecular-weight aromatic compounds in tomatoes, white wine, and herbs have also been implicated. Clinically, exacerbations of CSU resulting from dietary pseudoallergens gradually subside within 10 to 14 days on an exclusion diet in contrast to 1 to 3 days in acute allergic urticaria. Responders to a low pseudoallergen diet demonstrated normalization of gut mucosal permeability and skin symptoms. Although the underlying mechanisms for pseudoallergic reactions in CU remain unproven, an impaired gastroduodenal barrier function may be a contributory factor.

Bradykinin-Mediated Angioedema

The principal mediator of angioedema resulting from hereditary and acquired C1 esterase inhibitor (C1-INH) deficiency is bradykinin. Hereditary angioedema (HAE) attacks are characterized by overproduction of bradykinin. Angiotensin-converting enzyme inhibitor (ACEI)-induced angioedema results from reduced catabolism with accumulation of bradykinin. Idiopathic angioedema not responding to antihistamines may be mediated by bradykinin in some patients. Increased levels of bradykinin lead to activation of bradykinin β2 receptors, thereby increasing vascular permeability and causing vasodilatation and edema.[18]

CLINICAL CLASSIFICATION

CU may be spontaneous, inducible, or both.

Spontaneous Urticaria

Spontaneous urticaria is the most common presentation. The term "spontaneous" makes no assumption about etiology, which may include autoimmunity, allergy, pseudoallergy, or infection. Spontaneous urticaria can be further classified by duration of attacks into acute and CU.[2]

Acute Spontaneous Urticaria

Acute urticaria is defined as continuous disease lasting less than 6 weeks. Individuals with an atopic predisposition are at higher risk of acute urticaria; atopic diseases are found in about half the patients with acute urticaria. Most acute urticaria resolves spontaneously within 6 weeks, but in 10% of patients, it may progress to CSU.

Viral infections, foods, or drugs are common identifiable causes of acute urticaria. For example, urticaria was reported during or after the COVID-19 infection.[19] A specific cause of acute urticaria will not be found in about 50% of patients. Foods may cause acute urticaria in young children but are rarely responsible for acute urticaria in adults. In infancy, cow's milk allergy may cause acute allergic urticaria. Acute urticaria is the most common presentation of drug hypersensitivity, accounting for a quarter of all adverse drug reactions, with penicillin and NSAIDs being the most common causes of allergic and nonallergic drug-induced urticaria, respectively. Drug-induced urticaria is more likely in older adults, perhaps reflecting polypharmacy and age-related pharmacokinetic changes, and in patients with human immunodeficiency virus (HIV) infection and those with renal or liver diseases.

Chronic Spontaneous Urticaria

CSU is characterized by daily or almost daily itchy wheals on skin, with or without angioedema, for more than 6 weeks. CSU can be mediated by type I and type IIb autoimmunity (see Autoimmune Urticaria). However, in 50% of patients, the etiopathogenesis of the disease remains unknown (idiopathic) and seems to be endogenous. Whatever the primary cause, CSU appears to be aggravated by a variety of exogenous exposures, including acute infections (most commonly viral upper respiratory tract infections), NSAIDs, dietary pseudoallergens, menses in women, tiredness, and stress. The contribution of these factors probably explains why the clinical course is often erratic and unpredictable.

CSU affects women twice as often as men. There is a common link between CSU and thyroid autoimmunity as well as other autoimmune diseases, including Graves disease, vitiligo, systemic lupus erythematosus, pernicious anemia, and insulin-dependent diabetes.[20] Adult female patients with a family history and a genetic predisposition for autoimmune diseases are at an increased risk of autoimmune comorbidity.[20] CSU may be associated with inducible urticaria (e.g., dermographism or delayed-pressure urticaria [DPU]). Around 50% of patients present with both wheals and angioedema. CU can have a continuous or a relapsing course. CU of all types—including CSU—can cause serious disability in patients, including loss of sleep and energy, social isolation, altered emotional reactions, and difficulties in aspects of daily living similar in degree to patients with severe ischemic heart disease.

Chronic Inducible Urticarias

Chronic inducible urticarias (CIndUs) are common, accounting for around 25% of all cases of CU. These include physical urticarias, such as symptomatic dermographism, cold and heat urticarias, DPU, solar urticaria, and vibratory angioedema, in which symptoms are triggered reproducibly by an external mechanical or thermal/ultraviolet stimulus. CIndUs now also include cholinergic, adrenergic, aquagenic, and contact urticarias, in which the eliciting stimuli for MC degranulation are defined by a nonphysical exposure. CIndUs can coexist with CSU and more than one CIndU can occur in the same patient (e.g., dermographic and cholinergic urticarias), which may lead to difficulty in diagnosis. With the exception of DPU, CIndU develop rapidly after exposure to the relevant trigger and fade within an hour.[2]

The pathogenesis of CIndUs is unclear except in the case of allergic contact urticaria, which is caused by mucocutaneous exposure to an allergen in IgE-sensitized individuals, analogous to acute urticaria. A diagnosis of CIndUs is confirmed if the symptoms can be reproduced by challenge testing with the suspected stimulus.[21] Challenge testing can be used for monitoring threshold changes during treatment. In general, treatment of CIndUs involves avoidance of known triggers and taking antihistamines. Off-label treatment with omalizumab can be beneficial in patients with various CIndUs, with the strongest evidence for symptomatic dermographism, cold and solar urticarias.[22] Sometimes tolerance can be induced for cold and solar urticarias but these historical approaches to treatment are now rarely used.

Mechanical Urticaria

Symptomatic Dermographism

Dermographism is the most common CIndU, mainly affecting young people. Typical red, itchy, linear wheals are evoked within minutes of stroking, friction, rubbing, or scratching the skin (Fig. 46.6).

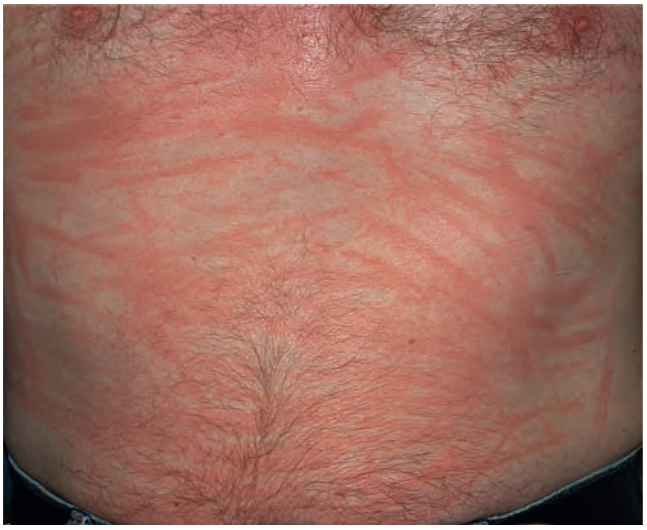


FIG. 46.6 Extensive Induced Dermographic Whealing on the Chest of a Patient as a Result of Scratching.

Delayed-Pressure Urticaria

Isolated DPU occurs in less than 2% of all patients with urticaria. DPU has been reported to coexist with CSU in up to 40% of patients but needs to be distinguished from pressure-aggravated CSU by provocation testing (below). DPU is the most debilitating of CIndUs and is triggered by sustained local pressure (e.g., wearing tight shoes, carrying heavy bags, long walks, sitting or leaning against firm objects, climbing ladders, jogging, driving, or clapping hands). Deep and painful swellings, clinically resembling angioedema, develop 30 minutes to 12 hours after pressure, and may be associated with 'flu-like symptoms, fever, arthralgia, and fatigue. The most frequently affected sites are hands, soles, buttocks, shoulders, and areas under straps and belts. DPU lesions last 12 to 48 hours and are usually painful rather than itchy, especially on the hands and feet. Laboratory investigation reveals transitory leukocytosis and elevated erythrocyte sedimentation rate (ESR). Hanging a heavy weight suspended on a narrow band over the forearm or thigh for 15 minutes may be used as a challenge test, but more reliable results can be obtained with special instruments. DPU is difficult to treat because it responds poorly to antihistamines.

Vibratory Angioedema

Vibratory angioedema is rare and has been linked to mutations in the *ADGRE2* gene.[23] Local swellings, or rarely wheals, develop several minutes to 6 hours after using vibrating machinery, lawn mowing, applauding, and jogging, for instance. Systemic symptoms may occur (headache, chest tightness, diffuse flare, facial flushing). Placing the elbow or hand on a laboratory vortex for 5 to 15 minutes is a useful challenge test.[21] Avoidance of the trigger is the only helpful treatment strategy. In rare cases, antihistamines can relieve the symptoms of vibratory angioedema.

Thermal or Ultraviolet-Induced Urticaria

Cold Urticaria

The incidence of cold urticaria (ColdU) is estimated at 0.05%. It occurs in both children and adults and is more common in cold climates, in women, and in atopic patients. The majority of cases are primary with no identifiable cause, but some cases are secondary to internal disease. Clinical manifestations can be local or generalized. Mucosal involvement may develop after drinking cold beverages. Systemic symptoms may be respiratory (laryngeal angioedema, tongue or pharyngeal swelling, wheezing), vascular (hypotension, tachycardia), GI (hyperacidity, nausea, diarrhea), or neurological (disorientation, headache). ColdU can be evoked by low ambient temperature, contact with cold objects, food or beverages, or immersion in cold water. Wheals develop during the cold exposure or, more commonly, on rewarming. The severity of ColdU depends on the intensity and duration of the cold stimulus. ColdU is potentially life threatening, with a risk of anaphylaxis and death on exposure of large skin areas to cold: for example, jumping into cold water and hypothermia in neurosurgical and cardiothoracic operations.[24]

In most studies, cryoproteins (mainly cryoglobulins) can be detected in less than 1% of patients (secondary ColdU). ColdU can be associated with infections (hepatitis C, infectious mononucleosis, syphilis, *Mycoplasma* infection), autoimmune diseases, and lymphoreticular malignancy (Waldenström macroglobulinemia, myeloma), although the evidence for causal relationships is weak. ColdU can precede these diseases by several years. Secondary ColdU can also be drug-related (penicillin, oral contraceptives).

The diagnosis of ColdU is confirmed by a 5-minute ice cube challenge or TempTest.[21] In patients with negative provocation tests for ColdU, atypical forms or familial cold autoinflammatory syndrome (FCAS) should be considered.

The clinical work-up in ColdU includes differential blood count, ESR, or C-reactive protein (CRP).[2,21] Additional work-up includes a search for underlying infections, if indicated by patient's history or required for the differential diagnosis. Cryoproteins can be measured, although there is no guidance on clinical relevance in ColdU. Patients should be cautious about bathing or swimming in cold water and consuming cold food or drinks. Antihistamine treatment is often helpful but does not prevent anaphylaxis caused by swimming in cold water. Omalizumab can be effective for patients with ColdU. In severe ColdU, tolerance induction by repeated cold exposure may be attempted but is very difficult to maintain.

Heat Urticaria

Heat-induced urticaria is very rare. It is induced by local heating of the skin at 38°C to 44°C. Antihistamines are only of limited therapeutic value, but omalizumab may be effective.

Solar Urticaria

Solar urticaria affects about 1% of all patients with urticaria and has a slight female predominance. It can be associated with erythropoietic porphyria. Wheals are caused by electromagnetic wavelengths ranging from 290 to 760 nm (ultraviolet B [UVB], UVA, and visible spectrum). In patients with solar urticaria, the action spectrum lies predominantly within longer UVA wavelengths and shorter visible light wavelength.[25] It develops within minutes or hours after sun exposure and fades within 24 hours. Lesions are usually confined to sun-exposed skin, although they can also develop under clothing. The severity of solar urticaria depends on the wavelength, intensity, and duration of irradiation. Short exposures induce flare and pruritus, whereas longer exposures cause whealing. In patients sensitive to the visible spectrum and UVA, reactions may occur through window glass.

Solar urticaria is diagnosed by phototesting with broadband UVR and monochromator sources. Patients are advised to use creams with a high sun protection factor (SPF), protective clothing, and protective window shields and to limit the time spent outdoors. Omalizumab may be effective. Phototherapy (PUVA and narrow band UV 311 nm) can also be used.

Other Patterns of Inducible Urticaria

Cholinergic Urticaria

Cholinergic urticaria (CholU) is the second most common CIndU and occurs mainly in adolescents, young adults who are usually atopic. CholU usually follows a rise in core temperature resulting from physical exercise, fever, or external passive heat (hot bath, shower, sauna), but may also be provoked by emotional stress and spicy food. The characteristic lesions are highly pruritic pinpoint pale wheals of 1 to 3 mm surrounded by a red flare. The wheals may occur anywhere except the soles and palms. Lesions usually begin on the trunk and the neck, extending outward to the face and limbs. As lesions progress, confluent areas of whealing and redness may develop. Severely affected patients may develop angioedema and even anaphylaxis. The

rash is triggered by activation of cholinergic sympathetic innervation of sweat glands, but the mechanism of activation remains unclear. Some patients with CholU have acquired generalized hypohidrosis/anhidrosis.[26] Decreased blood protease inhibitor levels have been reported, and this is the rationale for using anabolic steroids to treat severe, unresponsive disease. The prognosis is reasonably favorable, with spontaneous resolution within 8 years in most patients. However, 30% of patients are affected for over 10 years.

CholU can be confirmed by reproducing the rash with exercise to the point of sweating or passive heating in a hot bath or shower at up to 42°C. The condition may be refractory for up to 24 hours after provocation, and this may enable patients to prevent attacks by taking daily exercise. Treatment is primarily with antihistamines, but beta-blockers, danazol, ketotifen, and montelukast have also been used. Some patients with CholU may respond to omalizumab treatment although the evidence for its clinical efficacy is limited.[22]

Aquagenic Urticaria

Aquagenic urticaria is rare. It occurs in women more often than in men and is triggered by water contact but not after drinking water. Scattered small papular wheals, similar to CholU but with a larger flare, appear within 10 to 20 minutes of water contact and resolve in 30 to 60 minutes. Diagnosis is made by water immersion or applying a wet compress at body temperature for up to 10 minutes on whichever part of the body is usually affected. Associations with HIV and hepatitis B infection have been described.

Contact Urticaria

Contact urticaria occurs locally after skin or mucosal contact with the eliciting agent.[2] Reactions usually develop within a few minutes and resolve over 2 hours, although delayed-contact urticaria can occur with a latent period up to 48 hours. Contact urticaria can be caused by many organic and inorganic stimuli, such as latex, animal danders and secretions, foods, plants, topical drugs, and cosmetics.

Allergic contact urticaria occurs mainly in atopic subjects. Foods are the most common cause. Very rarely, contact with the allergen may induce anaphylaxis, especially in the context of latex allergy. The severity of contact reactions depends on many factors (area of exposure, duration of contact, amount and concentration of substance, patient reactivity, comorbidity, and concomitant treatment).

Nonallergic contact urticaria is common and is usually caused by low-molecular-weight chemicals. Nonallergic contact urticaria may occur on first exposure and may depend on dose and concentration of the chemical.

DIFFERENTIAL DIAGNOSIS OF URTICARIA

Urticarial Vasculitis

Urticarial vasculitis (UV) is characterized clinically by urticarial rash with or without angioedema and histologically by leukocytoclastic vasculitis on lesional skin biopsy. UV is idiopathic in most cases but has been reported in association with drugs, malignancy, autoimmunity, and infections. It is important to differentiate UV from CSU in terms of prognosis, approach to diagnostic evaluation, and therapy.

Angioedema Without Wheals

Angioedema occurs in nearly half the patients with CSU; in these patients, the disease tends to be more severe and more difficult to treat. Angioedema without wheals occurs in about 10% of patients with CSU but needs investigation to distinguish angioedema due to C1 inhibitor deficiency or drugs, such as ACEI, although many cases of angioedema without wheals remain unexplained (Fig. 46.7).

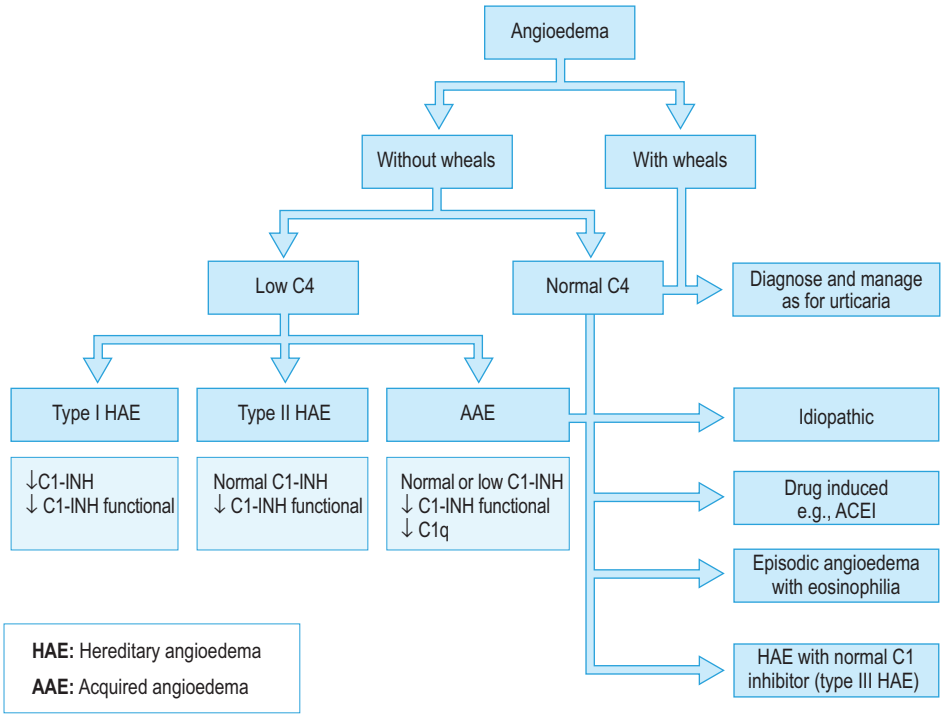


FIG. 46.7 Diagnostic Pathway for the Differential Diagnosis of Angioedema.

Angioedema Caused by C1 Inhibitor Deficiency

The diagnosis and clinical presentation of hereditary and acquired C1-INH deficiency are covered in Chapter 40.

Angioedema With Normal C1 Inhibitor

Histamine-Mediated Angioedema (without wheals). Histamine-mediated angioedema differs from bradykinin-mediated angioedema by the speed of onset, kinetics of resolution, prognosis, and response to treatment. Histamine-mediated angioedema develops rapidly, usually subsides within 48 hours, and responds well to antihistamines, corticosteroids, and epinephrine.

Idiopathic. Idiopathic angioedema is diagnosed when no cause can be established and can be histaminergic and non-histaminergic. It is generally self-limiting, but may follow a prolonged course. Treatment of idiopathic histaminergic angioedema should follow the same pathway as CSU. Antihistamines are the mainstay of therapy; corticosteroids (prednisolone 30 to 40 mg/day in adults) are useful for up to 3 days as rescue treatment to cover more severe episodes of the oropharynx, and epinephrine may be lifesaving in laryngeal angioedema.

Drug-induced. Histaminergic angioedema without wheals occasionally occurs with NSAIDs, usually within several hours of intake.

Bradykinin-Mediated Angioedema. Bradykininergic angioedema is usually characterized by a slower onset, a longer duration, and a little or no response to antihistamines, corticosteroids, and epinephrine. Bradykininergic angioedema can be idiopathic, due to hereditary or acquired C1 deficiency or ACEI therapy. Although the mortality due to angioedema is rare (0.34 deaths per 1,000,000 inhabitants), bradykininergic angioedema carries a 45 times higher mortality risk of mortality than angioedema due to MC-derived mediators.[27]

Hereditary. The prevalence of bradykininergic angioedema due to hereditary deficiency in C1 inhibitor is between 1.1 and 1.6 per 100,000.[28] HAE due to C1 inhibitor deficiency is characterized either by low antigenic and functional levels of C1 inhibitor (type1) or by normal (or elevated) antigenic but low functional C1 inhibitor levels (type 2). More than 500 mutations in the *SERPING1* gene have been reported. Some kindreds have been described with a dominant pattern of angioedema inheritance and normal C1 inhibitor, mainly in women, in whom a gain-of-function (GOF) mutation of the contact pathway of coagulation in *F12* has been found in about 20% of affected families. This used to be called type III HAE, but the term *hereditary angioedema with normal C1 inhibitor* is now preferred. HAE with normal C1 inhibitor can also be associated with mutations in plasminogen, kininogen 1, angiopoietin-1 or heparan sulphate 3-O-sulfotransferase 6 genes.

Drug-induced. The prevalence of bradykininergic angioedema due to ACEIs is between seven and 26 in 100,000.[28] It is thought to result from inhibition of kininase II, which breaks down bradykinin, as well as converting angiotensin I to angiotensin II in the renin–aldosterone pathway. It usually presents with episodic and unpredictable swellings of the head and neck, especially of the tongue and oropharynx. Although angioedema develops within the first week of treatment with ACEIs in most cases, symptom onset may occur after many years on treatment. Management involves discontinuation of ACEI therapy. Angioedema often recurs on re-exposure to ACEIs. ACEIs may also precipitate angioedema in patients with angioedema from other causes, including C1-INH deficiency. Rare instances of angioedema have also been reported with angiotensin II receptor antagonists. Additionally, dipeptidyl peptidase IV inhibitors (gliptins) used for the treatment of diabetes can affect bradykinin degradation and can cause angioedema. Current guidelines recommend close monitoring, airway intervention if needed, and lifetime avoidance of all ACEIs.

Acquired. Acquired deficiency in C1 inhibitor rarely occurs in association with lymphoproliferative or autoimmune disorders and is observed in patients over 40 years old and affects 0.15 per 100,000.[29] Acquired angioedema is often characterized by C1 inhibitor autoantibody or paraprotein and low C1q.

Idiopathic. Idiopathic non-histaminergic angioedema is bradykinin-mediated in approximately 10% of patients. A role for bradykinin has been suggested although the mechanisms remain unknown. Off-label therapeutic options in acute attacks of idiopathic angioedema include omalizumab, icatibant, C1 inhibitor, and ecallantide. Tranexamic acid is effective in more than 50% of patients with idiopathic angioedema without wheals. For prophylaxis, tranexamic acid, omalizumab, C1 inhibitor, and progestin were reported to be effective.

Autoinflammatory Syndromes Presenting With Urticarial Rash

Acquired

Schnitzler Syndrome. Schnitzler syndrome is a periodic fever syndrome with urticarial rash resembling CSU, bone pain, high ESR, and monoclonal IgM, or—very rarely—IgG gammopathy. Clinically, patients present with nonpruritic or mildly pruritic wheals, mainly affecting the trunk and limbs. The wheals are resistant to antihistamines, and angioedema is rare. Fever bouts may exceed 38°C, sometimes with chills and nocturnal sweating. Patients often suffer from bone pain, mainly in the pelvis or tibias, arthralgia, and sometimes full-blown arthritis. Lymphadenopathy, hepatomegaly, and splenomegaly may be present.

Monoclonal IgMκ and, less commonly, IgMλ or IgG paraproteins are found on serum electrophoresis. The ESR is persistently elevated at 60 to 100 mm/h, CRP (greater than 30 mg/L), with leukocytosis (greater than 10,000/mm^3), elevated platelet count, and anemia. Skin histology shows dermal neutrophils with a tendency to localize around appendages in most cases; monoclonal IgM is deposited in the epidermis around the keratinocytes and along basement membranes on direct immunofluorescence. Bone examination may demonstrate hyperostosis on radiography and hyperfixation on bone technetium scanning. Bone marrow examination shows normal results in most patients, but nonspecific lymphocytic, plasmocytic, or polyclonal infiltration is present in about 20%.

The pathophysiology of Schnitzler syndrome is still unclear, and the severity of urticarial rash does not depend on the paraprotein level. Evidence of activation of interleukin-1 (IL-1), increased IL-6, IL-18, granulocyte macrophage–colony-stimulating factor (GM-CSF), and confirming the efficacy of the IL-1 receptor antagonist anakinra, anti-IL1-β monoclonal antibody canakinumab, or fusion protein IL-1 antagonist rilonacept suggest that cytokines play a leading role in its pathogenesis. The outlook for maintaining symptom control is good, but long-term follow-up is recommended because patients may develop B-cell lymphomas 10 to 20 years after its onset. AA amyloidosis may also occur without effective treatment.

Hereditary (Cryopyrin-Associated) Periodic Syndromes

Several hereditary autoinflammatory urticarial syndromes show mutations of the *NLRP-3* gene on chromosome 1q44. *NLRP-3* encodes a protein called cryopyrin, which is involved in apoptosis and inflammation. These rare autosomal dominant disorders include FCAS, Muckle-Wells syndrome (MWS), and chronic infantile neurological, cutaneous, and articular (CINCA) syndrome, and newly described (FACAS)[9] and are grouped under the inclusive term *cryopyrin-associated autoinflammatory syndrome* (CAPS).[8] The diagnostic criteria for CAPS include raised inflammatory markers (CRP/serum amyloid A) in combination with at least two of six CAPS-typical symptoms: urticaria-like lesions, cold-triggered episodes, sensorineural hearing loss, musculoskeletal symptoms, chronic aseptic meningitis, and skeletal abnormalities.[8]

WORK-UP IN PATIENTS WITH URTICARIA

Evaluation of patients with urticaria requires a detailed history and physical examination.[2] The history is particularly important in patients with urticaria and should include a thorough inquiry for all potential comorbidities (e.g., autoimmune diseases), possible precipitating and aggravating factors, the timing of onset and duration of individual wheals, associated symptoms, as well as travel history, recent infection, occupational exposure, and food and drug intake. The duration of individual lesions can be very helpful in distinguishing the different clinical patterns of urticaria. Patients should be asked to complete Patient Related Outcome Measures and weekly urticarial activity scores for CSU.[2]

CLINICAL PEARLS

Diagnosis of Clinical Patterns of Urticaria

- The duration of individual wheals can help define the pattern of urticaria.
- Wheals lasting no more than 1 h are usually induced rather than spontaneous except delayed-pressure urticaria, which lasts longer.
- Localized wheals lasting up to 2 h may be caused by skin or mucosal contact with an allergen or other contactant.
- Wheals that take 1 to 24 h to fade are usually a presentation of chronic spontaneous urticaria.
- Wheals lasting more than 24 h may be caused by delayed-pressure urticaria or urticarial vasculitis.
- Urticarial rash in a neonate may be a symptom of cryopyrin-associated periodic syndromes.

Work-Up of Acute Urticaria

No routine testing is recommended for the patients with acute urticaria.[2] When allergens are suspected as the cause, there should be a close temporal relationship to the time of exposure—usually starting within minutes—a history of previous exposure causing sensitization, and prompt resolution on allergy withdrawal. Allergy testing can be performed if indicated.

Work-Up in Chronic Inducible Urticarias

When CIndU is suspected, appropriate challenge testing should be performed to confirm the diagnosis.[21] Generally, there is no need for further investigation, except for cold or solar urticaria.[21] In ColdU, a differential blood count, ESR or CRP, and cryoproteins are recommended and other photodermatoses should be considered in the diagnosis of solar urticaria.

Work-Up in Chronic Spontaneous Urticarias

A complete blood count with differential and ESR or CRP is routinely recommended for patients with CSU that is easily controlled by antihistamines, unless the history points to an underlying disease.[2] Studies have shown that random laboratory testing very rarely yields evidence of unsuspected internal diseases as a cause of CSU.

Extended laboratory evaluation can be considered in patients with CSU if indicated by the patient's history and/or physical examination.[2] Thyroid-stimulating hormone (TSH), thyroid antibodies, liver function tests, tryptase, and routine urinalysis will exclude most diseases associated with urticaria. Additional testing for infections, including *Helicobacter pylori*, can be performed if suggested by patient's history. Lesional skin biopsy can be considered if UV is suspected. Allergy testing should not be undertaken in chronic continuous urticaria unless the history indicates that allergy is likely.

The Diagnosis of Autoimmune Chronic Spontaneous Urticaria

The diagnosis of autoimmune CSU (type IIb autoimmunity) is not straightforward and involves in vivo and in vitro approaches. Autologous serum skin testing (ASST) is a simple and useful screening method for autoreactivity in patients with CU. The test is 80% specific and 70% sensitive as compared to a positive basophil histamine release assay (BHRA) and is, therefore, regarded as a useful test for autoimmune urticaria.

The diagnostic criteria for autoimmune CSU may include (1) positive in vivo autoreactivity (a positive ASST) as evidence of serum factors capable of inducing an inflammatory wheal response; (2) positive in vitro basophil reactivity as evidenced by BHRA or basophil activation test (BAT) using the patient's serum on healthy donor basophils or MCs; and (3) a positive immunoassay for specific identification of IgG autoantibodies against FcεRI and/or anti-IgE (Western blot or enzyme-linked immunosorbent assay [ELISA]. Being technically difficult, these assays are mainly confined to research centers. Low total IgE with positive thyroid antibodies may offer a useful surrogate for the BHRA. Results of nonfunctional immunoassays (Western blot, ELISA) based on binding of autoantibodies to relevant antigens (FcεRIα or IgE) do not correlate well with the results of functional assays.

MANAGEMENT OF URTICARIA

Finding effective treatment for urticaria can be challenging. Treatment should be tailored to the clinical pattern, duration, and severity of the urticaria. Management should include nonpharmacological measures and drug therapy with a stepwise approach.[2]

General Measures

Causes, triggers, and aggravating factors should be avoided or minimized, whenever possible. Patients with CSU should minimize exposure to nonspecific aggravating factors, such as overheating, wearing tight clothes and shoes, stress, alcohol, dietary pseudoallergens, and some drugs. NSAIDs aggravate up to 30% of patients with CSU and are best avoided. This does not apply to the physical urticarias, in particular DPU, where NSAIDs may be used as treatment. ACEIs are contraindicated in angioedema without wheals that may be mediated by kinins, but not

THERAPEUTIC PRINCIPLES

Management of Urticaria

- Eliminate infectious, drug, or food causes.
- Minimize nonspecific aggravators, including heat, stress, alcohol, nonsteroidal anti-inflammatory drugs, and pressure.
- Regular oral second-generation H_1 antihistamines are the first line of therapy for all spontaneous and inducible urticarias.
- Updosing of second-generation antihistamines is the second line of treatment for all spontaneous and inducible urticarias.
- Short courses of oral corticosteroids may be necessary as rescue medication for chronic spontaneous urticaria flare-ups, especially with oropharyngeal angioedema.
- Omalizumab (monoclonal anti-IgE) is often effective in chronic spontaneous urticaria refractory to H_1 antihistamines
- Immunosuppressive therapies should be reserved for patients with severe urticaria not responding to omalizumab. Careful monitoring for renal toxicity or hypertension is required with cyclosporine.

in other patterns of urticaria. Cooling lotions and creams, such as 1% menthol in aqueous cream, may help relieve pruritus. Some patients with spontaneous, but not inducible CU appear to respond to a low pseudoallergen diet. However, controlled clinical trials are lacking.

First-Line Therapy

Antihistamines are the cornerstone of treatment for acute and CSU. Second-generation antihistamines offer several advantages over classic H_1 antihistamines, including lack of sedation and impairment of performance, longer duration of action, and absence of anticholinergic side effects. Meta-analysis suggests that 63.2% of CSU patients respond to updosing of second-generation antihistamines although this approach is more effective against itching than whealing.[30] Second-generation antihistamines are inverse agonists of H_1 receptors, and stabilize H_1 receptors in the inactive conformation. Second-generation antihistamines are most effective in CSU when taken regularly for prophylaxis. The timing of antihistamine intake should be adjusted to suit the diurnal pattern of urticaria for each individual.

Second-Line Therapy

It has become common practice to increase second-generation antihistamines above their licensed doses up to fourfold when CU does not respond because clinical experience and updosing studies show that this achieves better control in some patients.

When urticaria does not respond to second-generation antihistamines, systemic corticosteroids may be used as short-term rescue therapy for acute urticaria or severe exacerbations of CSU. Long-term treatment with oral corticosteroids is not recommended because of safety concerns.

Third-Line Therapy

Omalizumab (anti-IgE) is licensed as add-on therapy for CSU not responding to antihistamines. The recommended dose is 300 mg by monthly subcutaneous injection, although clinical experience indicates that effective monthly doses may vary from 150 to 600 mg in different patients and the interval between treatments can be extended in good responders. Clinical experience and meta-analysis suggest that when used off-label, it may also be effective for many patients with CIndUs, with the strongest evidence for symptomatic dermographism, cold and solar urticarias.[22] Omalizumab was reported to rapidly reduce circulating free IgE, to downregulate high-affinity IgE receptors on MCs and basophils, resulting in a stabilizing effect. The mechanisms of omalizumab action in CSU remain certain but are likely to include reducing MC releasability, reversing basopenia, reducing activity of IgG autoantibodies against FcεRI and IgE, reducing activity of IgE autoantibodies against an as-yet undefined antigen or autoantigens, reducing the activity of intrinsically "abnormal" IgE and decreasing in vitro coagulation abnormalities.[31] Although experience to date suggests that omalizumab is often highly effective for control of urticaria symptoms, patients with serum autoreactivity as demonstrated by serum-induced BHRA and/ASST have a slower response.

Fourth-Line Therapy

The best-studied immunosuppressive therapy is cyclosporine, shown to be effective at low to moderate doses ranging from 1 to 4 mg/kg daily in patients with CSU. Patients must be monitored carefully for renal impairment and hypertension; treatment should normally be limited to 4 months. Cyclosporine is contraindicated in patients with previous malignant disease except nonmelanoma skin cancer.

MANAGEMENT OF HEREDITARY ANGIOEDEMA

The swellings of HAE are mediated by bradykinin rather than histamine, so the management of HAE is completely different from MC-dependent angioedema. The primary aim of therapy is to replace the missing functional C1 esterase inhibitor or stabilize the coagulation, fibrinolysis, complement, and kallikrein–kinin pathways. ACEIs should be avoided because ACE is a key enzyme involved in the breakdown of bradykinin. Exogenous estrogens (oral contraceptives and hormone replacement therapies) should be avoided in women as they activate kallikrein through activated factor XII. Lifestyle events that exacerbate HAE vary in their importance among individuals, but may include local trauma (e.g., dental extraction, sport activities), stress, tiredness, and intercurrent infections.

Attacks involving the extremities or abdomen are the most common. Up to 50% of patients will experience oropharyngeal swelling, with risk of asphyxiation at some point in their lives, so treatment for both emergencies and prophylaxis are essential. The intention of treatment is to reduce or curtail the severity of a swelling and to reduce the period of disability or disfigurement during or following the attacks.

Treatment of the Acute Attack

International guidelines strongly recommend early administration of C1 inhibitor concentrate, icatibant, or ecallantide for the treatment of the acute attacks.[32] Plasma-derived (pd-C1-INH) or recombinant C1 esterase inhibitor concentrate given by intravenous infusion is effective for oropharyngeal or GI attacks. Initial improvement in the swelling may be seen within 30 to 60 minutes, and time to clearance is usually in the order of 24 hours after pd-C1-INH. Self-administration can reduce the severity of attacks by allowing earliest treatment and should be encouraged.

If a C1 inhibitor concentrate, ecallantide, or icatibant is not available, solvent detergent-treated plasma should be used. If not available, up to three units of fresh frozen plasma (containing C1 inhibitor and its substrate, complement) may be given as an alternative in an emergency when C1-INH concentrate is not available.

Icatibant (a bradykinin 2 receptor antagonist) and ecallantide (a kallikrein inhibitor) are also highly effective for acute

episodes of HAE. Both are given by subcutaneous injection, which provides an easier route for administration than intravenous infusion. Ecallantide should be given under medical supervision, since there is a risk of allergic reactions. Icatibant is licensed for self-administration, which provides a potential advantage. Treatments developed for HAE with C1 inhibitor deficiency such as C1 inhibitor concentrate and icatibant are also effective in HAE with normal C1 inhibitor.

Laryngeal angioedema is a medical emergency and intubation or surgical airway intervention should be considered in the context of progressive respiratory compromise due to upper airway angioedema.

Antifibrinolytics (e.g., tranexamic acid) or androgens should not be used for acute treatment of HAE attacks.

Short-Term Prophylaxis

It has become common practice to give prophylactic treatment with pd-C1-INH before procedures involving local trauma to the oropharynx, including dental extraction, intubation for general anesthesia for invasive interventions. Common practice is to give 1000 units or a dose of 20 units/kg of pd-C1-INH as close as possible to the start of the procedure. Fresh frozen plasma can be used as a second-line agent for short-term prophylaxis. Another strategy is the prophylactic treatment with anabolic steroids for 5 days before and 2 to 3 days after the procedure. However, frequent short courses may lead to side effects. Specific guidance on dosing for adults and children can be found elsewhere. Tranexamic acid is no longer recommended for short-term prophylaxis.[32]

Long-Term Prophylaxis

Long-term prophylaxis is required in patients with frequent attacks. pd-C1-INH is the preferred first-line long-term prophylaxis for HAE attacks, usually given subcutaneously at doses of 40 to 60 units/kg bodyweight twice a week.[6]

Attenuated androgens are commonly used as second-line long-term prophylaxis as they increase intrinsic C1 inhibitor production by the liver in heterozygotes with a functioning allele and promote bradykinin degradation through an increase in aminopeptidase P. The dose should be titrated against the clinical response but not against blood levels of C1-INH to the lowest level that prevents or ameliorates the condition. Virilizing side effects may be a problem for women, and anabolic steroids are avoided in children due to growth retardation. Monitoring of liver function and lipid profiles should be undertaken periodically. Performing a liver ultrasound examination once every 2 years to screen for development of hepatoma is recommended in patients on long-term prophylaxis. Currently, tranexamic acid is not recommended for long-term prophylaxis given the lack of efficacy data and the availability of effective alternatives such as C1 inhibitors. However, tranexamic acid is preferred in children when C1 inhibitor concentrate is not available. Therapeutic advances include subcutaneous lanadelumab, a recombinant fully human monoclonal antibody against plasma kallikrein, which is currently licensed for HAE prophylaxis and berotralstat, a new oral kallikrein inhibitor.

Gene Therapy

Gene therapy using gene editing techniques or viral vector gene transfer is considered a potential treatment option for HAE patients in the future. The first report by Haslund and colleagues demonstrated an enhanced C1 inhibitor secretion in patient-derived fibroblasts transfected with wild-type *SERPING1* gene variant, pointing towards the potential of gene therapy in HAE.[33]

Anaphylaxis

Anaphylaxis is an acute, potentially life-threatening, systemic allergic reaction with variable clinical presentations. Its clinical features and management have been summarized in international guidelines.[34] A diagnosis of anaphylaxis is fulfilled if any one of the following are fulfilled: (1) acute onset (minutes to hours) with involvement of skin, mucosal tissue, or both with either respiratory involvement or reduced blood pressure (BP) and/or associated symptoms of end-organ dysfunction; (2) two or more of the following (skin-mucosal symptoms, respiratory involvement, reduced BP, GI symptoms) occurring rapidly on exposure to a culprit allergen; (3) reduced BP in response to exposure to a known allergen.[34]

Epidemiology of Anaphylaxis

The incidence of anaphylaxis in the general population in Europe varies from 1.5 to 7.9 cases per 100,000 per annum.[35] Lifetime prevalence of anaphylaxis is estimated at 0.3% in Europe[35,36] and varies between 1.6% and 5.1% in the United States. The mortality rate in anaphylaxis is estimated to be between 0.47 and 0.69 per million persons (1% of hospitalizations and 0.1% of emergency department visits).[37] There is a trend for increasing rates of anaphylaxis admissions in the UK, United States, Canada, and Australia.[37] The incidence of anaphylaxis in children ranges from 10.5 to 70 episodes per 100,000 persons-years.

Pathophysiology of Anaphylaxis

Although anaphylaxis can be mediated by immunological and nonimmunological mechanisms, the clinical presentation is similar in both, and most authorities no longer make a distinction. Immunological anaphylaxis is further classified as IgE-dependent and IgE-independent anaphylaxis. In IgE-mediated anaphylaxis, allergen cross-links allergen-specific IgE on the surface of MCs and basophils, leading to their degranulation. Release of mediators causes bronchoconstriction, mucus secretion, diminished cardiac contractility, increased vascular permeability, vasoconstriction of coronary and peripheral arteries, and vasodilation of venules, thereby producing clinical symptoms of anaphylaxis. IgE-mediated reactions occur in presensitized patients (e.g., those experiencing penicillin-, insulin-, latex-, or peanut-induced anaphylaxis). IgG- or IgM-related transfusion reactions should be classified as immunological, IgE-independent anaphylaxis. In contrast, opioids, radiocontrast media, vancomycin, and some muscle relaxants are capable of directly inducing HR from basophils and MCs without involvement of IgE. Although reactions to NSAIDs are considered pharmacological rather than immunological (because of the downstream effects of COX inhibition), an IgE-mediated mechanism has been suspected in some patients. In murine models IgG-mediated FcγRIII-dependent anaphylaxis elicited by a high dose of allergen has been described, however, there is not yet definitive evidence of IgG-mediated anaphylaxis in humans. Monocytes and macrophages are likely to play a role in in this type of anaphylaxis; however, the extent of their contribution is yet to be established. Cytokine storm–like reactions were reported in patients with chemotherapy-induced anaphylaxis.

Etiology of Anaphylaxis

Anaphylaxis is most commonly caused by foods, drugs, general anesthetic agents, insect stings, and, rarely, latex. Exercise can occasionally cause anaphylaxis either on its own (exercise-induced anaphylaxis) or after ingestion of a food in presensitized individuals (food- and exercise-induced anaphylaxis). The incidence and severity of anaphylaxis in patients with mastocytosis is four to six times higher than that in the general population. Up to 50% of patients with mastocytosis present with anaphylaxis during their lifetime.[38] Anaphylaxis occurs in the perioperative setting. Idiopathic anaphylaxis is diagnosed when no cause can be found.

Some anaphylactic cases are multifactorial. Cofactors are thought to lower the threshold for the induction of anaphylaxis and are implicated in about 30% of anaphylaxis cases in adults. The risk of severe anaphylaxis is increased in patients taking beta-blockers, ACEIs, or both. Rare causes include vaccines, semen, and aeroallergen inhalation.

The most common routes of allergen exposure are oral and parenteral, although inhalation of allergens (e.g., fish or legume allergens after cooking, latex particles in healthcare settings) or percutaneous penetration after skin contact can induce anaphylaxis in highly sensitized patients.

Food-Induced Anaphylaxis

According to a meta-analysis of 34 studies, food-induced anaphylaxis was reported to occur with an incidence of 0.14 cases per 100,000 person-years at all ages and up to seven per 100 person-years in children aged 0 to 4 years.[39] Fatal food-induced anaphylaxis causes from 0.03 to 0.3 deaths per million inhabitants per year.[40] According to the European Anaphylaxis Registry, foods are a predominant cause of anaphylaxis in children and young adults.[40] Young adults with a history of asthma, previously known food allergy, particularly to peanuts/tree nuts, are at higher risk of fatal food-induced anaphylaxis.

Peanuts, tree nuts, fish, and shellfish are the most frequent culprits in food-induced anaphylaxis, but almost any food can be implicated. Many cases of severe anaphylaxis are caused by unintended exposure to hidden food allergens. In addition, cofactors, such as alcohol, NSAIDs, exercise, and stress may increase the severity of a food-induced allergic reaction.[41]

Drug-Induced Anaphylaxis

Drug-induced anaphylaxis is more common in hospitalized patients than in the community. The incidence of drug-induced anaphylaxis is estimated at 0.04% to 3.1%[42] with the mortality rate of 0.65%. NSAIDs and antibiotics are the most common causes for drug-induced anaphylaxis. Anaphylaxis has been reported after treatment with monoclonal antibodies (basiliximab, rituximab, infliximab, omalizumab, etc.). IgE antibodies to the oligosaccharide galactose-α-1,3-galactose (α-gal) in mammalian meat and certain monoclonal antibodies (cetuximab, epidermal growth factor receptor inhibitor used for some cancers) may cause anaphylaxis with a delay of several hours on first exposure to cetuximab.

All routes of administration can potentially be fatal, including oral, intravenous, subcutaneous, intraarticular, intrauterine, inhalational, rectal, or topical, but the risk is greatest after parenteral administration. Patients older than 50 years with preexisting cardiovascular diseases are at higher risk of fatal drug-induced anaphylaxis.

Perioperative Anaphylaxis

The estimated incidence of perioperative anaphylaxis is 1:11,752.[43] Neuromuscular-blocking agents and antibiotics are the most common causes for perioperative anaphylaxis. Other causes may include blood and blood products, dyes, chlorhexidine, colloids or, rarely, natural rubber latex. Reactions to neuromuscular-blocking agents mostly occur on first exposure and have been associated with a 70% rate of cross-reactivity in this group; risk factors for fatal anaphylaxis include male gender, history of cardiovascular disease, emergency setting, and use of beta-blockers. The antiseptic chlorhexidine is increasingly recognized as a cause of IgE-mediated perioperative anaphylaxis.

Insect Sting–Induced Anaphylaxis

Insect sting–induced anaphylaxis is reported in 3% of adults and less than 1% of children.[44] Insect stings account for 10% to 20% of anaphylaxis but up to 50% of severe anaphylaxis.[45] Occupational anaphylaxis caused by venom allergy can occur in beekeepers, gardeners, forestry or greenhouse workers, farmers, truck drivers, and masons. The severity of reaction depends on the type of insect, amount of venom, location of sting, the patient's sensitivity, older age, preexisting diseases, previous less-severe systemic reactions, concomitant treatment, MC disorder, and elevated baseline tryptase. Allergen-specific immunotherapy with venom extracts has been shown to be safe and effective in patients with *Hymenoptera* venom allergy, providing some clinical protection within the first 8 weeks of treatment and a long-lasting effect after 3 to 5 years of maintenance treatment (see Chapter 47). Noteworthy, patients with systemic mastocytosis are at risk of potentially fatal anaphylaxis to insect stings even if they are not presensitized to venom: this may be attributed to venom components, such as phospholipase A_2, acting as MC liberators.

Other Rare Causes of Anaphylaxis

Anaphylaxis occurs during 1 in 20,000 to 47,000 transfusions of blood or blood products, especially in patients with IgA deficiency. IgA deficiency affects 1 in 500 to 700 Caucasians. One-third of these patients have circulating anti-IgA antibodies, which are associated with serious life-threatening anaphylactic reactions to blood products containing IgA. Seminal fluid allergy is extremely rare, mostly affecting young women with atopy, with 20% of cases developing life-threatening anaphylaxis. These reactions can be prevented with condom use or intravaginal desensitization with seminal fluid. Latex-induced anaphylaxis is now rarely reported but can be fatal. More than half the patients with latex allergy report allergic reactions to fruits, such as banana, avocado, kiwi fruit, chestnut, pear, pineapple, grape, and papaya. Bites by wild or laboratory animals can rarely cause anaphylaxis in sensitized individuals.

Anaphylaxis in Clonal Mast-Cell Disorders

There is a link between unexplained anaphylaxis and clonal mast-cell disease (systemic mastocytosis or monoclonal MC activation syndrome). In systemic mastocytosis, anaphylaxis may occur to *Hymenoptera* stings and drugs (NSAIDs, opioids, and drugs used in the perioperative setting, biologics, radiocontrast media, rarely vaccines). Patients with unexplained anaphylaxis should be evaluated for mast-cell clonality to exclude systemic mastocytosis or mast-cell activation syndrome.

Clinical Diversity of Anaphylaxis

Anaphylaxis can be preceded by prodromal symptoms, such as tingling and redness of the palms and soles, anxiety, sense of impending doom, and disorientation. Anaphylaxis most commonly begins in skin and mucous membranes, is followed by involvement of the respiratory and GI tracts and the cardiovascular system, and may finally proceed to cardiac and/or respiratory arrest. Generalized wheals with angioedema is the most common manifestation of anaphylaxis, observed in over 90% of cases, but may be absent. Respiratory symptoms may vary from rhinitis to potentially life-threatening laryngeal edema and airway obstruction. Cardiovascular manifestations in anaphylaxis include hypotension and/or cardiac arrhythmias. In adults, reduced BP is regarded as systolic BP of less than 90 mm/Hg or more than 30% decrease from that person's baseline values.[46] Some patients present with only cardiovascular collapse in the absence of other signs of anaphylaxis, especially during general anesthesia. Anaphylaxis is usually associated with tachycardia caused by increased cardiac sympathetic drive in response to a decreased effective vascular volume, but bradycardia may also occur. Anaphylaxis may result in up to 35% of intravascular fluid leaking into the extracellular space. A two-phase reaction to the hypovolemia may present with tachycardia as the first phase, followed by bradycardia when effective blood volume falls by 20% to 30%.

CLINICAL PEARLS

Diagnosis of Anaphylaxis

- Anaphylaxis—an acute, potentially life-threatening systemic allergic reaction with a wide range of clinical manifestations (see diagnostic criteria in Anaphylaxis section).
- It is nearly always accompanied by tachycardia, usually by flushing, urticaria, and panic, and sometimes by vomiting and diarrhea.
- Panic attacks do not cause airways obstruction, hypotension, or urticaria but may be accompanied by faintness or tetany of the hands as a result of hyperventilation.
- Vasovagal attacks present with fainting, nausea, slow pulse, and pallor without respiratory difficulty, diarrhea, or urticaria.

Diagnosis of Anaphylaxis

Measurement of blood tryptase is now widely used as a marker of MC degranulation for in vitro confirmation of anaphylaxis. Beta-tryptase is released from MCs, but not from basophils, and diffuses more slowly compared with histamine. The concentration of tryptase peaks 1 to 2 hours after the onset of reaction, with a half-life of approximately 2 hours. Samples for tryptase testing should be collected as soon as possible after emergency treatment of the patients and within 1 to 2 hours (but not later than 4 hours) of anaphylaxis onset, and again after 24 hours (baseline sample) to check that the value has returned to normal.

Normally, mature tryptase is below detection limits in the serum of healthy subjects, whereas it is elevated in most cases of anaphylaxis with vascular compromise, especially if it is parenterally induced. However, a normal tryptase result does not exclude anaphylaxis. Tryptase within the normal range during anaphylaxis is often observed in food-induced anaphylaxis and can occur in indolent mastocytosis. In infants, tryptase may not be elevated after anaphylaxis, although the baseline levels may be increased. The diagnostic value of the measurements of histamine, platelet-activating factor, chymase, carboxypeptidase A3, dipeptidyl peptidase I, basogranulin, and CCL-2 in anaphylaxis is under investigation.

In anaphylaxis, component-resolved diagnosis can be useful to stratify risk in certain clinical scenarios. For example, patients with wheat-dependent, exercise-induced anaphylaxis should be tested for omega-5-gliadin sensitization, whereas patients with anaphylaxis to vegetables, fruits, nuts, and cereals may have IgE to nonspecific lipid transfer proteins (mostly Pru p 3 and Tri a 14). In delayed anaphylaxis to mammalian meat or anaphylaxis to cetuximab, tests for IgE against galactose-α-1,3-galactose (α-gal) should be considered.

Differential diagnosis in anaphylaxis includes capillary leak syndrome, postural orthostatic tachycardia syndrome, carcinoid syndrome, neuroendocrine tumors, vasovagal reactions, seizures, and intoxications.

Management of Anaphylaxis

Early recognition of anaphylaxis facilitates removal of the cause and prompt institution of treatment. The patient with anaphylaxis should lie down with the lower limbs elevated to increase venous blood return and maintain cardiac output. Changes in posture may trigger decompensation and fatal outcome. In drug-induced or insect-induced anaphylaxis a tourniquet may be placed proximal to the site of the injection or insect sting to slow absorption of injected antigens. The tourniquet should be released for 3 minutes at 5-minute intervals, with the total duration of application not exceeding 30 minutes.

Epinephrine should be administered as first-line treatment by an intramuscular injection in the mid-outer thigh at the first sign of respiratory failure or cardiovascular collapse and repeated after 5 to 15 minutes if the response to the first injection is suboptimal. Epinephrine autoinjectors for self-administration are available, but a single epinephrine autoinjector may be insufficient to reverse severe reactions. Use of these epinephrine autoinjectors in anaphylaxis outside hospital can be lifesaving. Overall, prompt diagnosis of anaphylaxis, early administration of epinephrine, and fast transport to emergency rooms are crucial factors for successful management.

Epinephrine is both an α- and β-adrenergic agonist with cyclic adenosine monophosphate (cAMP)–mediated pharmacological effects on target organs. In patients with anaphylaxis, stimulation of α_1-adrenergic receptors increases peripheral vascular resistance, thereby improving BP and coronary perfusion, reversing peripheral vasodilation, and decreasing angioedema. Activation of β_1-adrenergic receptors increases myocardial contractility (inotropy, chronotropy) while stimulation of β_2-adrenoreceptors causes bronchodilation and decreases the release of inflammatory mediators from MCs and basophils.

Current guidelines recommend the intramuscular route for epinephrine administration because of faster absorption and higher plasma level of epinephrine after intramuscular injection compared with subcutaneous injection.[46] The appropriate dosage of epinephrine is 0.01 mg/kg of a 1:1000 (1 mg/mL) solution, to a maximum of 0.5 mg in adults and 0.3 mg in children (0.15 mg between 6 months and 5 years). Epinephrine has a rapid but short action, so the dose may need to be repeated every 5 to 15 minutes until symptoms improve. Intravenous administration of epinephrine should be reserved for severe anaphylaxis with profound life-threatening hypotension that is refractory to other treatment because of a risk of potentially fatal cardiac arrhythmias and myocardial infarction.

Common pharmacological adverse effects of epinephrine include anxiety, fear, headache, pallor, tremor, dizziness, and palpitation. In the event of an overdose, unwanted effects may include increased Q–T interval on electrocardiography, ventricular arrhythmias, angina, myocardial infarction, increased BP, pulmonary edema, and intracranial hemorrhage. Patients with cardiovascular diseases or thyrotoxicosis and cocaine users are particularly susceptible to the adverse effects of epinephrine.

The efficacy of epinephrine can be decreased by concomitant therapy with beta-blockers, which is associated with unopposed stimulation of α-adrenoreceptors and reflex vagotonic effects, leading to bradycardia, hypertension, coronary artery constriction, bronchoconstriction, and augmented mediator release. Anaphylaxis in patients on beta-blockers can be severe, protracted, and unresponsive to treatment. Patients treated with beta-blockers may require fluid replacement and treatment with glucagon, which increases intracellular cAMP independently of β-adrenergic receptors. Glucagon may improve hypotension in 1 to 5 minutes with maximal effect at 5 to 15 minutes. Side effects of glucagon include nausea and vomiting.

Corticosteroids are often administered in anaphylaxis to minimize the risk of recurrent anaphylaxis, although convincing evidence is lacking and it is still unclear how they may work for prevention of biphasic anaphylaxis.

If there is no response to epinephrine, life support measures should be instituted. The treatment choice depends on the clinical presentation. In hypotension, large volumes of fluids should be given rapidly using 0.9% normal saline to compensate for peripheral vasodilatation and for fluid loss into the extravascular space. Crystalloid normal saline should be preferred to colloids for intravenous fluid resuscitation. Other vasopressors (dopamine, glucagon) may be needed to reverse severe hypotension. Oxygen should also be administered in circulatory or respiratory failure. Bronchospasm should be treated with nebulized or inhaled β_2-agonists. If there is severe laryngeal edema, endotracheal intubation and even emergency tracheostomy may be needed to maintain the airway. Methylene blue, a selective nitric oxide cyclic guanosine monophosphate (cGMP) inhibitor, can prevent vasodilatation and has been reported to be effective in refractory anaphylaxis. Extracorporeal life support (extracorporeal membrane oxygenation) can be used in patients with refractory anaphylaxis.

Patients presenting with severe anaphylaxis or those who require more than one dose of epinephrine should be considered for longer observation time. Patients with a higher risk of biphasic anaphylaxis or risk factors for anaphylaxis fatality should be observed for up to 6 hours or longer.

Prevention of Anaphylaxis

The first step in prevention is to identify those at risk of anaphylaxis using predictors from epidemiological and clinical studies. Machine-learning approaches have been attempted for prediction of anaphylaxis.[47] Therefore, all patients with a history of anaphylaxis should be referred for assessment and undergo allergy evaluation. Patients should be instructed how to avoid culprit allergens and cross-reactive agents and should be advised on safe alternatives. The education of patients, their families, and, in the case of children, caregivers and school staff, about anaphylaxis and availability of first-aid measures is of primary importance. Written personalized emergency action plans should be provided to patients at special risk, such as school children. Emergency medications, such as epinephrine autoinjectors, should be dispensed, and patients should receive training on their correct use. Patients should be advised to carry an epinephrine autoinjector with them at all times. Immunotherapy is very effective for prophylaxis of bee- and wasp venom–induced anaphylaxis in sensitized patients and can be lifesaving. Drug-induced anaphylaxis can be prevented by avoidance of culprit drugs and cross-reacting agents. Current guidelines support the use of premedication with glucocorticosteroids and/or antihistamines to prevent anaphylaxis or infusion-related reactions for specific agents in chemotherapy protocols although premedication prior to radiocontrast media use is not supported. In rare cases, drug desensitization can be used for antibiotics, chemotherapeutic agents, insulin, vaccines, and biological agents. For food-induced anaphylaxis, avoidance of the culprit food is essential; oral immunotherapy is available in some allergy centers. In idiopathic anaphylaxis, patients with frequent episodes (more than six episodes per year or two or more episodes within 2 months) can be treated with steroids to prevent further episodes.

Omalizumab have been reported to be effective in preventing anaphylaxis in patients with systemic mastocytosis, exercise-induced anaphylaxis, intraoperative anaphylaxis, anaphylaxis during allergen-specific immunotherapy, and idiopathic anaphylaxis. IgE immunoadsorption has been reported to decrease the reactivity thresholds to foods in food-induced anaphylaxis.[48]

TRANSLATIONAL RESEARCH OPPORTUNITIES

ON THE HORIZON

- Novel insights into the autoimmune mechanisms and understanding the mast-cell activation signals in chronic spontaneous urticaria should improve clinical assessment and management of patient subgroups.
- Development of new bradykinin and kallikrein inhibitors for patients with hereditary angioedema should further improve the acute management of this rare but very important condition. Gene therapy for hereditary angioedema may be an attractive treatment option in the future.
- Understanding the full clinical spectrum of patients with cryopyrin-associated periodic syndrome with *NLRP-3* mutations should allow earlier identification and treatment of these individuals with interleukin-1 blockers to improve quality of life and prevent later complications.
- Novel prevention strategies may include off-label use of omalizumab for different types of anaphylaxes. Machine-learning approaches may improve the prediction of anaphylaxis in the future.

REFERENCES

1. Maurer M, Abuzakouk M, Bérard F, et al. The burden of chronic spontaneous urticaria is substantial: real-world evidence from ASSURE-CSU. *Allergy*. 2017;72:2005–2016.
2. Zuberbier T, Aberer W, Asero R, et al. The EAACI/GA²LEN/EDF/WAO guideline for the definition, classification, diagnosis, and management of urticaria. *Allergy*. 2018;73(7):1393–1414.
3. Fricke J, Ávila G, Keller T, et al. Prevalence of chronic urticaria in children and adults across the globe: systematic review with meta-analysis. *Allergy*. 2020;75:423–432.
4. Netchiporouk E, Sasseville D, Moreau L, et al. Evaluating comorbidities, natural history and predictors of early resolution in a cohort of children with chronic urticaria. *JAMA Dermatol*. 2017;153(12):1236–1242.
5. Losol P, Yoo H-S, Park H-S. Molecular genetic mechanisms of chronic urticaria. *Allergy Asthma Immunol Res*. 2014;6(1):13–21.
6. Brzoza Z, Grzeszczak W, Rogala B, et al. PTPN22 polymorphism presumably plays a role in the genetic background of chronic spontaneous autoreactive urticaria. *Dermatology*. 2012;224(4):340–345.

7. Ombrello MJ, Remmers EF, Sun G, et al. Cold urticaria, immunodeficiency, and autoimmunity related to *PLCG2* deletions. *N Engl J Med*. 2012;366(4):330–338.
8. Kuemmerle-Deschner JB, Ozen S, Tyrrell PN, et al. Diagnostic criteria for cryopyrin-associated periodic syndrome. *Ann Rheum Dis*. 2017;76(6):942–947.
9. Scheffel J, Mahnke NA, Hofman ZLM, et al. Cold-induced urticarial autoinflammatory syndrome related to factor XII activation. *Nat Commun*. 2020;11(1):179.
10. Kay AB, Ying S, Ardelean E, et al. Elevations in vascular markers and eosinophils in chronic spontaneous urticarial weals with low-level persistence in uninvolved skin. *Br J Dermatol*. 2014;171(3):505–511.
11. Varricchi G, Pecoraro A, Loffredo S, et al. Heterogeneity of human mast cells with respect to MRGPRX2 receptor expression and function. *Front Cell Neurosci*. 2019;13:299.
12. Schmetzer O, Lakin E, Topal FA, et al. IL-24 is a common and specific autoantigen of IgE in patients with chronic spontaneous urticaria. *J Allergy Clin Immunol*. 2018;142(3):876–882.
13. Kolkhir P, Church MK, Weller K, et al. Autoimmune chronic spontaneous urticaria: what we know and what we do not know. *J Allergy Clin Immunol*. 2017;139(6):1772–1781.
14. Pastore S, Berti I, Longo G. Autoimmune chronic urticaria: transferability of autologous serum skin test. *Eur J Pediatr*. 2013;172:569.
15. Schoepke N, Asero R, Ellrich A, et al. Biomarkers and clinical characteristics of autoimmune chronic spontaneous urticaria: results of the PURIST study. *Allergy*. 2019;74(12):2427–2436.
16. McNeil BD, Pundir P, Meeker S, et al. Identification of a mast cell specific receptor crucial for pseudoallergic drug reactions. *Nature*. 2015;519(7542):237–241.
17. Fujisawa D, Kashiwakura J, Kita H, et al. Expression of mas-related gene X2 on mast cells is upregulated in the skin of patients with severe chronic urticaria. *J Allergy Clin Immunol*. 2014;134(3):622–633.
18. 1Craig TJ, Bernstein JA, Farkas H, et al. Diagnosis and treatment of bradykinin-mediated angioedema: outcomes from an angioedema expert consensus meeting. *Int Arch Allergy Immunol*. 2014;165:119–127.
19. Veraldi S, Romagnuolo M, Benzecry V. Urticaria as a first clinical manifestation of COVID-19. *Eur J Dermatol*. 2020;30(6):737–738.
20. Kolkhir P, Borzova E, Grattan C, et al. Autoimmune comorbidity in chronic spontaneous urticaria: a systematic review. *Autoimmunity Rev*. 2017;16(12):1196–1208.
21. Magerl M, Altrichter S, Borzova E, et al. The definition, diagnostic testing, and management of chronic inducible urticarias—The EAACI/GA2LEN/EDF/UNEV consensus recommendations 2016 update and revision. *Allergy*. 2016;71(6):780–802.
22. Maurer M, Metz M, Brehler R, et al. Omalizumab treatment in patients with chronic inducible urticaria: a systematic review of published evidence. *J Allergy Clin Immunol*. 2018;141(2):638–649.
23. Boyden SE, Desai A, Cruse G, et al. Vibratory urticaria associated with a missense variant in *ADGRE2*. *N Engl J Med*. 2016;374(7):656–663.
24. Maltseva N, Borzova E, Fomina D, et al. Cold urticaria—what we know and what we don't know. *Allergy*. 2020 https://doi.org/10.1111/all.14674.
25. Haylett AK, Koumaki D, Rhodes LE. Solar urticaria in 145 patients: assessment of action spectra and impact on quality of life in adults and children. *Photodermatol Photoimmunol Photomed*. 2018;34(94):262–268.
26. Fukunaga A, Washio K, Hatakeyama M, et al. Cholinergic urticaria: epidemiology, pathophysiology, new categorization, and management. *Clin Auton Res*. 2018;28:103–113.
27. Crochet J, Lepelley M, Yahiaoui N, et al. Bradykinin mechanism is the main responsible mechanism for death by isolated asphyxiating angioedema in France. *Clin Exp Allergy*. 2019;49:252–254.
28. Aygoren-Pürsün E, Magerl M, Maetzel A, Maurer M. Epidemiology of bradykinin-mediated angioedema: a systematic investigation of epidemiological studies. *Orphanet J Rare Dis*. 2018;13:73.
29. Craig TJ, Bernstein JA, Farkas H, et al. Diagnosis and treatment of bradykinin-mediated angioedema: outcomes from an angioedema expert consensus meeting. *Int Arch Allergy Immunol*. 2014;165:119–127.
30. Guillén-Aguinaga S, Jáuregui Presa I, Aguinaga-Ontoso E, et al. Updosing nonsedating antihistamines in patients with chronic spontaneous urticaria: a systematic review and meta-analysis. *Br J Dermatol*. 2016;175(6):1153–1165.
31. Kaplan AP, Giménez-Arnau AM, Saini SS. Mechanisms of action that contribute to efficacy of omalizumab in chronic spontaneous urticaria. *Allergy*. 2017;72(4):519–533.
32. Maurer M, Magerl M, Ansotegui I, et al. The international WAO/EAACI guideline for the management of hereditary angioedema—the 2017 revision and update. *Allergy*. 2018;73:1575–1596.
33. Haslund D, Ryø LB, Seidelin Majidi S, et al. Dominant-negative SERPING1 variants cause intracellular retention of C1 inhibitor in hereditary angioedema. *J Clin Invest*. 2019;129(1):388–405.
34. Shaker MS, Wallace D, Golden DBK, et al. Anaphylaxis—a 2020 practice parameter update, systematic review, and Grading of Recommendations, Assessment, Development and Evaluation (GRADE) analysis. *J Allergy Clin Immunol*. 2020;145:1082–1123.
35. Panesar SS, Javad S, de Silva D, et al. The epidemiology of anaphylaxis in Europe: a systematic review. *Allergy*. 2013;68:1353–1361.
36. Wood RA, Camargo Jr CA, Lieberman P, et al. Anaphylaxis in America: the prevalence and characteristics of anaphylaxis in the United States. *J Allergy Clin Immunol*. 2014;133:461–467.
37. Turner PJ, Campbell DE. Epidemiology of severe anaphylaxis: can we use population-based data to understand anaphylaxis? *J Allergy Clin Immunol*. 2016;16(5):441–450.
38. Carter MC, Metcalfe DD, Matito A, et al. Adverse reactions to drugs and biologics in patients with clonal mast cell disorders: A Work Group Report of the Mast Cell Disorder Committee, American Academy of Allergy, Asthma & Immunology. *J Allergy Clin Imunol*. 2019;143:880–893.
39. Umasunthar T, Leonardi-Bee J, Turner PJ, et al. Incidence of food anaphylaxis in people with food allergy: a systematic review and meta-analysis. *Clin Exp Allergy*. 2015;45(11):1621–1636.
40. Pouessel G, Turner PJ, Worm M, et al. Food-induced fatal anaphylaxis: from epidemiological data to general prevention strategies. *Clin Exp Allergy*. 2019;48:1548–1593.
41. Worm M, Scherer K, Kohli-Wiesner A, et al. Food-induced anaphylaxis and cofactors—data from the anaphylaxis registry. *Allergologie*. 2017;1:21–27.
42. Montañez MI, Mayorga C, Bogas G, et al. Epidemiology, mechanisms, and diagnosis of drug-induced anaphylaxis. *Front Immunol*. 2017;8:614.
43. Harper N, Cook TM, Garcez T, et al. Anaesthesia, surgery and life threatening allergic reactions. Epidemiology and clinical features of perioperative anaphylaxis: the 6th National Audit Project. *Br J Anaesthesia*. 2018;121(1):159–171.
44. Golden DB. Anaphylaxis to insect stings. *Immunol Allergy Clin North Am*. 2015;35(2):287–302.
45. Turner PJ, Jerschow E, Umasunthar T, et al. Fatal anaphylaxis: mortality rate and risk factors. *J Allergy Clin Immunol Pract*. 2017;5(5):1169–1178.
46. Shaker MS, Wallace DV, Golden DBK, et al. Anaphylaxis—a 2020 practice parameter update, systematic review, and Grading of Recommendations, Assessment, Development and Evaluation (GRADE) analysis. *J Allergy Clin Immunol*. 2020;145:1082–1123.
47. Segura-Bedmar I, Colón-Ruíz C, Tejedor-Alonso MÁ, et al. Predicting of anaphylaxis in big data EMR by exploring machine learning approaches. *J Biomed Inform*. 2018;87:50–59.
48. Dahdah L, Ceccarelli S, Amendola S, et al. IgE immunoadsorption knocks down the risk of food-related anaphylaxis. *Pediatrics*. 2015;136(6):e1617–1620.

47

Allergic Reactions to Stinging and Biting Insects

Anna Gschwend and Arthur Helbling

ENTOMOLOGICAL ASPECTS[1,2]

Stinging hymenoptera belong to the suborder Aculeata, with the families Apideae, Vespidae, Formicidae (Table 47.1).

Apidae[1]

In this family, the honeybee (*Apis mellifera*) (Fig. 47.1A) is clinically the most important cause of allergies. Bumblebees (see Fig. 47.1B) may occasionally cause allergic sting reactions (see Table 47.1).

Vespidae[3]

The vespids are divided into the subfamilies Vespinae and Polistinae, which differ morphologically in the junction between thorax and abdomen (see Fig. 47.1C and D). The subfamily Vespinae contains the three genera *Vespula, Dolichovespula*, and *Vespa* (see Table 47.1). *Vespula* are called *wasps* in Europe and *yellow jackets* in the United States. Most species of *Dolichovespula* look very similar to *Vespula*, with black and yellow stripes on the abdomen and only a slightly larger size. They can be distinguished from *Vespula* by the larger distance between the eyes and the mandibles. The genus *Vespa* (European hornet) is easy to distinguish from other vespids due to its much larger size. The Asian hornet (*V. velutina*) is naturally distributed in Asia from Afghanistan to eastern China, Indochina, and Indonesia. *Vespa velutina* is one of the most aggressive Hymenoptera species in China, where it is known as killer-wasp. *Vespa velutina nigrithorax* (VVN) has spread rapidly across France and the Basque Country in the North of Spain, Portugal, Italy, and the UK. Anaphylactic Reactions due to VVN are an emerging problem in Spain.[4] The species *Polistes dominula* and *Polistes gallicus* are European paper wasps and of particular relevance in Mediterranean areas of Europe; *P. dominula* has also spread to the northeastern United States and been reported in Australia. Polybia wasps, particularly *P. paulista*, are present in South America.[3]

Ants (Formicidae)[2]

In South and Central America and in the southern states of the United States, fire ants (Fig. 47.2) are responsible for many systemic allergic sting reactions. Occasional allergic sting reactions to *Pogonomyrmex*, the North American harvester ant, have been described, and extremely rarely to the European red ant, *Formica rufa*. The species of Myrmicinae are an important cause of allergic sting reactions in southern Australia. Another group of aggressive ants is the *Pachycondyla*, in the Far East and Middle East (see Table 47.1).

ALLERGENS IN HYMENOPTERA VENOMS

The best-characterized venom is that of the honeybee, *Apis mellifera*, with more than 100 identified proteins. Among them, Api m1, Api m2, Api m3, Api m5 and Api m 10 can be defined as major allergens for which more than 50% of patients show IgE reactivity.[5] Prominent yellow jacket venom (YJV) allergens include phospholipase A1 (Ves v 1) and antigen 5 (Ves v 5). A detailed proteomic analysis of pure *P. dominula* venom (PDV) led to the identification of 47 proteins (unpublished data). These and other allergens of Hymenoptera Venoms are listed in Table 47.2.[5] The amount of venom injected during a sting varies between and within species, especially in vespids. Bees release 50 to 140µg venom per sting, and vespids between 2 and 17µg.

CLINICAL RELEVANCE

- Not every swelling after an insect sting is an allergy.
- Systemic allergic reaction to hymenoptera venom are usually caused by immunoglobulin E ((IgE)-mediated immunological mechanisms).
- Although hymenoptera venom allergy is not as prevalent as respiratory allergies, severe systemic reactions—even fatal—occur regularly worldwide.
- It is crucial to recognize specific risk factors for hymenoptera venom allergy and identify persons at risk for severe reactions or treatment failure.
- Every patient who has experienced a systemic allergic sting reaction should be investigated in consideration of venom specific immunotherapy.
- All patients with history of systemic reaction upon stinging should be provided with adrenaline auto-injectors (AAIs).

CLINICAL PICTURE

The clinical presentations following insect stings are classified into normal, large local, systemic allergic, systemic toxic, and unusual reactions (Table 47.3).

Normal Local Reactions

The normal local reaction of a nonallergic subject to a hymenoptera sting consists of a painful, sometimes itchy, local wheal flare reaction with swelling up to 5 to 10 cm in diameter, which unusually resolves within a few hours. The fire ant (*S. invicta*) attaches to the skin by means of its powerful mandibles and stings, releasing venom that produces a characteristic fire-like pain. After the ant stings, a vesicle is left behind, which later develops into a pustule that only heals after 1 to 2 weeks.[2]

TABLE 47.1 Taxonomy and characteristics of Hymenoptera

Families/Subfamilies	Genera	Species	Characteristic
Apidae Apinae	Apis Bombus	*A. mellifera* *B. terrestris*	• Apis stings occur in the vicinity of beehives or when the insect feels threatened • Apis stings can occur in spring, summer, and occasionally on warm winter days • Apis usually lose their barbed sting when stinging
Vespidae Vespinae Polistes	Dolichovespula Vespula Vespa	*D. maculata*, *D. arenaria*, *D. media* *V. germanica* *V. vulgaris*, *V. maculifrons* *V. carbo*, *V. orientalis* *P. dominula* *P. gallicus* *P. exclamans*, *P. annularis*, *P. fuscatus* Polybia	• Dolichovespula build their nests in tree branches or under the roofs of houses • Dolichovespula sting almost exclusively in the vicinity of their nests • Vespula stings occur near the nests and while victims are eating outdoors • Vespids do not usually lose their stinger • Vespula stings occur in summer and fall • Hornet stings are rare and occur almost exclusively in the vicinity of nests • Polistes commonly build nests on human habitation • Polistes although generally not aggressive, they can be provoked into defending their nests • Some Polybia species produce enough honey to be collected and eaten by local people
Formicidae Myrmicinae	Solenopsis Myrmicinae Pachycondyla	*S. invicta* *S. richteri* *M. pilosula* *P. chinesis* *P. senna* *P. arensis*	• Fire ants build their mounds in yards/playgrounds/fields and are responsible for many systemic allergic sting reactions • Jack-jumper ants are an important cause of allergic sting reactions in southern Australia

Large Local Reactions

Large local reactions (LLRs) are defined as swellings around the sting site exceeding 10 cm in diameter and lasting more than 24 hours. LLRs may cause significant discomfort, especially when they last for days or even weeks and involve a whole limb, eyelids, or lips. The development of local infection, abscesses, or a phlegmon at the sting site is usually inhibited by the bacteriostatic effect of hymenoptera venoms.

Systemic Reactions

Systemic reactions (SRs) are usually caused by immunoglobulin E ((IgE)-mediated immunological mechanisms). Affected organs include skin, the gastrointestinal tract, the respiratory tract, and the cardiovascular system (see Table 47.3). Cutaneous involvement is very frequently observed in both adults and children, accounting for 80% and more than 90% of SR reactions, accordingly. Respiratory involvement is observed in around half of SRs.[6] The hypotension and cardiovascular collapse might occur independently of other symptoms, especially in case of systemic indolent mastocytosis. Recurrence of symptoms after 4 to 12 hours from the resolution of the first anaphylactic episode, without reexposure to stings (biphasic anaphylaxis), is reported in 0.4% to 14.7% of cases. Known risk factors for biphasic reactions are history of previous anaphylactic episodes and delayed treatment with adrenaline.[7]

Several classifications were proposed to assess the degree of severity of anaphylaxis; the most used in clinical practice are Mueller's and Ring's.[6]

Systemic Toxic Reactions

Toxic reactions are dose-dependent. Toxic reactions and allergic reactions are etiologically different conditions. Clinically significant toxic effects appear after multiple stings—usually 50 to several hundred—and develop within hours to days (see Table 47.3). The number of stings needed to cause a fatal reaction in adults varies between 200 and 1000. In small children, however, fewer than 50 stings may be lethal. In most cases, death is not immediate but occurs after several days.

Unusual Reactions

Unusual sting reactions are rare and appear after hours to days, and more than half of them follow local or SRs.[8] Non–IgE-mediated immunological mechanisms are likely involved. The causal relation to the sting event often remains uncertain (see Table 47.3).

EPIDEMIOLOGICAL ASPECTS

Prevalence of Allergy to Hymenoptera

Cumulative lifetime sting rates of 61% to 95% have been reported in people aged 16 to 65 years. Of course, this can vary considerably in different regions of the world. Asymptomatic sensitization with the development of specific IgE to bee and wasp venoms is a frequent finding and has been reported in up to 40% of individuals.[9] Hymenoptera venom allergy (HVA) can occur at any age. In general, because of their higher level of outdoor activities, men are more frequently stung than women, and children are more frequently stung than adults. The reported cumulative lifetime prevalence of LLRs ranges from 2% to 26%, and that of SRs up to 8.9% in the adult European population. Among beekeepers, it varies between 14% and 32%. SRs are more frequently caused by vespids than by honeybees.[9] In the southern states of the United States and in Australia, ants are important causes of SR.[2]

Risk Factors for Developing Hymenoptera Allergy

The risk of developing a sting allergy increases with the number of stings, especially if two stings occur within a short period (up to 2 months). However, beekeepers stung less than 10 times a year have a much higher risk for SRs compared with those stung

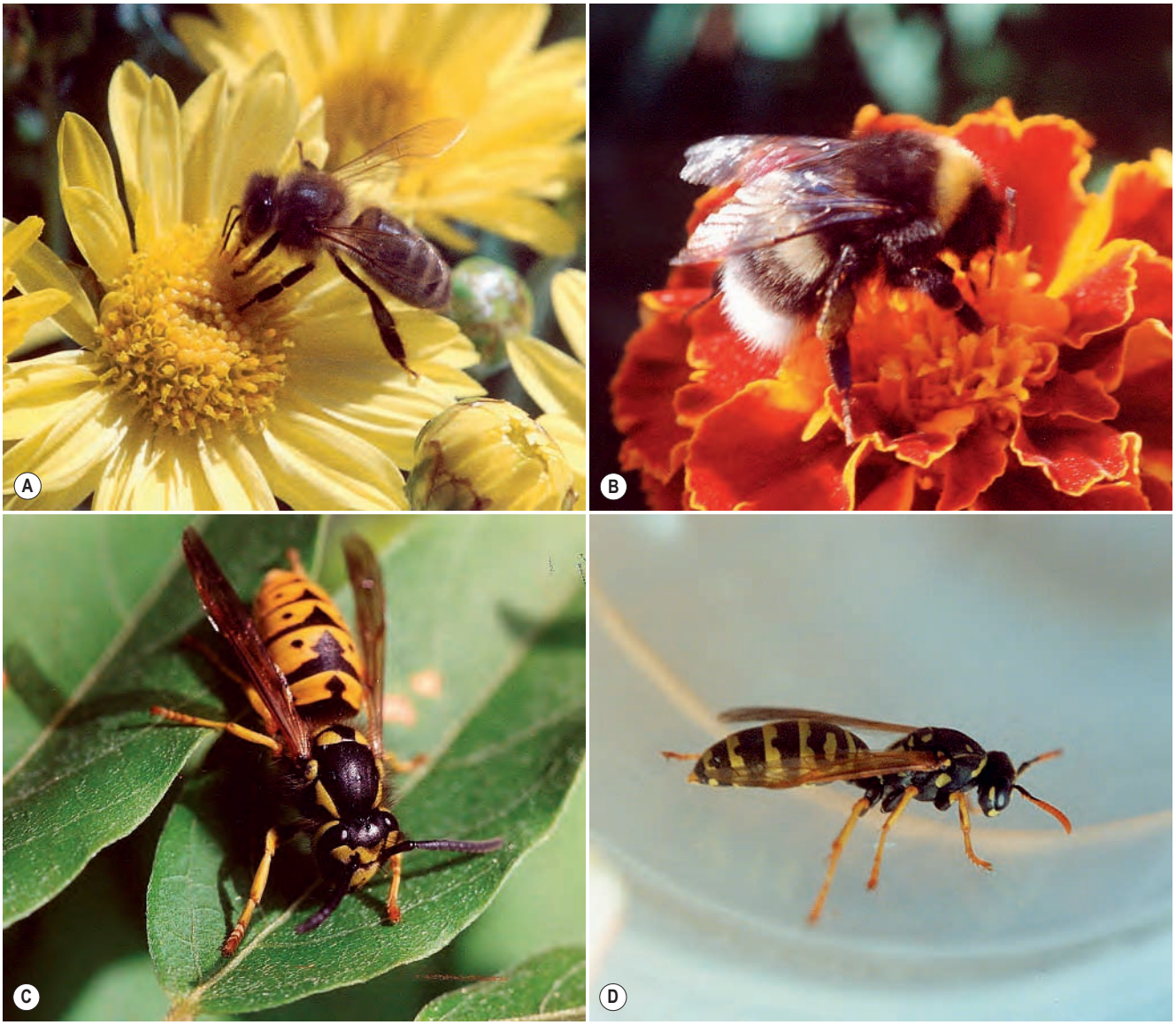


FIG. 47.1 Common hymenoptera. (A) Honeybee (*Apis mellifera*), (B) bumblebee (*Bombus terrestris*), (C) wasp/yellow jacket (*Vespula* spp.), (D) field wasp (*Polistes gallicus*). (Courtesy US Department of Agriculture.)

FIG. 47.2 Biting insect that may cause allergic reactions (fire ant—*Solenopsis invicta*).

more than 200 times a year.[10] Atopy is not more frequent in patients with hymenoptera sting allergy than in the general population. Systemic mastocytosis (indicated by elevated baseline serum tryptase) is a risk factor for developing a Hymenoptera venom allergy.[11]

Natural History of Hymenoptera Allergy and Risk Factors For Severe Reactions[9,12]

There is currently no reliable test that predicts the risk of future SRs to hymenoptera venom. The severity of an index sting reaction is an important factor determining the risk at re-exposure. Children are at lower risk of re-sting SR compared with adults. After a mild SR, the risk is between 15% and 30%, and after a severe SR, the risk is between 50% and 75%. The re-sting SR risk is definitely lower in patients with Vespula allergy than in those with bee venom allergy, probably because of the smaller and more variable amount of venom injected.

TABLE 47.2 Allergens of Hymenoptera Venoms

Recombinant Allergens

Apinae	Vespinae	Polistes	Ants
Bee	Vespen	Polistes	Solenopsis
Api m1 (PhospholipaseA2)	Ves v1 (Phospholipase A1)	Pol d1 (Phospholipase A1)	Sol i1
Api m2 (Hyaluronidase)	Ves v2 (Hyaluronidase)	Pol d2 (Hyaluronidase)	Sol i2
Api m3 (Acid phosphatase)	Ves v3 (Dipeptidylpeptidase)	Pol d3 (Dipeptidylpeptidase)	Sol i3
Api m4 (Melittin)	Ves v5 (Antigen 5)	Pol d4 (Serine protease)	Sol i4
Api m5 (Dipeptidylpeptidase)	Ves v6 (Vitellogenin)	Pol d5 (Antigen 5)	Myrmecia
Api m6 (Trypsin inhibitor)	Vespa	Pol a1 (Phospholipase A1)	Myr p1
Api m7 (CUBserine protease)	Vesp c1 (Phospholipase A1)	Pol a2 (Hyaluronidase)	Myr p2
Api m8 (Carboxylesterase)	Vesp ma2 (Hyaluronidase)	Pol a4 (Serine protease)	Myr p3
Api m9 (Carboxypeptidase)	Vesp c5 (Antigen 5)	Pol a5 (Antigen 5)	Pachycond.
Api m10 (Icarapin)	Dolichovespula	Polybia	Pac c3
Api m11.0101 (MRJP8)	Dol m1 (Phospholipase A1)	Poly p1 (Phospholipase A1)	
Api m11.0201 (MRJP9)	Dol m2 (Hyaluronidase)	Poly p2 (Hyaluronidase)	
Api m12 (Vitellogenin)	Dol m5 (Antigen 5)	Poly p5 (Antigen 5)	

MRJP, Major royal jelly protein.

TABLE 47.3 Clinical Manifestations Following Insect Stings

Type of reaction	Clinical Symptoms	Pathogenesis	Venom Immunotherapy
Normal local reactions	Local wheal and flare reaction, followed by a swelling of up to 5–10 cm in diameter	Non-allergic	Not indicated
Large local reactions	Swellings around the sting site exceeding 10 cm in diameter, lasting more than 24 h. In some cases, lymphadenopathy or lymphangitis	Mostly non-allergic	Not indicated
Systemic Toxic reactions	Rhabdomyolysis, intravascular hemolysis leading to acute renal failure, myocardial damage, hepatic dysfunction, coagulation disorders, brain edema, and/or necrosis	Cytotoxic/toxic effect	Not indicated. Allergologic workup may be considered
Systemic reactions	*Skin:* pruritus, urticaria, flush, angioedema *Gastrointestinal:* cramps, vomiting or diarrhea, dysphagia *Respiratory tract:* laryngeal edema, bronchial obstruction, pulmonary edema *Cardiovascular system:* arterial hypotension, shock, arrhythmias	IgE-mediated immunological mechanisms	Allergologic workup in view to treatment with VIT
Unusual Reactions	Serum sickness–like syndromes with fever, arthralgias, exanthema, lymphadenopathy, peripheral neuropathy, polyradiculomyelitis, extrapyramidal syndromes, acute disseminated encephalomyelitis, glomerulonephritis, interstitial nephritis, hemolytic anemia, thrombocytopenia, Henoch-Schönlein syndrome, and other forms of vasculitis	Non–IgE-mediated immunological mechanisms	Not indicated. Allergologic workup may be considered

IgE, Immunoglobulin E; *VIT*, venom immunotherapy.

The degree of sensitization does not correlate with the severity of the sting reaction. Neither the amount of specific IgE antibodies to whole venom extracts nor to major allergens were significantly associated with the severity of the sting reaction.[12] Generally, patients with repeated LLRs are reported as having minimal risk of progressing to SR. Nevertheless, a recent prospective study indicates a risk of 24% SR in adults and children after an initial LLR.[13] In this study, the risk of SR was reported to be higher in cases of skin test reactivity to *Apis mellifera* or Vespula species (OR 2.1 and 3.8, respectively), if positive at 0.001 µg/mL concentration (OR 13.4 and 16.5, respectively). There is overwhelming evidence of a strong association of mast cell clonality with both a higher incidence and an increased severity of sting-induced anaphylactic reactions.[11] Moreover, older age and cardiovascular diseases are very potent risk factors for severe anaphylactic reactions in general.[12] There is some evidence that the intake of antihypertensive medication does not have an impact on the outcome of field sting-induced anaphylaxis. However, data from the European Anaphylaxis Registry suggest an increased frequency of severe reactions during concurrent intake of ACE inhibitors or beta-blockers and an even higher risk if both drugs are combined (Fig. 47.3).

Mortality Caused by Hymenoptera Stings

The estimated incidence of fatal hymenoptera sting reactions in Europe is around 200 deaths per year.[10] Up to 60% of fatal sting reactions occur in individuals who did not have a notion of being allergic to hymenoptera venom. Delayed administration of adrenaline is considered to dictate an unfavorable outcome of anaphylactic reactions regardless of the elicitor.

Epidemiological Aspects of Allergic Reactions to Ant Stings

Nearly 50% of inhabitants are stung each year in areas of the United States that are fire ant endemic.[2] Many report LLRs, but up to 1% of patients who are stung by imported fire ants develop anaphylaxis, and some deaths have been reported.

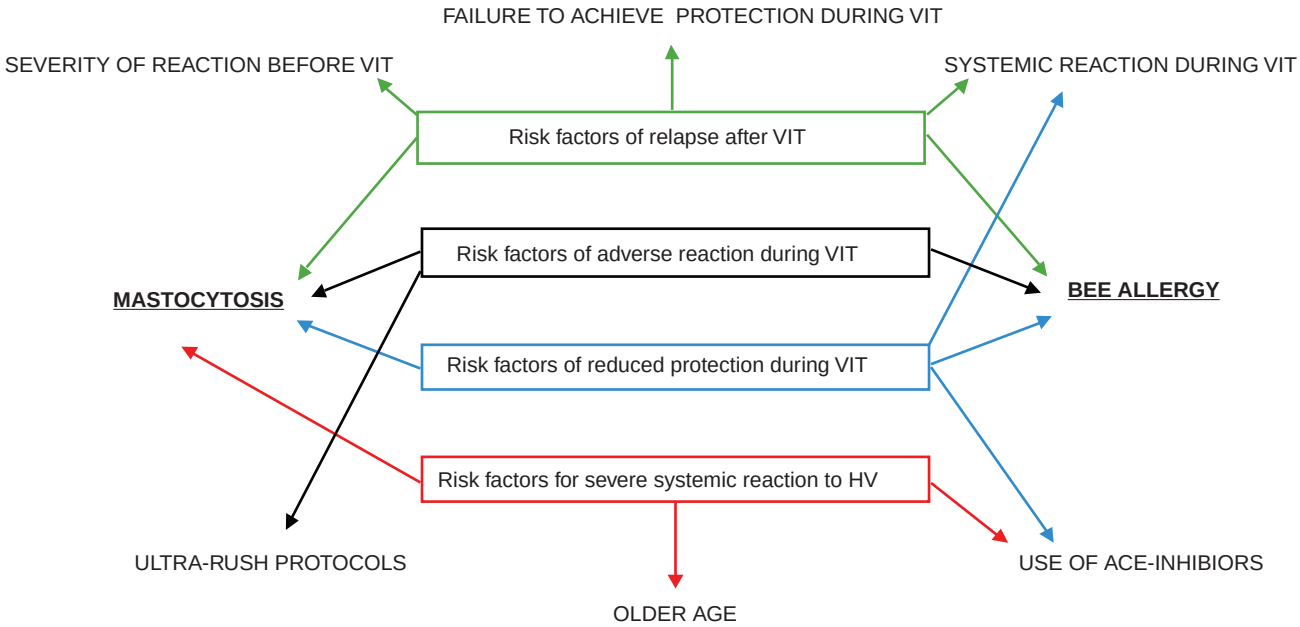


FIG. 47.3 Risk factors for severe systemic reactions to hymenoptera venom, adverse reactions, reduced protection, relapse during and after venom immunotherapy *(VIT)*.

DIAGNOSIS

Clinical History

As in all forms of allergy, the clinical history is key to diagnosing hymenoptera sting allergy. This includes the date, number, and circumstances of stings (*e.g.*, environment, activities); kind and severity of symptom; sting site; retained or removed stinger; interval to onset of symptoms; emergency treatment; risk factors for a particularly severe reaction (*e.g.*, comorbidity, drugs, elevated baseline serum tryptase/mastocytosis); tolerated stings after the first SR; reduction of the quality of life; and other allergies. For all patients with SRs, the following diagnostic tests are recommended: skin tests, venom-specific serum IgE antibodies, and baseline serum tryptase.[14]

Skin Tests[14]

Both skin tests and serologic tests should be performed in patients with a positive history of systemic reactions. If possible, skin tests should not be performed until at least 3 weeks after an SR, as false-negative results can occur (the refractory period). They are performed by intradermal or skin prick endpoint titration. Injection of 0.02 mL of venom solution is given intradermally in increasing concentrations—0.00001 to 1 µg/mL—into the volar surface of the forearm. For skin prick tests, concentrations of 0.01 to 300 µg/mL are used. However, even at 300 µg/mL, the sensitivity of skin prick is clearly lower than that of the intradermal test.

Venom-Specific Serum Immunoglobulin E Antibodies

Several different *in vitro* immunoassays for the detection of venom-specific serum IgE antibodies (sIgE) are commercially available. If no sIgE is found in a patient with a clear history of venom allergy, the test should be repeated after 2 to 4 weeks.[14]

Elevated sIgE levels after hymenoptera stings that cause a normal sting reaction are not predictive for future development of an SR. Specific IgE can be determined against the whole venom or components of the venoms. One of the most important disadvantages of whole venom extract is the under- representation of allergens that are present in low abundance. Increasing knowledge about the venom composition and availability of component-resolved diagnostics (CRD) have vastly augmented our ability to solve many diagnostic issues. Measuring sIgE to recombinant venom allergens has improved the precision of allergic diagnostic tests.[5]

Sensitivity and Specificity of Skin Tests and sIgE

The specific sensitization profiles obtained due to availability of CRD can clearly increase the specificity of HVA diagnosis. For instance, positive detection of 6 of the major allergens of bee venom (Api m 1 to 5 and Api m 10) increases the specificity of bee allergy diagnosis to 94.4%, compared to 84.4% if only two allergens are detected.[5] Similarly, concomitant Ves v 1 and Ves v 5 sensitization identifies 92% to 98% of Vespula spp. allergic patients. A major diagnostic problem is that currently available tests are not able to distinguish between asymptomatic sensitization and clinically relevant allergy; up to 40% of an unselected population have positive test results, whereas only 0.3% to 8.9% have a history of allergic sting reactions. Negative sIgE and negative skin tests have been reported in up to 15% of patients with systemic mastocytosis and history of a systemic reaction to insect stings.[6] For skin testing by ant-allergic patients, whole-body extracts are currently available.[2]

Cross-Reactivity

There is strong cross-reactivity between venom allergens within a family, such as among *Vespula, Dolichovespula*, and *Vespa*.[5] Immunotherapy with *Vespula* spp. in patients with anaphylaxis to *Vespa orientalis* und *Vespa carbo* has proven to be efficacious.[4] Conversely, the discrimination between *Vespula* spp. and *Polistes* spp. sensitization is more challenging. Some studies suggest that in clinical practice, assessing serum levels of Ves v 5 and Pol d 5 may be helpful to discriminate between sensitizations.[6] New recombinant antigens (*e.g.*, rPol d 3) for *Vespula-Polistes* could increase the diagnostic accuracy.[5]

Skin and/or serologic tests can frequently show positivity to multiple venoms (*i.e.*, *Vespula* spp. and *Apis mellifera* in 25% to 40% of the cases, *Vespula* spp. and *Polistes* spp. in over 50% of cases). For an accurate HVA diagnosis and treatment, it is important to discriminate between cross-reactivity and true sensitization.[5] Cross-reactivity between different venoms can occur due to high homology in the structural composition of allergenic molecules or cross-reactive carbohydrates. Most of the cross-reactive recombinant pairs (rApi m 2 / rVes v 2, rApi m 5 / rVes v 3, and rApi m 12 / rVes v 6) are commercially unavailable, thus preventing misleading diagnostic steps. Assessing IgE antibodies to species-specific nonglycosylated, recombinant major allergens (e.g., Api m 1, m 10 and Ves v 1, v 5) reduces the rate of double-positivity and is helpful in choosing venoms for immunotherapy. The cross-reactive carbohydrates can be detected in most major Hymenoptera venom allergens, and in many plant proteins (e.g., rapeseed pollen or bromelain). They are probably responsible for positive test results in individuals with no history of SRs.

Cellular Tests[15]

The basophil activation test (BAT) is based on flow-cytometric demonstration of an altered membrane phenotype of basophils stimulated by IL-3 and allergen exposure. At present, the most commonly used expression marker is CD63. The BAT is expensive and time consuming, but it may be used as an additional test to confirm the diagnosis in case of negative or inconclusive results of conventional diagnostic tests. The BAT should be exclusively performed by an experienced laboratory using a validated assay.[15]

Baseline Serum Tryptase

Tryptase is a serine protease produced and released by mast cells after IgE-mediated or non-IgE-mediated stimuli. Because elevated baseline serum tryptase levels (>11.4 μg/L) are associated with severe, systemic sting reactions, this enzyme should be determined in all patients with a history of venom allergy.[16] The basal tryptase value varies very little over time within the same individual and is determined by the genetic background, not by environmental factors. This information is useful, as even minimal variations in tryptase concentration in a single individual can be indicative of the presence of an anaphylactic event, even if values fall within the normal range. Serum levels of baseline tryptase (above 25 μg/mL) suggest clonal mast cell disorders (*e.g.*, systemic mastocytosis). In these patients, further diagnostic tests should be considered (*e.g.*, skin inspections/biopsy, testing for c-kit mutations). Of note, patients with severe anaphylaxis and absence of urticaria or angioedema due to sting should be investigated for mastocytosis, even in the presence of normal baseline tryptase level. The characteristic skin lesions, called urticaria pigmentosa, may be noticeable during skin inspections.[6]

Sting Challenge Tests

Sting challenge with a live insect is not recommended as a diagnostic tool in untreated patients. However, a sting challenge under well-supervised clinical conditions may be helpful in evaluating the efficacy of venom immunotherapy and improvement in health-related quality of life. A tolerated sting challenge does not definitively exclude a reaction to future stings after immunotherapy, especially if these are repeated.[16]

PREVENTION

All patients with a history of SRs should receive detailed instructions on the avoidance of future stings. Bee stings occur most often when walking barefoot on grass, and wasp stings occur when eating outdoors, in orchards with fallen fruits, and near open waste bins. While gardening, wearing long trousers, shirts with long sleeves, and gloves are recommended. Strongly scented perfumes, sun creams, or shampoos, as well as brightly colored garments, should be avoided.

Treatment of Large Local Reactions

Oral antihistamines and cooling of the sting site (*e.g.*, with ice cubes) reduces local swelling, pain, and itching. Anti-inflammatory ointments or topical corticosteroids may diminish the local inflammatory process. In cases of severe swellings, oral corticosteroids together with antihistamines over several days can be considered. Prescription of adrenaline auto-injectors (AAI) in LLR is usually not recommended.[16]

Treatment of Systemic Allergic Reactions

All patients experiencing SRs should seek medical advice and be medically observed until the symptoms resolve and blood pressure is stable. Mild reactions confined to the skin may be treated with rapid-acting oral antihistamines alone. If severe respiratory or cardiovascular symptoms occur, epinephrine (intramuscular injection in the vastus lateralis muscle of the thigh) must be given immediately (0.1 mg epinephrine/10 kg body weight), intravenous (IV) access should be established, and fluids, antihistamines, and corticosteroids given intravenously.[16]

All patients with a history of systemic reaction upon stinging should be provided with AAIs (dosage 0.3 mg adrenaline in adults, 0.15 mg in children with body weight <25 kg) and double-dose oral anti-histamines with two tablets of prednisolone of 50 mg or equivalent.[17] Double AAIs can be prescribed for patients with history of severe reactions, patients living, working or performing outdoor activities in out-of-reach locations or distant from emergency rooms, and patients with clonal mast cell disorders and/or elevated levels of baseline serum tryptase. These patients must be advised to bring AAIs and other rescue medications (*i.e.*, antihistamines and corticosteroids) along with them at all times, especially in situations at high risk of stinging (*i.e.*, outdoor activities) or in out-of-reach locations.[17]

Venom Immunotherapy[18]

Indications

Hymenoptera venom immunotherapy (VIT) is the only effective treatment able to prevent the onset (or to reduce the degree of severity) of systemic reaction at re-sting in subjects with HVA. VIT is indicated in children and adults with a history of severe SRs (grade III/IV), provided sensitization to the relevant venom has been demonstrated by skin and/or blood test. VIT is also recommended for patients who experience repeated mild, non–life-threatening reactions and are at high risk for re-exposure, such as beekeepers and their family members. VIT can be considered in adult patients if the quality of life is impaired. Concomitant cardiovascular disease and mastocytosis are relative indications for VIT in patients with non–life-threatening sting reactions.[18] VIT should not be considered in the context of

unusual reactions that are not attributed to type I hypersensitivity reactions (see Table 47.3).

Contraindications for VIT are the same as for immunotherapy with other allergens.[19] Beta-blockers had been considered to increase the risk of a systemic reaction during VIT, particularly in patients with cardiovascular diseases. However, a position paper published by the European Academy of Allergology and Immunology (EAACI) suggests that beta-blockers are no longer to be considered as contraindicated in VIT. Although angiotensin conversion enzyme inhibitors (ACEI) were previously associated with systemic reactions during VIT, the EAACI has now recommended that ACEI therapy can be continued during VIT as long as the patient is informed about possible risks.[19] Any use of anti-hypertensive drugs has been suggested as a significant risk factor of systemic reactions to VIT. However, in a retrospective study that focused on patients with cardiovascular diseases, cardiovascular medication did not impair the safety and/or the efficacy of VIT. Patients taking anti-hypertensive drugs should be evaluated carefully before starting VIT, based on an individual risk/benefit assessment.[18]

Dosage and Treatment Regimens

Venom extracts are commercialized either as aqueous or "retard" (adsorbed to alum hydroxide or other substances) preparations. The use of adsorbed (depot or retard) formulations overall showed less occurrence of reactions when compared to aqueous extracts. During VIT, the venom extract is administered subcutaneously following standardized protocols. These protocols always include two phases: an up-dosing phase that incrementally reaches the final dose, and a maintenance phase. This final dose may be reached within a few weeks to months (in outpatient clinics), days, or hours (ultra-rush or cluster protocols including several injections a day) depending on the protocol.[18]

The recommended maintenance dose for both children and adults is 100 µg of the venom. This maintenance dose is equivalent to approximately two bee stings or several vespid stings. A higher dose (*e.g.*, 200 µg) is recommended when SRs occur after re-exposure to a field sting or a sting challenge and in highly exposed subjects, such as beekeepers or professional gardeners. During the first year, maintenance VIT is given every four weeks. Subsequently, the intervals may be extended to six weeks during the second year and eight weeks between years three and five if VIT is well tolerated. In the case of lifelong therapy, 12-week intervals may still be safe and effective.[18]

Adverse Reactions to Venom Immunotherapy

The overall incidence of systemic adverse reactions to VIT varies between 5% and 40%.[20] VIT with bee venom causes side effects more often than with *Vespula* venom. Ultra-rush protocols are associated with a higher rate of side effects compared with conventional protocols (see Fig. 47.3). Most systemic side effects are mild, but life-threatening anaphylaxis and fatal anaphylaxis have been described. Uncontrolled asthma, a prior history of SRs, administration of subcutaneous immunotherapy (SCIT) injections during peak pollen season, and delay in administration of epinephrine are recognized factors of fatal reaction to VIT.[20] Some units routinely give antihistamines two hours before up-dosing injections until the maintenance dose has been repeatedly well tolerated. Levocetirizine decreased the rate of SSR and fexofenadine decreased the rate of LLR and cutaneous SSR. Importantly, premedication with antihistamines do not negatively influence the effect of the therapy. In patients with recurrent SRs during up-dosing, off-label use of anti-IgE (omalizumab) has made it possible to complete up-dosing and to continue maintenance VIT.[21]

Efficacy of Venom Immunotherapy

In addition to three prospective controlled trials, the efficacy of VIT has been confirmed by well-tolerated sting challenges during VIT in several uncontrolled prospective studies. Full protection is achieved in 80% to 85% of patients receiving bee venom and in 95% to 100% of patients receiving *Vespula* venom.[18] The higher risk of treatment failure in bee venom allergy may relate to differences in venom compositions. Important allergens, such as Api m10 or Api m 3, may be missing or present in low concentrations in some VIT products. It has therefore been suggested that patients with a predominant sensitization to these allergens may not be properly protected by VIT. Moreover, mastocytosis is a significant determinant of VIT failure (see Fig. 47.3).

Duration of Venom Immunotherapy

Lifelong treatment may be the safest recommendation, but in most allergy centers, VIT is given for up to a maximum of 5 years. If VIT is given for at least 3 years, greater than 80% of both adults and children are still protected when reassessed 1 to 7 years after discontinuation. Longer courses of treatment should be considered in high-risk patients, such as those with very severe systemic sting reactions, coexisting cardiovascular or pulmonary disease, systemic allergic reactions to VIT or stings during VIT, and in subjects with elevated basal serum tryptase levels. Lifelong VIT is advised for patients with cutaneous or systemic mastocytosis.[18]

ON THE HORIZON

- Addition of recombinant species-specific nonglycosylated major allergens from honeybee allergens to therapy solutions can increase the efficacy of venom immunotherapy (VIT) with bee venom, as these are lacking or insufficiently represented in currently available preparation.
- Further development of component resolved diagnostics and improvement in understanding of differences among venom allergens of different species (*A. melifera*/Bumblebee, Vespula/Vespa/Dolichovespa/Polistes).
- There is a need to improve the monitoring of VIT efficacy (*in vitro* test/s or biomarker/s to confirm tolerance and to identify treatment failure after stopping the VIT).
- Further research on immunological differences between insect venom-allergic patients with and without immunotherapy and asymptomatically sensitized subjects are needed.
- There is need for studies on biomarkers able to distinguish between asymptomatic and symptomatic sensitization.
- New VIT formulations and routes of administration would be helpful to make the therapy less invasive, potentially increasing the adherence and quality of life of patients.
- Allergic side effects of VIT might be reduced by using modified major allergens with reduced immunoglobulin E (IgE) binding but retained specific T-cell interactions.

Risk Factors for Recurrence of Systemic Reactions after Stopping Venom Immunotherapy

A number of risk factors have been identified for relapse of Hymenoptera venom allergy after discontinuation of VIT. In general, adults, especially older people, have a less favorable prognosis compared with children because of concomitant diseases. Patients with bee venom allergy have a higher risk of relapse compared with those allergic to *Vespula* venom. Patients with more severe pretreatment reactions and those with more SRs during VIT (to injections or field stings) have a higher risk for recurrent SRs to hymenoptera stings (see Fig. 47.3). The risk of relapse is lower after 5 years of VIT, compared with only three years of VIT.[18]

REFERENCES

1. Müller UR. Insect sting allergy. In: Barry Kay A, Bousquet J, Holt PG, Kaplan PA, eds. *Allergy and allergic diseases.* 2nd ed. Wiley-Blackwell; 2009.
2. Hoffman DR. Ant venoms. *Curr Opin Allergy Clin Immunol*. 2010;10(4):342–346.
3. Fernández J. Distribution of vespid species in Europe. *Curr Opin Allergy Clin Immunol*. 2004;4(4):319–324.
4. Vidal C, Armisén M, Monsalve R, et al. Anaphylaxis to Vespa velutina nigrithorax: pattern of sensitization for an emerging problem in Western countries. *J Investig Allergol Clin Immunol*. 2021;31(3):228–235.
5. Bilò MB, Ollert M, Blank S. The role of component-resolved diagnosis in Hymenoptera venom allergy. *Curr Opin Allergy Clin Immunol*. 2019;19(6):614–622.
6. Bilò MB, Tontini C, Martini M, et al. Clinical aspects of Hymenoptera venom allergy and venom immunotherapy. *Eur Ann Allergy Clin Immunol*. 2019;51(6):244–257.
7. Lee S, Sadosty AT, Campbell RL. Update on biphasic anaphylaxis. *Curr Opin Allergy Clin Immunol*. 2016;16(4):346–351.
8. Mingomataj EÇ, Bakiri AH. Unusual reactions to hymenoptera stings: What should we keep in mind? *Clin Rev Allergy Immunol*. 2014;47(1):91–99.
9. Bilò MB, Bonifazi F. The natural history and epidemiology of insect venom allergy: clinical implications. *Clin Exp Allergy*. 2009;39(10):1467–1476.
10. Müller UR. Bee venom allergy in beekeepers and their family members. *Curr Opin Allergy Clin Immunol*. 2005;5(4):343–347.
11. Ruëff F, Przybilla B, Biló B, et al. Predictors of severe systemic anaphylactic reactions in patients with Hymenoptera venom allergy: importance of baseline serum tryptase—a study of the European Academy of Allergology and Clinical Immunology Interest Group on Insect Venom Hypersensitivity. *J Allergy Clin Immunol*. 2009;124(5):1047–1054.
12. Stoevesandt J, Sturm GJ, Bonadonna P, et al. Risk factors and indicators of severe systemic insect sting reactions. *Allergy*. 2020;75(3):535–545.
13. Bilò MB, Martini M, Pravettoni V, et al. Large local reactions to Hymenoptera stings: outcome of re-stings in real life. *Allergy*. 2019;74(10):1969–1976.
14. Biló BM, Rueff F, Mosbech H, et al. Diagnosis of Hymenoptera venom allergy. *Allergy*. 2005;60(11):1339–1349.
15. Scherer K, Bircher AJ, Heijnen IAFM. Diagnosis of stinging insect allergy: utility of cellular in-vitro tests. *Curr Opin Allergy Clin Immunol*. 2009;9(4):343–350.
16. Golden DBK, Demain J, Freeman T, et al. Stinging insect hypersensitivity A practice parameter update 2016. *Ann Allergy, Asthma Immunol [Internet]*. 2017;118(1):28–54. Available from: https://doi.org/10.1016/j.anai.2016.10.031.
17. Bilò MB, Cichocka-Jarosz E, Pumphrey R, et al. Self-medication of anaphylactic reactions due to Hymenoptera stings—an EAACI Task Force Consensus Statement. *Allergy Eur J Allergy Clin Immunol*. 2016;71(7):931–943.
18. Sturm GJ, Varga EM, Roberts G, et al. EAACI guidelines on allergen immunotherapy: Hymenoptera venom allergy. *Allergy Eur J Allergy Clin Immunol*. 2018;73(4):744–764.
19. Pitsios C, Demoly P, Bilò MB, et al. Clinical contraindications to allergen immunotherapy: an EAACI position paper. *Allergy*. 2015;70(8):897–909.
20. Mosbech H, Müller U. Side-effects of insect venom immunotherapy: results from an EAACI multicenter study. European Academy of Allergology and Clinical Immunology. *Allergy*. 2000;55(11):1005–1010.
21. Stretz E, Oppel EM, Räwer HC, et al. Overcoming severe adverse reactions to venom immunotherapy using anti-IgE antibodies in combination with a high maintenance dose. *Clin Exp Allergy*. 2017;47(12):1631–1639.

48

Atopic and Contact Dermatitis

Mark Boguniewicz, Luz Fonacier, and Donald Y.M. Leung

Atopic dermatitis (AD) and contact dermatitis (CD) are common inflammatory skin diseases.[1,2] The complex pathophysiology of AD involves both underlying skin barrier abnormalities and immune dysregulation. Its course is augmented by environmental influences, including stress, allergen exposure, and microbial infection. The systemic nature of AD has been increasingly recognized with type 2 immunity central across clinical phenotypes and endotypes and provides a rationale for treating patients with moderate-severe AD who do not respond to topical therapy with biologics targeting the key immune abnormalities.

CD is a skin disorder caused by contact with an exogenous substance that elicits an allergic and/or irritant response. Allergic contact dermatitis (ACD) accounts for 20%, whereas irritant contact dermatitis (ICD) accounts for 80% of all cases of CD. ACD is a delayed-type, T cell–mediated immune response consisting of an afferent limb and an efferent limb. Irritants cause direct activation of the innate immune system through hyperproduction of cytokines and chemokines. The management of ACD includes identification of the allergen, avoidance, pharmacological intervention, and prevention. This chapter reviews the clinical and mechanistic aspects of both AD and CD.

CLINICAL ASPECTS OF ATOPIC DERMATITIS

Epidemiology

Prevalence of AD has been documented in up to 24% of children, with no apparent difference seen between urban and rural districts or between the two sexes. Children from families with atopy have a significantly higher risk of developing AD. Data from an international study showed that the prevalence of current eczema in children varies widely among countries, ranging from 0.9% to 22.5%.[3] Prevalence in adults responding in the Atopic Dermatitis in America survey was found to be 7%.[4] These data point to AD as a global health problem in both developed and developing countries and suggest that the ultimate presentation of an atopic disease may depend on a complex interaction between environmental exposures and end-organ response in genetically predisposed individuals.

Natural History

AD typically manifests in early childhood, with onset in the first year occurring in more than 50% of patients and before 5 years of age in approximately 90% of patients. Although AD can present at any age, other diseases need to be considered especially in adults with new-onset dermatitis, without a history of childhood eczema, asthma, or allergic rhinitis (Table 48.1). Severity and atopic sensitization are major determinants of long-term prognosis. A registry of patients with mild-moderate AD showed that it was not until age 20 years that 50% had experienced at least one 6-month period free of symptoms and treatments.[5]

Clinical Features

AD has no pathognomonic skin lesions or unique laboratory parameters; the diagnosis is, therefore, based on the presence of major and associated clinical features (Table 48.2). The principal features include pruritus, a chronically relapsing course, typical morphology and distribution of skin lesions, and a history of atopic disease. Patients usually have dry (xerotic) skin, and those with mutations of the gene encoding the epidermal barrier protein filaggrin (*FLG*) typically have prominent scaling and hyperlinear palms. Acute AD is characterized by pruritic, erythematous papules associated with excoriations, vesiculations, and serous exudation. Subacute AD is characterized by erythematous, excoriated, scaling papules, whereas chronic AD is characterized by lichenification and fibrotic papules. During infancy, AD involves primarily the face, scalp, and the extensor aspects of the extremities; the diaper region is typically spared, although infants can have flexural involvement typical of older patients (Fig. 48.1).

Complicating Features

Ocular Problems

Atopic keratoconjunctivitis is bilateral, with intense pruritus, burning, tearing, and copious mucoid discharge. It is frequently associated with eyelid dermatitis and chronic blepharitis and may result in visual impairment from corneal scarring. Patients may also develop keratoconus from persistent rubbing of the eyes or anterior subcapsular cataracts.

Hand Dermatitis

Patients with AD may have a nonspecific irritant hand dermatitis aggravated by repeated wetting, especially in an occupational setting.

Infections

Patients with AD have an increased susceptibility to infection or colonization with a variety of organisms.[2] *Staphylococcus* (*S.*) *aureus* can be cultured from the skin of the majority of patients with AD. Preferential adherence of this organism in AD may

TABLE 48.1 Differential Diagnosis of Atopic Dermatitis

Congenital Disorders
- Netherton syndrome
- Familial keratosis pilaris

Chronic Dermatoses
- Seborrheic dermatitis
- Contact dermatitis (allergic or irritant)
- Nummular eczema
- Psoriasis
- Ichthyoses

Infections and Infestations
- Scabies
- Human immunodeficiency virus (HIV)–associated dermatitis
- Dermatophytosis

Malignancies
- Cutaneous T-cell lymphoma (mycosis fungoides/Sézary syndrome)
- Letterer-Siwe disease

Autoimmune Disorders
- Dermatitis herpetiformis
- Pemphigus foliaceus
- Graft-versus-host disease (GvHD)
- Dermatomyositis

Immunodeficiencies
- Wiskott-Aldrich syndrome (WAS)
- Severe combined immunodeficiency syndrome (SCID)
- Hyperimmunoglobulin E (IgE) syndrome (HIES)
- Dedicator of cytokinesis 8 (DOCK8)–associated immunodeficiency
- Tyrosine kinase 2 (TYK2) deficiency
- Immune dysregulation, polyendocrinopathy, enteropathy X-linked (IPEX) syndrome

Metabolic Disorders
- Zinc deficiency
- Pyridoxine (vitamin B6) and niacin deficiency
- Multiple carboxylase deficiency
- Phenylketonuria

TABLE 48.2 Clinical Features of Atopic Dermatitis

Major Features
- Pruritus
- Facial and extensor involvement in infants and children
- Flexural involvement at any age
- Chronic or relapsing dermatitis
- Personal or family history of atopic disease

Minor Features
- Xerosis
- Cutaneous infections
- Nonspecific dermatitis of the hands or feet
- Ichthyosis, palmar hyperlinearity, keratosis pilaris
- Pityriasis alba
- Nipple eczema
- White dermatographism and delayed blanch response
- Anterior subcapsular cataracts
- Elevated serum immunoglobulin E (IgE) levels
- Positive immediate-type allergy skin tests

Modified from Hannifin JM, Rajka G. Diagnostic features of atopic dermatitis. *Acta Derm Venereol (Stockh).* 1980;92:44–47.

be related to expression of adhesins, such as fibronectin and fibrinogen, in inflamed skin. Recurrent pustulosis has become a significant problem for a number of patients, especially with the emergence of methicillin-resistant *S. aureus* (MRSA) as an important pathogen in AD.

Viral infections in AD include herpes simplex, molluscum contagiosum, and human papillomavirus infections. Patients with eczema herpeticum (ADEH) have more severe disease, increased body surface area affected, and an increase in biomarkers, including circulating eosinophil counts, serum immunoglobulin E (IgE), thymus and activation-regulated chemokine, and cutaneous T cell–attracting chemokine, compared with patients with AD who do not have a history of EH. Patients with ADEH also have more cutaneous infections caused by S. aureus or molluscum contagiosum virus and are more likely to have a history of asthma, as well as food and inhalant allergies. Patients with ADEH have reduced interferon-γ (IFN-γ) production. IFNγ and IFNγR1 SNPs are significantly associated with ADEH and may contribute to an impaired immune response to herpes simplex virus. In addition, genetic variants in IFN-γ regulatory factor 2 (IFN-γR2) have been shown to be associated with ADEH and may contribute to abnormal immune responses to herpes simplex virus.

Superimposed dermatophytosis may contribute to cutaneous inflammation in AD. The opportunistic yeast *Malassezia sympodialis* (previously *Pityrosporum ovale*) has been associated with a predominantly head and neck distribution of AD.

Systemic Complications

A significant number of patients with AD do not outgrow their disease and can have systemic inflammation with increased risk for systemic diseases, although association with, for example, cardiovascular diseases, needs to be evaluated in prospective studies.

Psychosocial Implications

Patients with AD frequently respond to stress or frustration with itching and scratching. Stimulation of the central nervous system (CNS) may intensify cutaneous vasomotor and sweat responses and contribute to the itch–scratch cycle. Scratching can be associated with significant secondary gain or with a strong habitual component. Severe disease can lead to problems with self-esteem and social interactions. Sleep abnormalities are common and can contribute to impaired quality of life of patients and family members, even when the skin disease appears to be in remission.

Differential Diagnosis

A number of diseases can be confused with AD (see Table 48.1). Immunodeficiency with eczematous rash includes immune dysregulation, polyendocrinopathy, and enteropathy X-linked (IPEX) syndrome. IPEX results from mutations of *FOXP3*, a gene located on the X chromosome that encodes a DNA-binding protein required for development of regulatory T cells (Tregs). Patients may present with enteropathy, type 1 diabetes, thyroiditis, hemolytic anemia, and/or thrombocytopenia. Wiskott-Aldrich syndrome (WAS) is an X-linked recessive disorder characterized by an eczematous rash, associated with thrombocytopenia along with variable abnormalities in humoral and cellular immunity and severe bacterial infections. Hyper-IgE syndrome (HIES) with mutations in the signal transducer and activator of transcription 3 gene (*STAT3*) is an autosomal dominant multisystem

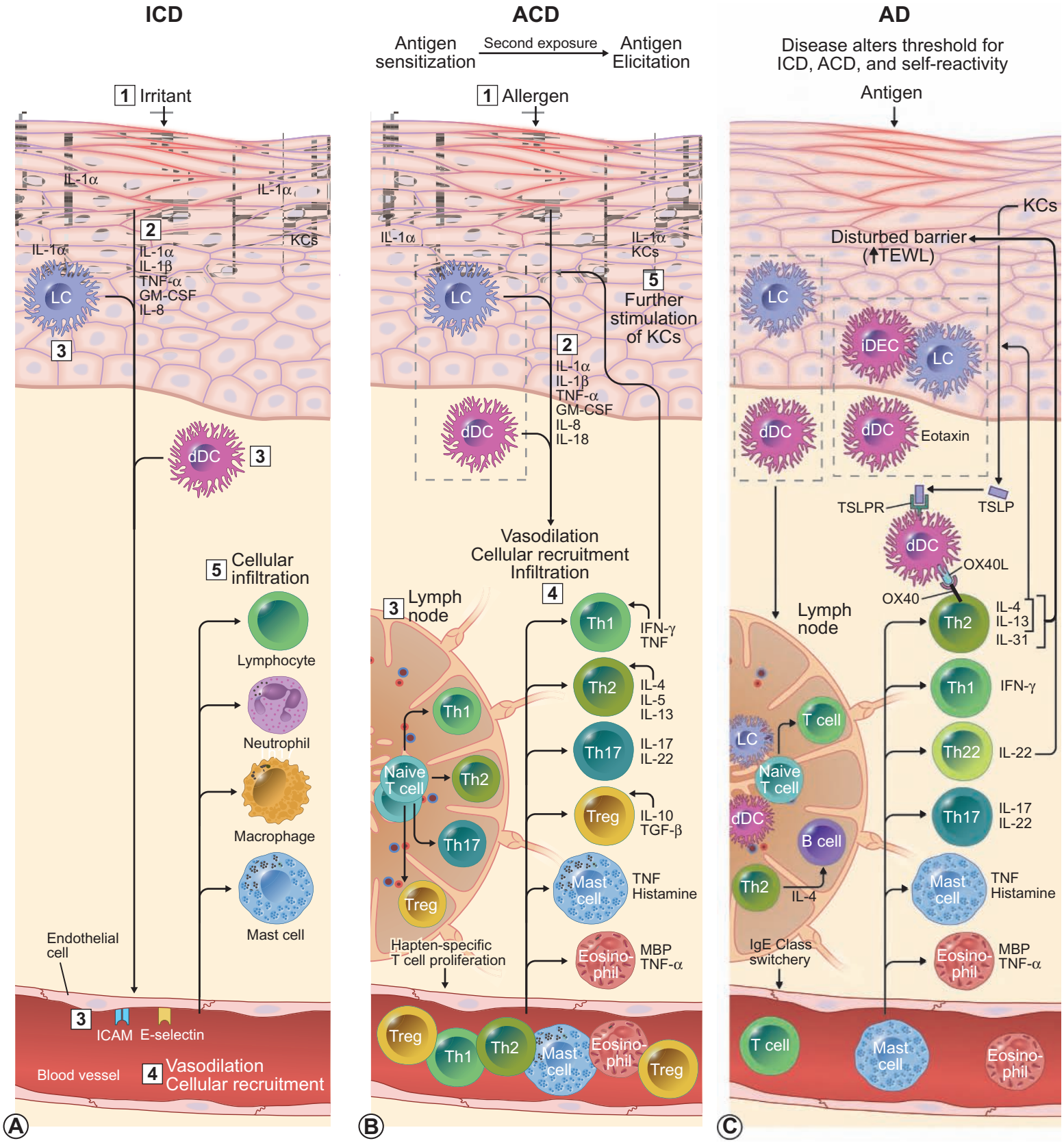

ICD
1 Irritant
IL-1α
IL-1α
IL-1α
2
IL-1α
IL-1β
TNF-α
GM-CSF
IL-8
KCs
LC
3
dDC
3
5 Cellular infiltration
Lymphocyte
Neutrophil
Macrophage
Mast cell
Endothelial cell
3
ICAM
E-selectin
Blood vessel
4 Vasodilation
Cellular recruitment
A
ACD
Antigen sensitization
Second exposure
Antigen Elicitation
1 Allergen
IL-1α
IL-1α
KCs
5 Further stimulation of KCs
LC
2
IL-1α
IL-1β
TNF-α
GM-CSF
IL-8
IL-18
dDC
Vasodilation
Cellular recruitment
Infiltration
3 Lymph node
4
Th1
IFN-γ
TNF
Th2
IL-4
IL-5
IL-13
Th17
IL-17
IL-22
Treg
IL-10
TGF-β
Naive T cell
Th1
Th2
Th17
Treg
Mast cell
TNF
Histamine
Eosino-phil
MBP
TNF-α
Hapten-specific T cell proliferation
Treg
Th1
Th2
Mast cell
Eosino-phil
Treg
B
AD
Disease alters threshold for ICD, ACD, and self-reactivity
Antigen
KCs
Disturbed barrier (↑TEWL)
LC
iDEC
LC
dDC
dDC
Eotaxin
TSLPR
TSLP
dDC
OX40L
OX40
Lymph node
Th2
IL-4
IL-13
IL-31
Th1
IFN-γ
Th22
IL-22
Th17
IL-17
IL-22
T cell
LC
Naive T cell
dDC
B cell
Th2
IL-4
Mast cell
TNF
Histamine
IgE Class switchery
Eosino-phil
MBP
TNF-α
T cell
Mast cell
Eosino-phil
C

FIG. 48.1 (A) In patients with irritant contact dermatitis (ICD), exposure to an irritant exerts toxic effects on keratinocytes (KCs), activating innate immunity with release of interleukin (IL)-1α, IL-1β, TNF, granulocyte macrophage–colony-stimulating factor (GM-CSF), and IL-8 from epidermal keratinocytes. In turn, these cytokines activate Langerhans cells (LCs), dermal dendritic cells (DCs), and endothelial cells, all of which contribute to cellular recruitment to the site of keratinocyte damage. Infiltrating cells include neutrophils, lymphocytes, macrophages, and mast cells, which further promote an inflammatory cascade. (B) In the sensitization phase of allergic contact dermatitis (ACD), similar to ICD, allergens activate innate immunity through keratinocyte release of IL-1α, IL-1β, TNF, GM-CSF, IL-8, and IL-18, inducing vasodilation, cellular recruitment, and infiltration. LCs and dermal DCs encounter the allergen and migrate to the draining lymph nodes, where they activate hapten-specific T cells, which include T-helper cell-1 (Th1), Th2, Th17, and regulatory T cells (Tregs). These T cells proliferate and enter the circulation and site of initial exposure, along with mast cells and eosinophils (EOS). On re-encountering the allergen, the elicitation phase occurs, in which the hapten-specific T cells, along with other inflammatory cells, enter the site of exposure and, through release of cytokines and consequent stimulation of keratinocytes, induce an inflammatory cascade. (C) In patients with atopic dermatitis (AD), a disturbed epidermal barrier leads to increased permeation of antigens, which encounter LCs, inflammatory dendritic epidermal cells (IDECs), and dermal DCs, activating Th2 T cells to produce IL-4 and IL-13. DCs then travel to lymph nodes, where they activate effector T cells and induce immunoglobulin E (IgE) class-switching. IL-4 and IL-13 stimulate keratinocytes to produce thymic stromal lymphopoietin (TSLP). TSLP activates OX40 ligand–expressing dermal DCs to induce inflammatory Th2 T cells. Cytokines and chemokines, such as IL-4, IL-5, IL-13, eotaxin, CCL17, CCL18, and CCL22, produced by Th2 T cells and DCs stimulate skin infiltration by DCs, mast cells, and eosinophils. Th2 and Th22 T cells predominate in patients with AD, but Th1 and Th17 T cells also contribute to its pathogenesis. The Th2 and Th22 cytokines (IL-4/IL-13 and IL-22, respectively) have been shown to inhibit terminal differentiation and contribute to the barrier defect in patients with AD. Thus both the barrier defects and immune activation alter the threshold for ICD, ACD, and self-reactivity in patients with AD. MBP, Major basic protein. (Figure adapted from Gittler JK, Krueger JG, Guttman-Yassky E. Atopic dermatitis results in intrinsic and immune abnormalities: implications for contact dermatitis. *J Allergy Clin Immunol*. 2013;131:300–313.)

disorder characterized by recurrent deep-seated bacterial infections, including cutaneous cold abscesses and pneumonias caused by *S. aureus*. Patients with mutations in the gene encoding dedicator of cytokinesis 8 protein (*DOCK8*) have an immunodeficiency with eczema, recurrent viral infections, including some with CNS involvement, and many have associated food allergies. Patients with tyrosine kinase 2 (TYK2) deficiency can also present with an eczematous rash with high serum IgE and recurrent cutaneous staphylococcal infections. Other diseases to consider in the differential diagnosis of AD include cutaneous T-cell lymphoma, especially in adults with no history of childhood eczema and without other atopic features, human immunodeficiency virus (HIV) infection, and CD (see section on ACD). Contact allergy can also complicate AD, especially in patients whose AD does not respond to or worsens with therapy.

PATHOGENESIS OF ATOPIC DERMATITIS

Genetics

The genetics of AD are complex, with key roles played by skin barrier or epidermal differentiation genes and immune response or host defense genes.

KEY CONCEPTS

Pathogenesis of Atopic Dermatitis

- Patients with AD have abnormalities of skin barrier and immune dysregulation.
- Patients with AD can have decreased filaggrin in the skin based on mutations in the gene encoding filaggrin (*FLG*), copy number variation of *FLG*, or secondary to type 2–mediated cytokine (e.g., IL-4, IL-13) suppression of *FLG*.
- *FLG* mutations are associated with early-onset, severe, persistent AD with increased risk for asthma and allergic sensitization.
- Normal-appearing skin in patients with AD is associated with immune activation and epidermal terminal differentiation abnormalities.
- Patients with AD are predisposed to *Staphylococcus aureus* colonization or infection, and a subset are prone to eczema herpeticum.

Loss-of-function mutations of *FLG* are a major predisposing factor for AD. Patients with *FLG* gene mutations have early-onset, severe, and persistent AD and are at increased risk for developing asthma, as well as food and inhalant allergies.[6] Other skin barrier proteins implicated in AD include loricrin and involucrin, which are both significantly decreased in lesional and nonlesional skin of patients with AD. Candidate gene approaches have also implicated variants in the *SPINK5* gene, and haplotype tagging of SNPs for the claudin-1 gene (*CLDN1*) has revealed associations with AD. These observations have established a key role for impaired skin barrier function in the pathogenesis of AD that allows increased transepidermal water loss and penetration of allergens, antigens, and chemicals from the environment, resulting in cutaneous inflammation. Additional genome-wide significant susceptibility loci identified through the genome-wide association study (GWAS) and immunochip analysis continue to be reported and suggest a role for epidermal barrier function, innate-adaptive immunity, interleukin-1 (IL-1) family signaling, Tregs, the vitamin D pathway, and the nerve growth factor pathway in the pathogenesis of AD. Thus, a combination of varied genetic factors may influence a wide range of phenotypes of AD among individuals.

Immune Abnormalities in Atopic Dermatitis

Immunohistology

On the basis of clinical appearance and duration of illness, AD skin can be characterized as nonlesional AD, acute AD lesions (3 days or fewer after onset), and chronic skin lesions. Nonlesional AD skin is not normal and is characterized by a sparse perivascular T-cell infiltrate, suggesting the presence of minimal inflammation at baseline. Langerhans cells (LCs) also exhibit surface-bound IgE molecules, which have enhanced capacity to present allergen to T cells. In acute lesions, there is epidermal spongiosis with an increased infiltration of activated memory T cells bearing the skin-homing cutaneous lymphocyte-associated antigen (CLA^+). Eosinophils, basophils, and neutrophils are rare. Mast cells are seen in various stages of degranulation.

Chronic lichenified AD lesions are characterized by epidermal hyperplasia, prominent hyperkeratosis, and an increased number of dendritic cells (DCs) bearing surface IgE. Macrophages dominate the dermal mononuclear cell infiltrate, but lymphocytes remain prominent. Although intact eosinophils are rarely seen, eosinophil product deposition can be readily identified, suggesting they contribute to allergic skin inflammation.

Immune Pathways in Atopic Dermatitis

AD is associated with a combination of defective innate responses to microbes and altered adaptive responses to environmental allergens (reviewed in references).[1,2,7] A key difference between epidermal keratinocytes found in AD skin, compared with normal skin, is the presence of thymic stromal lymphopoietin (TSLP) and IL-33 in the AD epidermis. TSLP and IL-33 are key cytokines secreted by epithelial cells that induce DCs to drive T-helper 0 (Th0) cells into the Th2-cell differentiation pathway. Nonlesional AD and acute AD skin lesions are predominantly associated with expression of IL-4, IL-5, IL-13, IL-25, and IL-33. These type 2 cytokines are present in all stages of AD and can be secreted by multiple cell types, including innate lymphoid type 2 cells (ILCs), mast cells, and basophils, which are present in AD skin lesions and contribute to substantial redundancy in allergic inflammation. As such, cytokine targeting, as opposed to cell targeting, is considered a more effective approach in the treatment of AD. It is noteworthy that experimental studies demonstrate pretreatment with IL-4, and IL-13 dampens responses to IFNs and IL-17, suggesting that once the early AD lesion is exposed to IL-4 and IL-13, there is a long-lasting persistent effect.

Besides Th2, other cytokine pathways are also activated during the evolution of AD. The IL-22–IL-17 pathway is of particular interest, since it, along with IL-4 and IL-13, can inhibit terminal keratinocyte differentiation, including filaggrin expression. Since DC-derived IL-23 enhances IL-22/IL-17 cell differentiation, all of these cytokines are being closely examined for their potential role in AD. It is interesting that IL-4 and Il-13 can enhance IL-23 production by DCs. Furthermore, blockade of the IL-4 and IL-13 pathway, leading to improvement in AD, is also associated with reduced IL-23 and IL-17 expression in AD skin. When acute AD skin lesions become chronic, there is an increase in Th1 cytokines, such as IFN-γ, which potentiates AD skin inflammation. TSLP can be detected in at-risk infants before onset of AD, suggesting that the TSLP–Th2–ILC2 pathway plays a critical role in initiation of AD.

Epidermal Barrier Dysfunction

The clinically dry skin and increased transepidermal water loss (TEWL) in AD reflect underlying skin barrier dysfunction and loss of natural moisturizing factors that play an important role in the pathogenesis of AD. However, only a minority of patients have *FLG* null mutations. Other genetic variants in the epidermal differentiation complex (EDC) and tight junctions are even rarer. The majority of patients with AD likely have immune-mediated reduction in epidermal terminal differentiation, which leads to reduced generation of various epidermal structural proteins, filaggrin breakdown products, lipids, and antimicrobial peptides. TSLP, IL-4, and IL-13 are the most potent cytokines downregulating filaggrin expression by keratinocytes.[6] IL-17, IL-22, IL-25, and IL-33 can act synergistically with IL-4 and IL-13 to further downregulate expression of epidermal proteins and lipids. This combination of events, along with activation of proteases and lipases, creates defective epidermal barrier function, alters epidermal acidification, and leads to loss of moisturization in AD, thereby contributing to enhanced allergen and microbial penetration met by the host immune response and the clinical appearance of AD.

MANAGEMENT OF ATOPIC DERMATITIS

Management of AD is addressed in both an AD practice parameter update[8] and AD guidelines[9] with a more recent AD yardstick including newer therapeutic options.[10] Recognition of skin barrier and immune abnormalities suggests the need for barrier repair and maintenance along with antiinflammatory measures.

THERAPEUTIC PRINCIPLES

Atopic Dermatitis

- Fundamentals of skin care in AD include avoidance of irritants and proven allergens, along with proper skin hydration and use of moisturizers.
- Patients and caregivers have concerns about topical steroids and calcineurin inhibitors and tend to underuse prescribed medications.
- If eczema cannot be cleared and the patients keep relapsing, they may benefit from proactive intermittent antiinflammatory therapy.
- The diagnosis of AD should be reconsidered in patients not responding to conventional therapy, as the differential diagnosis includes a number of immunodeficiencies with eczematous rash.
- Systemic therapies should be reserved for patients with moderate-to-severe AD that does not respond to conventional therapy.

Identification and Elimination of Exacerbating Factors

Irritants

Patients with AD have a lowered threshold of irritant responsiveness. Recognition and avoidance of irritants is integral for successful management of this disease.

Allergens

Identification of clinically relevant allergens involves taking a careful history and doing selective testing, when appropriate. Negative skin tests with proper controls have a high predictive value for ruling out a suspected allergen. Positive skin tests have a lower correlation with clinical symptoms and may reflect sensitization.[8] Patients with severe AD who are avoiding foods based on *in vitro* testing results can tolerate many of the foods previously avoided when they are introduced under supervision. Environmental control measures aimed at reducing dust mite allergen can also improve AD in sensitized patients.

Psychosocial Factors

Recognizing and addressing sleep disturbance in patients is critical in managing a chronic, relapsing disease such as AD. Counseling together with relaxation, behavioral modification, and biofeedback may all be of benefit, especially for patients with habitual scratching.

Patient Education

Patients and caregivers need to be educated regarding the chronic relapsing nature of AD, its natural history, exacerbating

factors, and treatment options. Recognizing that normal-appearing skin in patients with AD is, in fact, *not* normal is a difficult concept to understand but has important therapeutic implications. Patients should be counseled about prognosis and receive appropriate vocational counseling.

Hydration

The skin of patients with AD shows enhanced TEWL and lipid abnormalities that result in reduced water-binding capacity, higher TEWL, and decreased water content.[7] Skin hydration by soaking the affected area or bathing and applying an occlusive agent to retain absorbed water can help restore barrier function.[8] Bathing can also remove allergens, reduce *S. aureus* colonization, and act as relaxation therapy.

Moisturizers and Occlusives

Use of moisturizer or occlusive, especially when combined with hydration therapy, helps restore and preserve the skin barrier and can result in decreased need for topical corticosteroids. Twice-daily emollient application can improve barrier function and protect the skin from *S. aureus* proliferation while preserving microflora biodiversity. Patients with AD have been shown to have a ceramide deficiency of the stratum corneum; barrier repair may be accelerated by increasing the ratio of ceramides, cholesterol, and either the essential fatty acid linoleic acid or the nonessential palmitic or stearic fatty acids.

Corticosteroids

Corticosteroids reduce inflammation and pruritus in acute and chronic AD, acting on multiple resident and infiltrating cells, primarily through suppression of inflammatory genes. Topical corticosteroids have been the mainstay of conventional therapy: when they are appropriately used, side effects are infrequent. Thinning of skin, telangiectasia, bruising, hypopigmentation, acne, striae, and secondary infections may occur. The face, particularly the eyelids, and the intertriginous areas are especially sensitive to these adverse effects. If topical corticosteroids are used on the face, this can lead to perioral dermatitis, characterized by erythema, scaling, and follicular papules and pustules around the mouth, in the alar creases, and sometimes on the upper lateral eyelids.

An important concept with translational applications is the recognition that nonlesional skin in AD shows evidence of both immunological dysregulation and skin barrier abnormalities.[11] This observation provides a rationale for the use of topical corticosteroids in a proactive manner.[8] If the eczema can be cleared or almost cleared but has a relapsing course, long-term control can be maintained with twice-weekly applications of topical corticosteroid to areas that had previously been involved but now appear normal, with fewer relapses and less need for topical corticosteroids compared with treating eczema in a reactive manner.

In addition to their anti-inflammatory properties, topical corticosteroids can decrease *S. aureus* colonization in patients with AD. Failure to show clinical improvement with topical corticosteroids may be due to inadequate potency or amount of medication used, superinfection, steroid allergy, steroid resistance, or, more commonly, nonadherence with the treatment regimen, emphasizing the need for both education and alternative therapies.

Systemic corticosteroids, including oral prednisone, should be avoided in the management of a chronic disorder such as AD.[8,12] Improvement observed with systemic corticosteroids may be associated with flaring of AD after discontinuation. If a short course of oral corticosteroids is given, topical skin care should be intensified during the taper to suppress rebound flaring of AD.

Topical corticosteroid treatment in patients with AD can result in improvements of the AD genomic signature. Cytokine levels (IL-12p40, IL-13, IL-22, CCL17, CCL18, peptidase inhibitor 3 [PI3]/elafin, and S100As) were consistently reduced, with corresponding improvements in epidermal disease markers (keratin 16 and loricrin) in lesional skin from responders. Even low-potency corticosteroids can affect a broad array of immune and barrier responses in patients with AD.

Topical Calcineurin Inhibitors

Tacrolimus ointment (0.03% and 0.1%) and pimecrolimus cream (1%) are nonsteroidal topical calcineurin inhibitors (TCIs) approved for the treatment of AD.[8–10] Both drugs have proven effective, with a good safety profile even when used over extended periods, including in infants treated with pimecrolimus. Treatment with TCIs is not associated with skin atrophy and may also be useful in treating patients with steroid insensitivity. A common side effect with TCIs is burning or stinging sensation of skin. Ongoing surveillance and recent reports have not shown a trend of increased frequency of viral infections or problems with response to childhood vaccinations. Although there is no evidence of a causal link to cancer and the use of TCIs, the US Food and Drug Administration (FDA) issued a "box warning" because of lack of long-term safety data. The labeling states that these drugs are recommended as second-line treatments and that their use in children under the age of 2 years is currently not recommended. However, a review of epidemiological and clinical data concluded that the published data did not demonstrate any causal relationship between TCI use and malignancy or lymphoma risk. Proactive treatment with tacrolimus ointment may benefit patients with a relapsing course.[8,10]

Phosphodiesterase-4 Inhibitor

Crisaborole is a phosphodiesterase 4 (PDE4) inhibitor approved as a 2% topical ointment for adults and children 3 months and older with mild to moderate AD.[10] The unique configuration of boron within the crisaborole molecule enables selective targeting and inhibition of PDE4, increasing cAMP levels and reducing inflammation. In addition, the boron molecule allowed for synthesis of a low-molecular-weight compound (251 daltons), thereby facilitating penetration of crisaborole through human skin.[13] *In vitro* crisaborole inhibits cytokine production by peripheral blood mononuclear cells similar to other PDE4 inhibitors and distinct from corticosteroids.

Antiinfective Therapy

Systemic antibiotic therapy may be necessary when a secondary infection with *S. aureus* is present.[2] Recolonization after a course of anti-staphylococcal therapy occurs rapidly. Maintenance antibiotic therapy should be avoided because it can result in colonization by MRSA. Bleach baths with dilute sodium hypochlorite may reduce skin infections, although data showing clinical benefit versus plain water baths for eczema have been questioned.[14]

Patients with disseminated EH require treatment with a systemic antiviral, such as acyclovir.[2] Recurrent cutaneous herpetic

infections can be treated with prophylactic oral acyclovir. Superficial dermatophytosis can be treated with topical or, rarely, with systemic antifungals.

Antipruritic Agents

Pruritus is the most common and least tolerated symptom in AD. Mediators other than histamine, including neuropeptides and cytokines, especially IL-31, can contribute to pruritus, and centrally acting agents, such as opioid receptor antagonists, have been shown to be effective against the itch of AD. Treatment with cyclosporine, which results in decreased transcription of a number of proinflammatory cytokines, leads to rapid improvement in pruritus in many AD patients. Monoclonal antibodies (mAbs) against IL-31 receptor are currently in phase 3 clinical trials in patients with AD, and trials with dupilumab (discussed below) showed significant reduction in pruritus.[15] Systemic antihistamines and anxiolytics may be most useful through sedative effects and should be used primarily at bedtime to avoid daytime somnolence. Short-term use of a sedative to allow adequate rest may be appropriate. Treatment with topical antihistamines and topical anesthetics should be avoided because of the potential for allergic sensitization.

Biologic therapy

Dupilumab is a fully human mAb directed against the IL-4 receptor α subunit that blocks signaling of both IL-4 and IL-13. Results from randomized, double-blind, placebo-controlled trials involving adults with moderate-to-severe AD showed that treatment with dupilumab resulted in rapid and dose-dependent improvements in clinical indexes, biomarker levels, and disease transcriptome.[15] Of the patients in the dupilumab group, 85% had a 50% reduction in the Eczema Area and Severity Index (EASI) score compared with 35% in the placebo group; 40% of patients in the dupilumab group had an investigator's global assessment score of 0 to 1 (clear or almost clear) compared with 7% in the placebo group; and pruritus scores decreased by 55.7% in the dupilumab group versus 15.1% in the placebo group. In a 52 week study of dupilumab with concomitant topical steroids, 100% of the patients in the dupilumab group had a 50% improvement in EASI, compared with 50% of those in the placebo group, even though the patients who received dupilumab used less than half the amount of topical steroids compared with those in the placebo group.[15] Adverse events, such as skin infection, occurred more frequently with placebo; nasopharyngitis and headache were the most frequent adverse events with dupilumab. In addition, conjunctivitis was reported in approximately 10% of patients in the phase 3 trials, although this adverse event did not require discontinuation in the majority of patients. Of note, treatment with dupilumab can correct both cutaneous and systemic abnormalities.[16] Dupilumab was well tolerated with no dose-limiting toxicity. Studies in adolescents confirmed both safety and efficacy.[17] Dupilumab is indicated for the treatment of patients 6 years or older with moderate-to-severe AD whose disease is not adequately controlled with topical prescription therapies or when those therapies are not advisable. Dupilumab can be used with or without topical corticosteroids. Dosing in adults is 600 mg loading dose subcutaneously followed by 300 mg every other week. Dosing for patients 6-17 years is weight dependent with patients 60kg or more receiving adult dosing, patients 30kg to less than 60kg receiving 400mg loading dose followed by 200mg every other week, and patients 15kg to less than 30kg receiving 600mg loading dose followed by 300mg every 4 weeks. Currently, no laboratory monitoring is required.

RECALCITRANT DISEASE

Patients failing conventional therapy should be evaluated to confirm the diagnosis of AD and review treatment regimen as well as triggers and confounding problems. An approach to such patients with an annotated figure was the subject of a recent review.[18]

Hospitalization

Patients who experience erythroderma, appear toxic, or fail conventional therapies may require hospitalization. Removing the patient from environmental allergens or stressors, intense education, and assurance of adherence with therapy usually result in marked clinical improvement. The patient can also undergo appropriately controlled provocative challenges to identify potential triggering factors.

Wet Wrap Therapy

Wet wrap therapy (WWT), in which a layer of wet clothing or bandages is placed over topical corticosteroid with a dry layer on top, can reduce inflammation and pruritus and acts as a barrier to trauma associated with scratching.[18] WWT can aid epidermal barrier recovery that persists even after this therapy is discontinued, and clinical benefit can be demonstrated after discontinuation. Overuse can result in chilling, maceration of skin, or secondary infection. WWT should be employed for acute exacerbations of AD or for selective use in areas of resistant eczema, rather than as a maintenance treatment. The package inserts for topical calcineurin inhibitors recommend that they should not be used under occlusive dressings.

Systemic Immunosuppressive Agents

Oral cyclosporine has been shown to be effective in placebo-controlled studies in severe AD, although it is not approved for treatment of AD in most countries. A systematic review of randomized controlled trials evaluating systemic immunomodulating treatments for moderate-to-severe AD found that, of 12 different interventions studied in 34 randomized controlled trials, strong recommendations were only possible for the short-term use of cyclosporine.[19] Patients need to be appropriately monitored while on this medication.[12] Other systemic immunosuppressive drugs that have been used off-label to treat severe AD include methotrexate (a folic acid antagonist that interferes with purine and pyrimidine synthesis), and less commonly mycophenolate mofetil (a purine biosynthesis inhibitor) and azathioprine (a systemic immunosuppressive agent affecting purine nucleotide synthesis and metabolism).[12,18] All of these drugs require careful monitoring of patients for potential serious adverse effects.

Phototherapy and Photochemotherapy

Ultraviolet (UV) light therapy can be a useful treatment for chronic, recalcitrant AD.[10,12] In patients with moderate-to-severe chronic AD treated with narrow-band UVB phototherapy, gene expression, and immunohistochemistry studies of both lesional and nonlesional skin showed that Th2, Th22, and Th1 pathways were suppressed and that measures of epidermal hyperplasia and differentiation were normalized. Clinical improvement was associated with decrease in Th2/Th22–associated cytokines and chemokines and, importantly, normalized expression of epidermal

barrier proteins. UVB phototherapy has also been shown to significantly decrease levels of toxin-producing *S. aureus* on the skin of pediatric patients with AD. Potential long-term adverse effects include premature skin aging and cutaneous malignancies.

Allergen-Specific Immunotherapy

Allergen-specific immunotherapy (AIT; see Chapter 88) may be effective for a subset of patients with AD and aeroallergen sensitivity.[8] However, a Cochrane review of randomized controlled trials of AIT in AD found results confounded by high loss to follow-up and lack of blinding.[20] Subgroup analyses did not identify if a particular allergen, age, or level of disease severity predicted treatment success.

Investigational Therapies

Intravenous Gammaglobulin

Because chronic inflammation and T-cell activation appear to play a critical role in the pathogenesis of AD, high-dose intravenous immunoglobulin (IVIG; see Chapter 82) could have immunomodulatory effects in this disease. IVIG could also interact directly with infectious organisms or toxins involved in the pathogenesis of AD. IVIG has been shown to contain high concentrations of staphylococcal toxin–specific antibodies that inhibit the *in vitro* activation of T cells by staphylococcal toxins. Treatment of severe refractory AD with IVIG has yielded conflicting results. Studies have not been controlled and have involved small numbers of patients. A systematic review of systemic therapies found that IVIG was not efficacious in the treatment of moderate-to-severe AD.[19]

Omalizumab

Treatment of patients with AD with omalizumab off label has mainly been reported in case reports and case series, showing both clinical improvement and lack of benefit.[10] No specific markers have been identified that define responders, although one study suggested that adult patients with AD that respond to treatment have wild-type *FLG* mutations. In addition, patients receiving omalizumab have been shown to have decreased levels of TSLP, OX40L, thymus and activation-regulated chemokine (TARC), and IL-9 and marked increase in IL-10 compared with placebo. A systematic review and meta-analysis of omalizumab in AD found that fewer than 50% of the patients treated with this biologic achieved a significant clinical improvement.[21] In the two randomized controlled trials in that review, patients failed to show any significant clinical improvement with omalizumab, or their clinical response was comparable to that of the control group. However, the authors noted that 43% of patients treated with omalizumab had a good response, suggesting that a subset of patients with AD, possibly those with an urticarial component to their disease might still benefit from this therapy. Furthermore, a recent randomized clinical trial in children with severe AD found that omalizumab significantly reduced disease severity and topical steroid use.[22]

Recombinant Human Interferon-γ

Interferon-γ suppresses IgE synthesis and inhibits Th2 cell function. In patients with AD, treatment with subcutaneous recombinant human interferon-γ (rhIFN-γ) resulted in reduced clinical severity and decreased total circulating eosinophil counts.[8,18] rhIFN-γ may act primarily on the allergic inflammatory response as opposed to IgE synthesis, and a subset of patients treated with rhIFN-γ could respond to individualized titration of their treatment dose. In pediatric patients with ADEH, IFN-γ and IVIG were thought to be less likely to enhance the cutaneous viral susceptibility of these patients compared with systemic immunosuppressive therapies.

Probiotics

Clinical trials of probiotics in patients with AD show varying results. Probiotic supplementation during the prenatal and the postnatal period may reduce the incidence of AD in infants and children. In contrast, the most recent Cochrane review found no benefit for probiotics as a treatment for eczema.[23]

Other Biologics and Small Molecules in Clinical Trials

A number of biologics are currently in trials for moderate-to-severe AD (reviewed in reference 24). These include anti–IL-13 (lebrikizumab, tralokinumab), anti–IL-31-receptor (nemolizumab), and anti-OX40 (GBR 830). In addition, several Janus kinase inhibitors in both topical (ruxolitinib, delgocitinib) and oral (abrocitinib, baricitinib, upadacitinib) formulations are being studied in AD.[24] Ruxolitinib cream is approved by the FDA for patients 12 years and older with mild-moderate AD. New insights into AD phenotypes and endotypes will likely lead to improved selection of appropriate patients for more targeted therapies.

PREVENTION

Because infants at risk for AD could be identified as early as day 2 of life by TEWL measurement, independent of parental atopy or *FLG* mutation status, early treatment with emollients has been suggested as a strategy to reduce incidence of AD. While preliminary studies suggested that emollient therapy started in early infancy in at-risk neonates reduced the incidence of AD, recent large trials failed to reproduce these results.[25,26]

CONTACT DERMATITIS

CD is a skin disorder caused by contact with an exogenous substance that elicits an allergic and/or irritant response. The vast majority of cases of contact-induced skin reactions are attributable to CD. However, there are other less-well-defined contact reactions, including contact urticaria, contact urticaria syndrome, and protein contact dermatitis. ACD affects approximately 7% of the general population, 13.3% to 24.5% of pediatric patients,[27] and up to 65% among children with suspected ACD. Numerous studies have reported an increased frequency of ACD in patients with AD, likely as a result of the increased exposure to products and chemicals used to treat AD, the barrier defect, and immunological changes in AD, which predispose a patient to both ICD and ACD.[28]

PATHOGENESIS OF ALLERGIC CONTACT DERMATITIS

The Genes

ACD is a multifactorial condition in which genetic background plays an important part, as shown by twin and family studies.

An association has been demonstrated between loss of function mutations R501x and 2282del4 in the filaggrin (*FLG*) gene and contact sensitization against nickel II-sulfate, combined with an intolerance to fashion jewelry but not with other contact allergens. *FLG* mutations have further been shown to lower the age of onset of nickel sensitization.[29] Thus FLG defects may be a risk factor for contact sensitization to allergens.

The Allergens

Most contact allergens are *haptens*, that is, simple chemicals that bind to carrier proteins present in skin to form a complete antigen (see Chapter 6). To be allergenic, the chemical must be able to penetrate the principal barrier in skin (stratum corneum) and reach the living cells of the epidermis. Only molecules with molecular mass of less than 500 daltons (Da) are capable of penetrating the stratum corneum. Lipid solubility promotes transit through the stratum corneum. Thus, most contact allergens are small, lipophilic molecules. Once in the epidermis, the nature of the protein carrier for the hapten is very important because if the contact sensitizer is complexed to nonimmunogenic carriers, this may induce tolerance rather than sensitization.

The Immune Response

KEY CONCEPTS

Pathogenesis of Contact Dermatitis

- Allergic contact dermatitis (ACD) is a delayed-type, T cell–mediated response with an afferent limb or sensitization phase and an efferent or elicitation phase.
- Irritant contact dermatitis (ICD) is caused by irritants exerting toxic effects on keratinocytes, causing a direct activation of the innate immune system through hyperproduction of cytokines and chemokines and inducing an inflammatory skin reaction.
- Patients with ICD are more susceptible to the development of ACD, indicating that the activation of innate immunity by irritants likely reduces the threshold for development of ACD.
- There is an increased frequency of ACD in patients with AD likely because of the disturbed skin barrier allowing increased allergen penetration on an already amplified adaptive response in AD.
- Although ACD, ICD, and AD are independently defined diseases, they frequently interact and coexist.

The immune response of ACD requires completion of both an afferent (sensitization) and an efferent (elicitation) phase. In the afferent limb, the hapten enters the epidermis and activates keratinocytes to release inflammatory cytokines and chemokines, including tumor necrosis factor (TNF), granulocyte macrophage–colony-stimulating factor (GM-CSF), IL-1α, IL-1β, IL-8, IL-10, IL18, and macrophage inflammatory protein-2 (MIP-2). The activation of LCs, other DCs, and endothelial cells can lead to recruitment of even more DCs at the site of antigen contact. The release of IL-1β by epidermal LCs promotes their egress from the epidermis. LCs process the antigen while migrating to regional lymph nodes, where they present it to naïve T cells. This phase is influenced by multiple factors. Defects in the integrity of the stratum corneum allow greater penetration of allergen and increase the chances of activating antigen-presenting cells (APCs) in skin. The availability and viability of APCs in skin, as well as the presence or absence of cytokines produced by keratinocytes, can promote or hinder APC–T-cell engagement.

In the draining lymph nodes, LCs present the peptides to T cells and activate CD4 and CD8 antigen-specific T cells. An important property of LCs and DCs is their ability to present exogenous antigens on both major histocompatibility complex (MHC) class I and class II molecules. This cross-priming leads to the activation of both CD4 and CD8 hapten-specific T cells. Although classic delayed-type hypersensitivity reactions are mediated primarily by CD4 cells, contact dermatitis to haptens is mediated primarily by CD8 cells with a Th1-type cytokine profile. LCs activate hapten-specific T cells, which include Th1, Th2, Th17, and Treg subsets.[30] This preexisting mix of T-cell subtypes specific for the antigen influence the outcome of this process. The higher the frequency of cells of an effector subtype, the higher the likelihood that dermatitis will result, whereas a higher frequency of cells of a regulatory subtype may limit or prevent the development of dermatitis.

The efferent phase of ACD occurs on subsequent contact of skin with the hapten. Antigen-specific memory T cells and other inflammatory cells leave vessels and enter skin through sequential activation of a number of adhesion molecules by cytokines. Memory T cells constitutively express CLA. E-selectin, the ligand for CLA, is induced on vascular endothelium by inflammatory mediators, such as IL-1 and TNF. This interaction causes memory T cells to slow down and roll along the endothelial surface as a prelude to migration to sites of inflammation. Firm adhesion and migration of leukocytes to the endothelium are mediated by T-cell very late antigen-4 (VLA-4)/leukocyte function–associated antigen-1 (LFA-1) and endothelial cell vascular cell adhesion molecule-1 (VCAM-1)/intercellular adhesion molecule 1 (ICAM-1), respectively (see Chapter 16). Subsequently, LFA-1^+ T cells migrate toward ICAM-1^+ epidermal cells.

Mast cells can also participate in the elicitation phase. Mast cells contain preformed TNF, which may regulate the adhesion molecules involved in the early recruitment of Th cells. The net result is an influx of sensitized T cells homing to the hapten-provoked skin site, releasing their inflammatory mediators, IL-2 and IFN-γ, and this results in an enhanced immune response through activation and recruitment of more inflammatory cells, producing spongiosis and the inflammatory dermal infiltrate characteristic of ACD.

Although Th1 cells have been classically considered the primary effector cells of ACD (responses to haptens, such as nickel, were dominated by IFN-γ–producing cells), recent studies have indicated that Th2 cells also participate in the development of contact hypersensitivity.[31]

Recently, both murine models and human studies have suggested the potential role of Th17 cells in the immunopathogenesis of ACD. In patients with ACD, Th17-associated mediators, such as IL-17A, IL-17F, IL-22, IL-23, chemokine receptor 6 (CCR6), IL-22 receptor, and the Th17 transcription factor retinoic acid–related orphan receptor γ (ROR-γ) were shown to be produced by nickel-specific T cells and were upregulated in ACD lesional skin and positive patch test biopsy specimens.[32] Nickel exposure was reported to induce production of IL-23 by keratinocytes, promoting a Th17-mediated response, as detected by the presence of IL-17–producing T cells in peripheral blood from patients with nickel allergy.[33] The role of IL-17 in ACD lesions includes induction of keratinocyte release of cytokines and chemokines (i.e., IL-8 and IL-6) and promotion of T cell–induced apoptosis of keratinocytes.

The immunological mechanism of ACD appears to be hapten-specific with allergens like nickel inducing Th1/Th17 polarization, while fragrance induces Th2/Th22 polarization.[34]

The concept that ACD may differ mechanistically according to allergen is likely to have significant therapeutic relevance as a consideration in the choice of future targeted therapy.

The skin of patients with AD is at increased risk for the development of ACD, and this may be attributed to multiple factors. First, AD skin is exposed to a large number of different chemicals used to treat AD, including moisturizers, topical corticosteroids, and TCIs. Second, the disturbed barrier system allows increased permeation of allergens leading to greater access of surface antigens to LCs. Third, AD skin has a heightened immunological status with existing activation of innate immunity and selective upregulation of the Th2 adaptive immune response. In patients with AD, cutaneous contact with irritants and allergens leads to amplification of innate immunity and enhanced adaptive immune responses, including Th2 and Th17 in patients with acute AD and Th22 and Th1 in patients with chronic disease. Just as innate immune activation stimulated by an irritant permits a lower threshold of ACD elicitation, the amplified adaptive responses in lesional and nonlesional skin promote increased expression of ACD and ICD in patients with AD.

PATHOGENESIS OF IRRITANT CONTACT DERMATITIS

ICD accounts for 80% of cases of CD. The clinical presentation of ICD is usually restricted to skin that is directly in contact with the offending agent, with little or no extension beyond the site of contact. The inflammatory response is dose- and time-dependent. Any impairment to the epidermal barrier layer (e.g., fissuring, overhydration) renders skin more susceptible to an irritant effect. Contact with agents such as detergents, solvents, alcohol, creams, lotions, ointments and powders, and environmental factors such as wetting, drying, perspiration, and temperature extremes, can abrade or irritate skin. After exposure to skin irritants, perturbation of the skin barrier and disorganization of the lipid bilayers with lamellar body lipid extrusion in the epidermis is seen with an associated increase in TEWL.[35]

Irritants cause direct activation of the innate immune system through hyperproduction of cytokines and chemokines, such as IL-1α, IL-1β, IL-6, IL-8, and TNF, which further induce a cytokine cascade and inflammatory reaction with infiltration of inflammatory cells. Epidermal keratinocytes have been identified as key effector cells in the initiation and propagation of contact irritancy. Keratinocytes can release both preformed and newly synthesized cytokines and can upregulate MHC class II molecules and induce adhesion molecules in response to irritants.[30] These mediators can cause direct tissue damage, activating LCs, dermal DCs, and endothelial cells, which contribute to further cellular recruitment, including neutrophils, lymphocytes, macrophages, and mast cells, which also contribute to the inflammatory cascade. The "final" cellular damage results from inflammatory mediators released by activated nonsensitized T cells.

Sodium lauryl sulfate (SLS), a common ingredient of cleaning products, is directly toxic to keratinocytes (an experimental model of ICD) and has been demonstrated to induce LC mobilization and consequent migration to the draining lymph nodes. Other irritants have been shown to cause direct activation of the innate immune system through a set of membranous and intercellular receptors called Toll-like receptor 7 (TLR-7) and nucleotide-binding oligomerization domain (NOD)–like receptors (NLRs) that activate the inflammasome and nuclear factor-κB (NF-κB) pathways, inducing release of many cytokines and chemokines.

Although few immunohistopathological differences have been noted between ICD and ACD, both ICD and ACD begin with activation of the innate immune system.

ICD does not require prior sensitization, and immunological memory is not involved. The cellular infiltrate includes CD4 T cells with a Th1-type profile. Irritant reactions predispose to allergic reactions, making patients with ICD more susceptible to the development of ACD. In human studies, when patch test sites were pretreated with SLS, the threshold elicitation concentrations of contact allergens, such as cobalt, nickel, and colophony, were significantly decreased, indicating that the activation of innate immunity by irritants likely reduces the threshold for the development of ACD.

In summary, although ACD, ICD, and AD are independently defined diseases, they interact and frequently occur in combination. In patients with AD, if the activation of an innate immune response is preceded by exposure to irritants and allergens, there is altered permeation of contact allergens, which are introduced into an already activated innate immune system, resulting in amplification of innate immunity and enhanced adaptive immune responses. The amplified adaptive responses in AD promote increased ACD and ICD. Conversely, the innate immune activation stimulated by an irritant permits a lower threshold of ACD elicitation (see Fig. 48.1).

CLINICAL MANIFESTATIONS OF CONTACT DERMATITIS

The clinical and histological findings of ACD are characteristic, but not diagnostic. ACD should be suspected in patients with acute and chronic eczematous or noneczematous dermatitis and is diagnosed on the basis of the clinical appearance of the lesions, the distribution of the dermatitis, the presence of pruritus, and the absence of other etiologies. Acute CD is characterized by erythematous papules, vesicles, oozing, and crusted lesions (Fig. 48.2). Recurrence and persistence of the dermatitis may lead to subacute and chronic lesions. Subacute CD manifests as erythema, scaling, fissuring, or a parched, scalded appearance, and chronic inflammation may have more skin thickening, hardening, scaling, fissuring, and lichenification.

Although the location of the dermatitis serves as an important clue to the source of the offending chemical, multiple factors contribute to the distribution of ACD. Spread from the principal site of exposure can involve distant sites, either by inadvertent contact or by autosensitization. Areas of the scalp, palms, and soles have thicker skin, whereas the eyelid, face, and genital areas have thinner skin that is more sensitive to contact allergens (Figs. 48.3 and 48.4). A geographical approach can be very helpful in identifying the causal allergen.

The hand is a common body site involved in contact dermatitis.[36] Hand dermatitis may be due to ICD, ACD, AD, dyshidrosis, or psoriasis. Because the skin on the palm is much thicker than that on the dorsum of the hands, ACD is often associated with vesicles, occurring most often on the thinner skin on the dorsum of the hands, the fingertips, nail folds, and less commonly involves the palms (Fig. 48.5). Patch tests in patients with hand eczema showed that relevant allergens included nickel sulfate, potassium dichromate, rubber accelerators, and cobalt

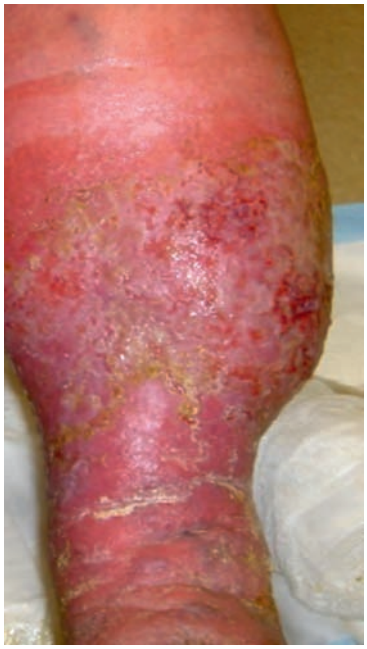


FIG. 48.2 Acute Allergic Contact Dermatitis.

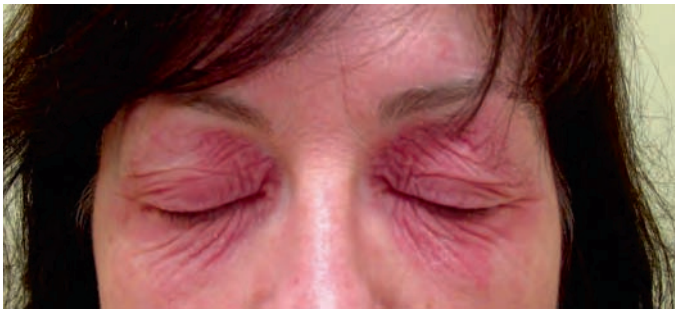


FIG. 48.3 Allergic Contact Dermatitis of the Eyelid.

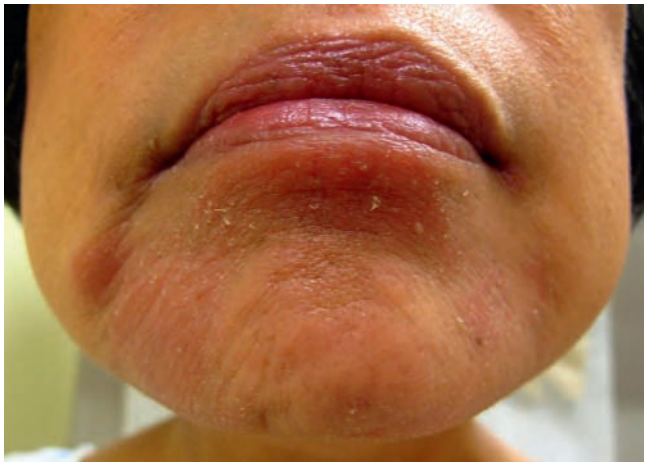


FIG. 48.4 Allergic Contact Dermatitis of the Face Caused by Fragrance.

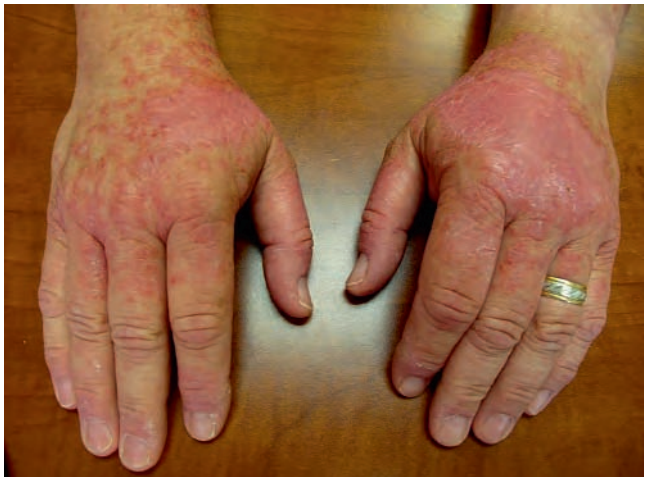


FIG. 48.5 Allergic Contact Dermatitis of the Hand Caused by Rubber Accelerators.

chloride.[37] The prevalence of hand eczema in patients with AD is 2- to 10-fold higher than in nonatopics.

Facial presentation of dermatitis ranks third by location in the results from the NACDG from 2015 to 2016.[36] Symmetrical, central, or patchy distribution tends to be seen with cosmetics and personal products (moisturizers, sunscreens, foundations, and powders) whereas peripheral face (pre-auricular, submental and jawline areas) involvement can be due to shampoos, conditioners, and facial cleansers. Involvement of the lateral neck could be due to perfumes/colognes, ectopic transfer of nail cosmetics, or as a "run-off" pattern from cosmetics applied to face, scalp, or hair. Although the regional distribution may be helpful in some cases, dermatitis with scattered generalized distribution, which lacks the characteristic distribution that gives a clue as to the possible etiology of ACD, is actually the second most common pattern of dermatitis in both children and adults, as reported by the North American Contact Dermatitis Group in 2018.[36] Systemic contact dermatitis (SCD), specifically the "baboon syndrome," is a diffuse eruption involving flexural and intertriginous areas following oral, intravenous, or transcutaneous exposure to the allergen in a contact-sensitized individual. Aside from the baboon syndrome, SCD could also manifest as a recall reaction (reactivation of a previous site of dermatitis or a previous positive patch test), dyshidrotic hand eczema, flexural dermatitis, exanthematous rash, erythroderma, and even vasculitis-like lesions. The most common causes of SCD are (1) metals, such as mercury, nickel, and gold; (2)medications, including aminoglycoside antibacterial, topical corticosteroids, and aminophylline; and (3) plants and herbal products, including Compositae and Anacardiaceae families and Balsam of Peru (also known as *Myroxylon pereirae* resin).

Histologically, CD demonstrates intercellular edema of the epidermis (spongiosis) with varying degrees of acanthosis (thickening of the epidermal stratum basale and stratum spinosum) and superficial perivascular, lymphohistiocytic infiltrates. It is often difficult to distinguish ACD (see Fig. 48.5) from ICD (Fig. 48.6) on the basis of physical examination or histological findings.

MANAGEMENT OF ALLERGIC CONTACT DERMATITIS

The management of ACD includes identification of the allergen, avoidance, pharmacological intervention, and prevention.

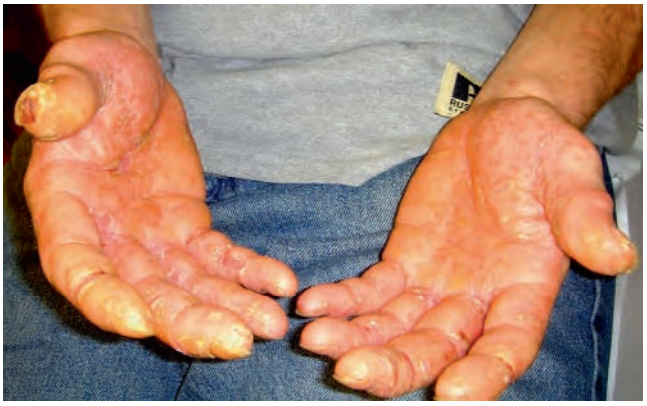


FIG. 48.6 Irritant Contact Dermatitis.

THERAPEUTIC PRINCIPLES

Contact Dermatitis

- Patch testing is the procedure of choice to confirm the diagnosis of allergic contact dermatitis (ACD) and to identify offending contact allergens.
- The interpretation of patch tests requires both experience and judgment on its relevance.
- Once identified, the key to management of ACD is prevention by avoiding substances containing the allergens or irritants that have been identified and providing patients with alternatives and substitutes of products that do not contain their allergen.
- Medical treatments such as topical and/or systemic corticosteroids can be used to relieve ongoing dermatitis.

Identification of the Allergen

The most widely acceptable and available method for confirming the diagnosis of ACD is patch testing. Nickel remains the most common contact sensitizer and is more common in women than in men, likely because of greater exposure to nickel in jewelry and body piercing practices. In cosmetic and personal products, fragrances, preservatives, and emulsifiers are the most common causative allergens. In addition, paraphenylenediamine (in hair dye), cocamidopropyl betaine (in shampoos and soaps), and medications (e.g., neomycin, benzocaine, corticosteroids) are commonly reported allergens in personal hygiene and medical products. Methylisothiazolinone is a preservative widely used in cosmetics, toiletries, and household products including shampoos, conditioners, baby soaps, detergents, dishwashing liquid, stain removers, and fabric softeners. It is the second most common positive patch test (second only to nickel) in the 2018 NACDG patch test results.[36] The results of patch tests must be interpreted in the context of the patient's experience; exposure and relevance should be established.[38]

Allergen Avoidance and Treatment

Once the allergen is identified, avoidance of contact to the allergen and education of patients and/or families is the mainstay of treatment for ACD. All other measures are palliative and temporary. Patients and/or caregivers must be educated regarding the nature of the dermatitis, triggering agents and irritant factors, and avoidance of further contact with the offending agent. A list of the offending agents, potential exposure alternatives, and substitutes should be offered to the patient. The names of the allergens are long, complex, and confusing. There are also numerous synonyms and cross reactions, thus a list of "safe products" to use can increase compliance.

The dimethylglyoxime test (nickel spot test) and the cobalt spot test (based on disodium-1-nitroso-2-naphthol-3, 6-disulfonate) can be used to detect nickel or cobalt released from metal objects and dermal exposure, thus aiding in avoidance of contact in sensitized patients.

Cool compresses with aluminum subacetate (Burrow's solution), calamine, or colloidal oatmeal may help acute, oozing lesions. Excessive handwashing should be discouraged in patients with hand dermatitis, and nonirritating or sensitizing moisturizers must be used after washing. Soaps and nonalkaline cleansers should be avoided.

In addition to avoidance of exposure, appropriate adjunct medical treatment can be prescribed. First-line therapy is with topical corticosteroids, and second-line treatment includes phototherapy, oral retinoids, and immunosuppressant agents. Topical corticosteroids may be sufficient for localized lesions. Patients with dermatitis that is acute, extensive (particularly if involving greater than 10% of total body surface), or severe (such as poison ivy) may benefit from systemic therapy. Due to its high side effect profile, their use for chronic allergic contact dermatitis is not recommended.

Contact sensitization to the corticosteroid itself, the vehicle, or other ingredients in the topical corticosteroid should be suspected if symptoms worsen, initially improve but then worsen with continued treatment, or do not respond to treatment at all. Several topical T-cell selective inhibitors (tacrolimus and pimecrolimus) have been used successfully in the treatment of AD, but their efficacy in ACD or ICD has not been established.

Antihistamines may offer some benefit in contact urticaria and some relief from pruritus.

Phototherapy (UVB) may be effective in the treatment of ACD, especially in chronic ACD of the hands. Other immunomodulating agents such as cyclosporin, methotrexate, azathioprine, and mycophenolate mofetil have been used, especially in patients who could not avoid their allergens.

There are conflicting reports on whether dupilumab is effective or exacerbates some cases.

New evidence is emerging to support that dupilumab (IL-4RI) may be a treatment option in some patients with recalcitrant ACD.[39]

Mechanical barriers such as protective gloves and clothing and barrier creams may help with avoidance of exposure, especially in the occupational setting.[40] For nickel-allergic patients, barriers such as gloves, covers for metal buttons, and identification of nickel by the dimethylglyoxime test can be prescribed, but results can be disappointing.

Patient education regarding the nature of the dermatitis, triggering agents and irritant factors, plus instruction for avoidance and appropriate substitutes will not only aid in clearing the dermatitis but also prevent or minimize recurrences. At present, hyposensitization of patients with ACD is not a viable therapy.

PERSPECTIVES IN ATOPIC DERMATITIS AND ALLERGIC CONTACT DERMATITIS

Looking into the future, as we define the immune pathways that cause AD, novel biologics and molecules currently in clinical trials may provide more targeted therapy in AD.

ON THE HORIZON

- Novel biologicals and small molecules currently in clinical trials may provide more targeted therapy in atopic dermatitis (AD).
- Characterization of unique clinical phenotypes of AD may lead to a better understanding of AD pathophysiology and a precision-medicine approach to treatment.
- Identification of "at risk" patients will facilitate preventative strategies.
- A large number of chemicals are in everyday usage: new, rare, and emerging allergens should be considered in the assessment of suspected ACD. Patch testing to these allergens has yet to be standardized and irritant and elicitation concentration to patch testing determined.
- Data about contact allergen sensitization in children with AD are limited and are continually expanding. Frequency and patterns of CD in children with AD and results of patch testing have yet to be determined.

Identification of the endotypes involved in the pathogenesis of unique clinical phenotypes of AD will be required for a precision-medicine approach to treatment. In the absence of a low-cost cure for AD, biomarkers to identify "at-risk" patients could facilitate preventative strategies.

In the area of ACD, a large number of chemicals are being evaluated for their role, and new, rare, and emerging allergens should be considered in the assessment of suspected ACD. Patch testing to these allergens has yet to be standardized; irritant and elicitation concentrations to patch testing need to be determined. Data about contact allergen sensitization in children with AD are limited but are continually expanding. Frequency and patterns of CD in children with AD and the results of patch testing have yet to be determined.

REFERENCES

1. Goleva E, Berdyshev E, Leung DY. Epithelial barrier repair and prevention of allergy. *J Clin Invest*. 2019;129(4):1463–1474.
2. Boguniewicz M, Leung DY. Recent insights into atopic dermatitis and implications for management of infectious complications. *J Allergy Clin Immunol*. 2010;125(1):4–13.
3. Odhiambo JA, Williams HC, Clayton TO, et al. Global variations in prevalence of eczema symptoms in children from ISAAC Phase Three. *J Allergy Clin Immunol*. 2009;124(6):1251–1258.
4. Fuxench ZC, Block JK, Boguniewicz M, et al. Atopic dermatitis in America study: a cross-sectional study examining the prevalence and disease burden of atopic dermatitis in the US adult population. *J Invest Dermatol*. 2019;139:583–590.
5. Margolis JS, Abuabara K, Bilker W, et al. Persistence of mild to moderate atopic dermatitis. *JAMA Dermatol*. 2014;150(6):593–600.
6. Irvine AD, McLean WH, Leung DY. Filaggrin mutations associated with skin and allergic diseases. *N Engl J Med*. 2011;365(14):1315–1327.
7. Boguniewicz M, Leung DY. Atopic dermatitis: a disease of altered skin barrier and immune dysregulation. *Immunol Rev*. 2011;242(1):233–246.
8. Schneider L, Tilles S, Lio P, et al. Atopic dermatitis: a practice parameter update 2012. *J Allergy Clin Immunol*. 2013;131(2):295–299.
9. Eichenfield LF, Tom WL, Chamlin SL, et al. Guidelines of care for the management of atopic dermatitis: section 1. Diagnosis and assessment of atopic dermatitis. *J Am Acad Dermatol*. 2014;70(2):338–351.
10. Boguniewicz M, Fonacier L, Guttman-Yassky E, et al. Atopic dermatitis yardstick: practical recommendations for an evolving therapeutic landscape. *Ann Allergy Asthma Immunol*. 2018;120:10–22.
11. Suárez-Fariñas M, Tintle SJ, Shemer A, et al. Nonlesional atopic dermatitis skin is characterized by broad terminal differentiation defects and variable immune abnormalities. *J Allergy Clin Immunol*. 2011;127(4):954–964.
12. Sidbury R, Davis DM, Cohen DE, et al. Guidelines of care for the management of atopic dermatitis: section 3. Management and treatment with phototherapy and systemic agents. *J Am Acad Dermatol*. 2014;71(2):327–349.
13. Guttman-Yassky E, Hanifin JM, Boguniewicz M, et al. The role of phosphodiesterase 4 in the pathophysiology of atopic dermatitis and the perspective for its inhibition. *Exp Dermatol*. 2019;28(1):3–10.
14. Chopra R, Vakharia PP, Sacotte R, et al. Efficacy of bleach baths in reducing severity of atopic dermatitis: a systematic review and meta-analysis. *Ann Allergy Asthma Immunol*. 2017;119:435–440.
15. Boguniewicz M. Biologic therapy for atopic dermatitis: moving beyond the practice parameter and guidelines. *J Allergy Clin Immunol Pract*. 2017;5:1477–1487.
16. Guttman-Yassky E, Bissonnette R, Ungar B, et al. Dupilumab progressively improves systemic and cutaneous abnormalities in patients with atopic dermatitis. *J Allergy Clin Immunol*. 2019;143:155–172.
17. Simpson EL, Paller AS, Siegfried EC, et al. Efficacy and safety of dupilumab in adolescents with uncontrolled moderate to severe atopic dermatitis: a phase 3 randomized clinical trial. *JAMA Dermatol*. 2020;156:44–56.
18. Brar K, Nicol NH, Boguniewicz M. Strategies for successful management of severe atopic dermatitis. *J Allergy Clin Immunol Practice*. 2019;7:1–16.
19. Roekevisch E, Spuls PI, Kuester D, et al. Efficacy and safety of systemic treatments for moderate-to-severe atopic dermatitis: a systematic review. *J Allergy Clin Immunol*. 2014;133(2):429–438.
20. Tam H, Calderon MA, Manikam L, et al. Specific allergen immunotherapy for the treatment of atopic eczema. *Cochrane Database Syst Rev*. 2016;2:CD008774.
21. Wang HH, Li YC, Huang YC. Efficacy of omalizumab in patients with atopic dermatitis: a systematic review and meta-analysis. *J Allergy Clin Immunol*. 2016;138:1719–1722.
22. Chan S, Cornelius V, Cro S, et al. Treatment effect of omalizumab on severe pediatric atopic dermatitis: the ADAPT randomized clinical trial. *JAMA Pediatr*. 2020;174:29–37.
23. Makrgeorgou A, Leonardi-Bee J, Bath-Hextall FJ, et al. Probiotics for treating eczema. *Cochrane Database Syst Rev*. 2018;11:CD006135.
24. Renert-Yuval Y, Guttman-Yassky E. New treatments for atopic dermatitis targeting beyond IL-4/IL-13 cytokines. *Ann Allergy Asthma Immunol*. 2020;124:28–35.
25. Skjerven HO, Rehbinder EM, Vettukattil R, et al. Skin emollient and early complementary feeding to prevent infant atopic dermatitis (PreventADALL): a factorial, multicentre, cluster-randomised trial. *Lancet*. 2020;395:951–961.
26. Chalmers JR, Haines RH, Bradshaw LE. Daily emollient during infancy for prevention of eczema: the BEEP randomised controlled trial. *Lancet*. 2020;395:962–972.
27. Goldenberg A, Mousdicas N, Silverberg N, et al. Pediatric contact dermatitis registry inaugural case data. *Dermatitis*. 2016;27(5):293–302.
28. Aquino M, Fonacier M. The role of contact dermatitis in patients with atopic dermatitis. *J Allergy Clin Immunol Pract*. 2014;2(4):382–387.
29. Ross-Hansen K, Menné T, Johansen JD, et al. Nickel reactivity and filaggrin null mutations--evaluation of the filaggrin bypass theory in a general population. *Contact Dermatitis*. 2011;64(1):24–31.
30. Gittler JK, Krueger JG, Guttman-Yassky E. Atopic dermatitis results in intrinsic barrier and immune abnormalities: implications for contact dermatitis. *J Allergy Clin Immunol*. 2013;131(2):300–313.
31. Minang JT, Troye-Blomberg M, Lundeberg L, et al. Nickel elicits concomitant and correlated in vitro production of Th1-, Th2-type and regulatory cytokines in subjects with contact allergy to nickel. *Scand J Immunol*. 2005;62(3):289–296.
32. Zhao Y, Balato A, Fishelevich R, et al. Th17/Tc17 infiltration and associated cytokine gene expression in elicitation phase of allergic contact dermatitis. *Br J Dermatol*. 2009;161(6):1301–1306.
33. Larsen JM, Bonefeld CM, Poulsen SS, et al. IL-23 and T(H)17-mediated inflammation in human allergic contact dermatitis. *J Allergy Clin Immunol*. 2009;123(2):486–492.
34. Dhingra N, Shemer A, Da Rosa JC, et al. Molecular profiling of contact dermatitis skin identifies allergen-dependent differences in immune response. *J Allergy Clin Immunol*. 2014;134(2):362–372.
35. de Jongh CM, Jakasa I, Verberk MM, et al. Variation in barrier impairment and inflammation of human skin as determined by sodium lauryl sulphate penetration rate. *Br J Dermatol*. 2006;154(4):651–657.
36. DeKoven JG, Warshaw EM, Zug KA, et al. North American contact dermatitis group patch test results: 2015–2016. *Dermatitis*. 2018;29(6):297–309.

37. Warshaw EM, Ahmed RL, Belsito DV, et al. Contact dermatitis of the hands: cross-sectional analyses of North American Contact Dermatitis Group Data, 1994–2004. *J Am Acad Dermatol*. 2007;57(2):301–314.
38. Fonacier L. A practical guide to patch testing. *J Allergy Clin Immunol Pract*. 2015;3(5):669–675.
39. Puza CJ, Atwater AR. Positive patch test reaction in a patient taking dupilumab. *Dermatitis*. 2019;29(2):89.
40. Fonacier L, Bernstein DJ, Pacheci K, et al. Contact dermatitis: a practice parameter-updated 2015. *J Allergy Clin Immunol Pract*. 2015;3(Suppl_3):S1–S39.

Food Allergy

J. Andrew Bird and A.Wesley Burks

Adverse reactions to foods may be mediated by immunological mechanisms (food allergy) and nonimmunological mechanisms (food intolerance). Immune-mediated reactions to foods most commonly involve immunoglobulin E (IgE), an antibody that binds to mast cells and basophils. Upon exposure to an allergen, they release inflammatory mediators, such as histamine, prostaglandins, and leukotrienes, producing symptoms ranging from localized oral itching to anaphylaxis, the latter being a potentially fatal systemic reaction. Non–IgE-mediated immune reactions to foods include such diseases as celiac disease and food protein–induced enterocolitis syndrome (FPIES), while other diseases, such as eosinophilic esophagitis (EoE) and eosinophilic gastroenteritis (EGE), involve both IgE-mediated and non–IgE-mediated mechanisms. Food allergies have a significant negative effect on quality of life, and the burden on the health system amounts to approximately $25 billion annually.[1] Current management of food allergy relies on avoidance, but ongoing research into treatment for food allergy may offer interventional approaches to provide additional protection for allergic patients. This chapter reviews the basics of food allergy diagnosis, management, and natural history, with a specific focus on IgE-mediated food allergies.

KEY CONCEPTS

Common Characteristics of Most Frequently Encountered Food Allergens

- A relatively small molecular weight (<70 kDa)
- Abundant source of the relevant allergen
- Glycosylation residues
- Water solubility
- Most are resistant to heat and digestion

PREVALENCE

The true prevalence of food allergy is difficult to establish. This is most likely because (1) studies have focused on only the most common food allergens; (2) the incidence and prevalence of food allergy may be increasing and changing with time, with studies showing an increasing prevalence over the past 10 to 20 years; and (3) studies of prevalence are difficult to compare because of inconsistencies and deficiencies in study design.[2] Population-based surveys indicate that about 10% of the US population (1 in 10 adults and 1 in 12 children) suffer from at least one IgE-mediated food allergy.[3] More than 170 foods have been reported to trigger allergic reactions, though the most common foods that trigger reactions in US children are peanuts (2.2%), milk (1.9%), shellfish (1.3%), tree nuts (1.2%), egg (0.9%), finned fish (0.6%), wheat (0.5%), soy (0.5%), and sesame seed (0.2%).[3] Prevalent allergens vary in different cultural groups, with milk, egg, peanut, and tree nut allergens topping the lists in the Americas, Australia, and Western Europe, whereas fish and shellfish allergies are more common in Asia.[4] Food allergies commonly co-occur with other atopic diseases, such as atopic dermatitis (AD), asthma, and allergic rhinitis.[5] The most common food and plant allergens are listed in Table 49.1.

SPECTRUM OF DISEASE

Immunoglobulin E–Mediated Food Allergies

Immediate reactions to foods compose the largest proportion of food-induced allergic diseases. IgE-mediated food allergy occurs when allergenic proteins cross-link allergen-specific IgE bound to mast cells or basophils, leading to release of histamine and other inflammatory mediators. Symptoms of IgE-mediated food allergies usually occur within minutes of ingestion of the provoking food allergen, and do not present more than 2 hours later, except in rare circumstances (*e.g.*, delayed anaphylaxis related to red meat ingestion).

Symptoms may be severe (*e.g.*, anaphylaxis) or localized (*e.g.*, pollen-food allergy syndrome). Characteristic signs of an immediate allergic reaction manifest through skin and subcutaneous tissues (*e.g.*, urticaria and/or angioedema), the respiratory system (*e.g.*, bronchospasm), the gastrointestinal system (*e.g.*, vomiting and/or diarrhea), and/or the cardiovascular system (*e.g.*, increased vascular permeability leading to hypotension). Anaphylaxis is the most severe symptom of food allergy and may result in death (Chapter 46).

Pollen-food allergy syndrome (PFAS, or oral allergy syndrome) symptoms are restricted to the lips, throat, and mouth and are most commonly elicited due to fruit and vegetable proteins cross-reacting with antibodies against pollen proteins in individuals with pollen allergy. Affected individuals will typically complain of pruritus and/or tingling of the lips, tongue, roof of the mouth, and throat with or without swelling. PFAS reactions are unlikely to progress to a systemic reaction.

Because of protein similarities between allergens, cross-reactivity occurs. Patients allergic to certain foods should be counseled to avoid cross-reactive food proteins. It is common to find cross-reactivity among tree nuts; in particular, cashews and pistachios share common allergen binding sites, as do walnuts and pecans. Between 25% and 50% of patients with a peanut allergy are also allergic to tree nuts, with particular

TABLE 49.1 Food Allergens by Family

Food Allergen Family	Foods Containing Allergen	Description
Tropomyosins	Crustacean shellfish (*e.g.*, shrimp, lobster, crab), mollusks (*e.g.*, oyster, scallop, squid)	Invertebrate tropomyosins are a family of muscle proteins that share homology across invertebrate species and therefore may act as panallergens. They do not share homology with vertebrate tropomyosins. They are generally heat stable and highly cross-reactive.
Parvalbumins/EF-hand proteins	Vertebrate fish and frogs	Muscle proteins that possess a calcium-binding domain referred to as an EF-hand motif. This is the second-largest family of allergens, and these allergens are considered highly cross-reactive panallergens.
Casein	Mammalian milk	Function to bind calcium and stabilize it in micellar form. There is high sequence homology between cow's milk and other mammalian milks, such as goat's milk and sheep's milk. Other animal milks such as human milk, horse's milk, donkey's milk, and camel's milk have caseins with approximately 60% homology, which may account for less allergenicity than seen with cow's milk.
Prolamin superfamily	Seeds, tree nuts, legumes (including peanut), fruits, vegetables, wheat, corn, rice	This family contains the highest number of plant food allergens and is characterized by rich disulfide bonds and a core of eight conserved cysteine residues, providing stability and resistance to digestion. Families within this superfamily include 2S albumin seed storage proteins, nonspecific lipid transfer proteins, and α-amylase/trypsin inhibitors.
Cupin superfamily	Legumes, nuts, seeds	A large and functionally diverse superfamily of proteins termed seed storage globulins that share a β-barrel structural core domain. Seed storage globulins may be grouped into two families: vicilins and legumins.
Bet v 1 superfamily	Apple, pear, stone fruits, celery, carrot, soybean, peanut	Bet v 1 is the major birch pollen allergen and is a member of the pathogenesis-related protein 10 family within this superfamily. Symptoms of Bet v 1 hypersensitivity typically present with pollen food allergy syndrome (also known as oral allergy syndrome), which is caused by IgE cross-reactivity between Bet v 1 and homologous allergens from plant foods.

From Sampson HA, Aceves S, Bock SA, et al. Food allergy: a practice parameter update—2014. *J Allergy Clin Immunol.* 2014;134:1016–1025.e1043.

cross-reactivity noted between peanut allergens and tree nut allergens in almonds, walnuts, pecans, hazelnuts, and Brazil nuts. Tropomyosins found in crustacean shellfish are a panallergen,[5] and approximately 75% of individuals who are allergic to one crustacean (*e.g.*, shrimp) are also likely to react to another crustacean (*e.g.*, lobster).[8] Parvalbumins found in vertebrate fish are commonly cross-reactive on testing, but clinical relevance of cross-reactivity varies. Studies have shown that an individual who is allergic to one fish species has approximately a 50% likelihood of reacting to another species of fish.

Although most reactions occur immediately after ingestion, some individuals may experience delayed anaphylaxis following ingestion of mammalian meat.[5] Delayed allergy to mammalian meats has been linked to the production of IgE to α-gal in susceptible subjects. α-Gal is an immunogenic oligosaccharide, and sensitization is believed to occur via a tick bite. Symptoms of urticaria, angioedema, and anaphylaxis can occur 3 to 6 hours after eating beef, pork, lamb, and venison. The mechanism(s) underlying the delayed reaction is poorly understood.

Mixed Immunoglobulin E/Non–Immunoglobulin E– and Non–Immunoglobulin E–Mediated Food Allergies

Delayed gastrointestinal (GI) reactions to foods include such diseases as EoE, EGE, FPIES, and eosinophilic proctocolitis. The most common food proteins involved in these diseases include milk, egg, wheat, and soy. Celiac disease is a non–IgE-mediated food allergy triggered by ingestion of gluten-containing grains (*e.g.*, wheat, barley, and rye). Human leukocyte antigen (HLA) DQ2- or DQ8-restricted CD4 T cells, which recognize gluten selectively in affected persons, are critical to the pathogenesis of celiac disease.

EoE is a clinicopathological diagnosis, based on symptoms of esophageal dysfunction (including dysphagia, vomiting, feeding disorders, and abdominal pain) together with pathological findings of at least 15 eosinophils per high-power field on light microscopy.[9] The precise role of food allergy in EoE is not well defined; IgE-mediated and non–IgE-mediated mechanisms may be involved in the pathogenesis. The most common food allergens implicated in the pathogenesis include milk, egg, wheat, and soy. Numerous other foods have been implicated in the pathogenesis of EoE, and a common approach to treatment is initial elimination of milk, egg, wheat, soy, peanut, tree nuts, fish and shellfish. If dietary elimination is not successful or not feasible for the patient, then topical (swallowed aerosol) inhaled steroids (*e.g.*, fluticasone or budesonide) may be swallowed rather than inhaled to treat the inflammation.

EGE is less common than EoE and, like EoE, it is believed that its pathogenesis involves both IgE-mediated and non-IgE-mediated mechanisms. Common symptoms of EGE include vomiting, abdominal pain, diarrhea, and failure to thrive/weight loss. Multiple food allergens are often implicated, although response to dietary elimination of the most common food allergens (milk, egg, wheat, soy, peanut, tree nuts, fish, and shellfish) is typically less successful than dietary elimination reported in patients with EoE. Topical (swallowed aerosol) inhaled steroids (*e.g.*, fluticasone and budesonide) may provide some benefit; however, systemic steroids are often necessary for disease control.

FPIES is a non–IgE-mediated disease usually occurring in infants.[10] Characteristic symptoms of FPIES manifest as repetitive emesis with or without diarrhea accompanied by lethargy occurring 2 to 4 hours after ingestion of the offending food protein. The risk of abrupt volume loss, hypotension, and potential for bowel perforation makes this a medical emergency. Treatment is reliant on rehydration. Ondansetron may be helpful in managing acute FPIES reactions. Milk and soy are most commonly implicated, along with less common food allergens such as rice, oats, fruits, or vegetables. FPIES is outgrown in the majority of affected children by 3 years of age but may be protracted for many years in a smaller subset of patients.

Celiac disease is an immune-based reaction to gluten, a storage protein for wheat, barley, and rye. The small intestine is typically affected in genetically predisposed individuals, and symptoms resolve with gluten avoidance. Symptoms of celiac disease are variable and may include diarrhea, steatorrhea, weight loss, bloating, flatulence, abdominal pain, and also non-GI symptoms, such as abnormal results on liver function tests, iron deficiency anemia, bone disease, and skin disorders. Celiac disease is detected with serological testing of celiac-specific antibodies and confirmed by duodenal mucosal biopsy, both of which should be performed while the patient is on a gluten-containing diet.[11]

PATHOPHYSIOLOGY

Food allergy results from a breakdown of oral tolerance (or failure to develop it); foods that are ordinarily harmless may then trigger an immune response that results in harmful adverse symptoms upon exposure. Maintaining tolerance requires a delicate balanced effort from multiple arms of the immune system. Deviation from the protective response may result in the development of an allergic response.

PROPERTIES OF FOOD ALLERGENS

An intact GI mucosal barrier is required to maintain tolerance. The first line of defense against the mucosal immune system is a hydrophobic layer of mucin oligosaccharides, which serve to trap antigen.[12] Secretory IgA is also a part of the outer layer of the intestinal defense against dietary antigens. Dietary antigens must then penetrate the intestinal epithelium, which is maintained by epithelial junction complexes (adherens junctions) and tight junctions. Intestinal epithelial barrier dysfunction may play a role in food allergen sensitization. Alterations in the integrity of junctional complexes may be induced by calcineurin inhibitors, and this can result in food allergen sensitization. Genetic defects, such as those in individuals with filaggrin mutations, a protein that binds to keratin and is important for epithelial cell integrity, may predispose individuals to increased risk of EoE. Other factors that have been shown to affect intestinal permeability include viruses, alcohol, and nonsteroidal anti-inflammatory drugs (NSAIDs). These environmental exposures may alter intestinal epithelial integrity, thus allowing antigen to interact with the next layer of defense, mucosa associated lymphoid tissue (MALT; Chapter 24).

KEY CONCEPTS

Risk Factors for Fatal Food-Induced Allergic Reactions

- Peanut and/or tree nut allergy
- Delay in administration of auto-injectable epinephrine
- Preexisting and/or poorly controlled asthma
- Concomitant use of beta-blocker medications
- Teen and young adult age groups

MALT is composed of lymphocytes, antigen-presenting cells (APCs), stromal cells, and other immune cells in the lamina propria. It is within MALT that dendritic cells (DCs) interact with dietary antigens.

There are several common characteristics among the most commonly allergenic foods: (1) relatively small molecular weight, generally less than 70 kilodaltons (kDa); (2) an abundant source of the relevant allergen; (3) glycosylation residues; (4) water solubility; and (5) resistance to heat and digestion. These characteristics allow the proteins to stay intact until reaching the small intestine, wherein they initiate a T-helper-cell type 2 (Th2) response that results in production of specific IgE and eventual allergic disease. *Glycosylation* refers to the reaction by which carbohydrates are attached to molecules; in food allergens, the carbohydrate is most often attached to a protein. Carbohydrate residues surrounding proteins may be important in initiating the immune response. For example, interaction with DC-specific intercellular adhesion molecule–grabbing nonintegrin (DC-SIGN), a c-type lectin expressed on APCs that identifies conserved carbohydrate residues, has been shown to mediate recognition of the major peanut protein, Ara h 1. This interaction allows DC activation and Th2 skewing of naïve human T cells.

Once the Th2 response is initiated, it is strengthened through the induction of interleukin-4 (IL-4) signaling. IL-4 signals B cells to undergo class-switch recombination and begin producing IgE. Basophils have been implicated as a likely contributor of early IL-4 production and may play an important role in priming the T-cell response to allergens.

The Allergic Response

Allergenic food proteins that survive the initial stages of digestion are taken up by the APCs in MALT. Mucosal DCs may encounter antigen through (1) extending dendrites through the paracellular space between epithelial cells to sample luminal contents; (2) directly interacting with the epithelial cells; and (3) taking up antigen in the Peyer patch. Once contact with the antigen is established, the antigen is processed and loaded onto major histocompatibility complex (MHC) class II molecules on the cell surface, costimulatory molecules necessary for T-cell activation are upregulated, and chemotaxis to the draining lymph node occurs. Once a DC encounters a T-cell receptor with the same specificity as the peptide antigen, an immune response ensues. In the presence of cytokines, such as IL-4, IL-5, and IL-13, the responding T cell is programmed to be a Th2 cell. The Th2 cell will then signal B cells to generate IgE antibodies.

Soluble IgE that is produced by B cells circulates and binds to the surface of mast cells and basophils. Mast cells are found in skin, the gut, and the respiratory tract and are located adjacent to nerves and blood vessels. When an allergen is encountered and recognized by cell-bound IgE, calcium influx ensues, activating the mast cell. Once activated, the mast cell degranulates and releases vasoactive compounds and proteases, including histamine, platelet-activating factor, tryptase, chymase, carboxypeptidase, and heparin, resulting in the characteristic symptoms of an allergic reaction: urticaria, angioedema, flushing, nausea, vomiting, abdominal pain, diarrhea, wheezing, coughing/bronchospasm, rhinorrhea, and hypotension/syncope. Symptoms may occur alone or in combination and typically appear within minutes of ingestion.

Natural History

The majority of food allergies are outgrown. Allergies to peanuts, tree nuts, fish, and shellfish are more likely to persist.[13] Clinical characteristics and laboratory measures may help predict which food allergies will be outgrown and which are more likely to be lifelong.

CLINICAL PEARLS

Important Diagnostic Considerations

- The patient's history should support a diagnosis of immunoglobulin E (IgE)–mediated food allergy before performing either serum-specific IgE testing or the skin prick test.
- 95% predictive probability cutoffs have been established for only a few foods, including cow's milk, hen's egg, and peanut.
- Even with negative specific IgE serum testing or skin testing, the patient could be allergic if he or she has a convincing history. In this case a physician-supervised oral food challenge (OFC) may be required to confirm the presence or absence of an IgE-mediated food allergy.
- Testing for food allergies should be limited to the food(s) in question, since positive IgE tests are not always clinically relevant. Unnecessary avoidance of foods may lead to nutritional deficiencies.

Cow's milk is typically one of the first foods introduced to infants in the form of infant formula; it is present in diets across cultural groups and is one of the most common allergens globally. Fortunately, cow's milk allergy is typically outgrown without intervention. Studies of natural resolution vary but about 50% of children with a milk allergy develop tolerance between 5 and 10 years of age.[13] High levels of milk-specific IgE generally indicate a higher likelihood of persistent disease, but as many as 60% of children whose milk-specific IgE level peaks at over 50 kU/L will achieve natural tolerance by age 18 years. Up to 75% of children with reactions to uncooked milk can tolerate baked milk products. Consumption of heated milk products has been associated with accelerated acquisition of tolerance.[14]

Hen's egg allergy is another common food allergen across cultural groups. Most of the allergenic proteins in hen's eggs are in the egg white. Allergy typically develops in the first year of life, while in some children, especially those with atopic dermatitis, it may develop before 4 months of age.[13] Similar to cow's milk allergy, egg allergy usually resolves during childhood without intervention. Roughly 50% of individuals with egg allergy in infancy develop natural tolerance between 6 and 9 years of age. Baked egg is tolerated by approximately 70% of children with egg allergy.[15] Individuals who can tolerate baked egg are likely to develop tolerance to lightly cooked egg sooner compared with individuals who are unable to tolerate baked egg.[16]

Although the majority of individuals who are allergic to peanuts will remain reactive throughout life, approximately 20% of individuals with peanut allergy may develop natural tolerance. Favorable prognostic factors include low levels of peanut-specific IgE antibodies in the first 2 years of life and decreasing levels of IgE sensitization by 3 years. Those with peanut-specific IgE $\geq$ 3 kU/L and skin prick test (SPT) wheal diameter >6 mm before 2 years of age are more likely to have persistent a peanut allergy.[17]

Allergy persistence, regardless of the food allergen, has been associated with the following factors: (1) earlier age at diagnosis; (2) concomitant presence of other allergic diseases (e.g., allergic rhinitis, asthma, and eczema); (3) severity of those allergic diseases; (4) symptom severity after ingestion; and (5) lower threshold dose required to elicit a reaction. The higher the food-specific IgE level, the more likely it is that the allergy will persist. In clinical practice, food-specific IgE levels are typically checked yearly except in patients whose specific IgE levels remain high and unchanged over several years.

DIAGNOSIS

The diagnosis of food allergy begins with obtaining a detailed medical history.[6] Food-induced allergic reactions result in reproducible characteristic symptoms, as described above. Validated testing modalities are only available for IgE-mediated food allergies and celiac disease. If the clinical history does not support either diagnosis, then food allergy serum or skin testing should not be done, since there is a risk of finding sensitization to allergens that are not clinically relevant and multiple studies have shown the dangers of unnecessary dietary avoidance. When the clinical history does support a diagnosis of food allergy, this can be confirmed by SPTs and detection of specific IgE in serum.

SPTs to food allergens can be performed in the office setting and are both safe and effective, with results being available within minutes. A positive result of the SPT reflects the presence of specific IgE bound to the surface of cutaneous mast cells, but as with serum IgE testing, a positive test result does not always indicate clinical reactivity. A positive test result is generally interpreted as 3 mm larger than the negative SPT control, and the larger the SPT mean wheal diameter, the more likely it is indicative of a clinically relevant response. Negative SPT results have been associated with a high negative predictive value and may lead the physician either to offer an observed OFC or to counsel the patient on dietary reintroduction, depending on the clinical history and circumstances.

Serum-specific IgE testing is useful in providing an objective measure of food-specific IgE antibody, especially if the patient cannot stop antihistamine therapy or has extensive skin disease making it impossible to perform the SPT, and serum IgE testing may be helpful in counseling patients on the natural history of their food allergy. Predictive values have been established for a limited number of foods. Higher specific IgE levels are more likely to be associated with clinical reactivity, but the predictive value of specific IgE levels varies across patient populations and is affected by such factors as the patient's age, ethnicity, and time since last ingestion of allergen. Specific IgE levels may also help physicians decide when an OFC is or is not appropriate.

The component-resolved diagnostic (CRD) test uses allergenic proteins derived from recombinant DNA technology or purification from natural sources to identify the patient's specific IgE reactivity to individual allergenic proteins rather than to the whole allergen. Diagnostic accuracy can be enhanced in specific circumstances (*e.g.*, for peanuts and hazelnuts).[18] However, CRD is not routinely used for diagnosis and has not been shown to provide significant additional clinical information for most allergens. CRD for peanuts and hazelnuts provides additional diagnostic information that is helpful to the clinician, but standardized decision-making cutoffs have not yet been established.

The basophil activation test (BAT) uses flow cytometry to detect upregulation of cell-surface molecules, such as CD63 and CD203c, after stimulation with allergen.[19] BAT has been reported to be superior to the SPT, the CRD test, and the whole-allergen-specific IgE test for diagnosis of peanut allergy; however, testing has not been standardized. Further research is needed to standardize the BAT and validate results with various food allergens.

The OFC remains the gold standard for the diagnosis of food allergy.[20] OFCs can be conducted in an open manner; with a placebo control, where the patient is blinded to the product

being given; or in a double-blinded manner, with both the physician and patient blinded to the food being given to the patient. An open OFC is most commonly performed in clinical practice, while the double-blind, placebo-controlled food challenge (DBPCFC) is considered the diagnostic standard typically reserved for research studies. During the OFC, a standard serving size of the allergen is divided into 4 to 7 servings and administered over 60 to 90 minutes, with each dose being given 15 to 20 minutes apart. The initial amount fed to the patient is typically a very small proportion of the total serving, and each successive dose administers a larger amount of protein. At the first sign of an objective reaction, the OFC is stopped and appropriate treatment administered. In cases where anxiety or subjective symptoms may affect the interpretability of the OFC, a single-blind or DBPCFC may be preferred.

MANAGEMENT

The patient with a food allergy must maintain strict avoidance of the food allergen to prevent an allergic reaction. Accidental ingestion is common, with reports showing that as many as 50% of children with a peanut allergy may experience an adverse reaction in a 2-year period, while up to 75% have this experience over 10 years.[6] Individuals with food allergies and their caregivers must read ingredient labels carefully, prevent cross-contact, communicate with restaurant staff when eating outside of the home, and be prepared to treat a reaction, when necessary.

Food allergy labeling laws in the United States require that the presence of the most common allergens (milk, egg, peanut, tree nuts, wheat, soy, fish, and crustacean shellfish) must be declared in simple English on the ingredient labels of all packaged foods. Individuals allergic to foods other than the eight most common allergens may have more difficulty with interpretation of ingredient labels. Ingredient labels may report "spices" or "natural flavors," which could include a multitude of foods or food products not covered by food allergy labeling laws. Statements such as "may contain [allergen]" and "manufactured on shared equipment with [allergen]" are voluntary and not regulated. Allergen content in such products is unknown, and it is typically recommended that individuals with allergies avoid products with "may contain" labeling.

Children with milk allergy or with two or more food allergies have been shown to be at particular risk of growth deficiency. Nutritional counseling with a registered dietitian is encouraged for these patients. A registered dietitian will help educate the patient and his or her family on avoidance of food allergens, in addition to providing guidance on nutrient supplementation to avoid potential dietary deficiencies.

TREATMENT OF A REACTION

An acute reaction must be recognized and treated expeditiously. Food-induced fatalities are most commonly reported from ingestion of peanuts and tree nuts, but any food allergen can induce a severe reaction. Fatalities have been associated with delay in administration of auto-injectable epinephrine, preexisting and/or poorly controlled asthma, and concomitant use of beta-blocker medications; there is increased mortality in teen and young adult age groups. Intramuscular auto-injectable epinephrine must be readily available to patients with IgE-mediated food allergies and is the first-line treatment for a food-induced allergic reaction. Patients with food allergies are encouraged to have a written emergency action plan that lists the signs and symptoms of an allergic reaction and details treatment of those symptoms.

PREVENTION OF FOOD ALLERGY

Exposure to antigen early in life is likely important for shaping the appropriate immune response to foods. Primary exposure through the oral route is believed to predispose patients to the development of a tolerogenic response, whereas primary exposure through skin may result in sensitization.[21] Support for the theory of prevention through primary oral exposure has been strongly supported through epidemiological studies showing that some cultural groups that introduce peanuts to their children in the first year of life have a lower incidence of peanut allergy.[22] This theory has been strengthened by evidence demonstrating that children identified as being at high risk of developing peanut allergy (severe atopic dermatitis and/or egg allergy) are substantially protected against the development of peanut allergy if they regularly ingest peanuts from between 4 and 11 months of age through 60 months of age, compared with matched controls who avoid peanuts.[23] Based on the results of this study, it is globally recommended to introduce peanut in an infant-safe form into the diet of high-risk infants (as defined above) between 4 and 11 months of age, and both Australian and British allergy societies also recommend early introduction of egg into the diet of high-risk infants to prevent the development of egg allergy.[23] Data to support introduction of other allergens into the diet to prevent development of those food allergies is lacking; however, there is no evidence that suggests delayed introduction of allergenic solids is beneficial. Ongoing studies may reveal whether reduction in the risk of AD development may prevent food allergy development.

The role of the microbiome's contribution to food allergy development is an intense area of study. Microbial products in the gut flora interact with innate immune receptors, such as Toll-like receptors (TLRs) and relay signals implicated in the activation of regulatory T cells (Tregs), which are important in the promotion of tolerance. Activation of a specific TLR using nonpathogenic bacteria (probiotics) could conceivably prevent allergic disease. Unfortunately, data on probiotic supplementation for food allergy is insufficient at this time to offer any specific recommendations.[24]

EXPERIMENTAL INTERVENTIONAL THERAPIES

The standard of care for IgE-mediated food allergy is avoidance of the potentially triggering allergen, treatment of a reaction with autoinjectable epinephrine, and dietary supplementation of potentially deficient nutrients in the diet of the patient with a food allergy. Allergen-specific immunotherapies are currently under investigation utilizing the oral, sublingual, and epicutaneous routes for the application of the allergen. The US Food and Drug Administration (FDA) has approved a regulated oral immunotherapy product for peanut allergy at the time of this writing, and it is anticipated that more FDA-approved products will soon follow.

Oral Immunotherapy

Oral immunotherapy (OIT) is accomplished by mixing the allergenic food into a vehicle food, initially giving doses below the level that would trigger reactions in an allergic individual and

gradually increasing the amount of protein ingested over time. The buildup phase of therapy typically lasts several months; once a maintenance dose of allergen is achieved, the patient must ingest the allergen for a certain period (typically ≥1 year, possibly indefinitely) to maintain a protected, desensitized state. Most studies have focused on achievement of desensitization, which refers to a temporary increase in the threshold of allergen required to elicit a reaction and is dependent on regular exposure to the allergen.

OIT will induce significant desensitization in most patients who are able to tolerate therapy. The largest OIT trial to date showed that 67.2% of participants by intention-to-treat were desensitized (defined by ability to ingest 600 mg of peanut protein without dose-limiting symptoms) after 6 months of daily ingestion of 300 mg of peanut protein, compared to 4% of placebo-treated subjects.[25] Most individuals undergoing OIT will experience adverse reactions at a greater frequency than patients practicing avoidance, though adverse reactions are typically mild and tolerable. Transient abdominal pain and oral pruritus are the most common problems reported across OIT studies; reactions do not typically require any treatment. Severe reactions, such as anaphylaxis, may develop during therapy; predisposing factors include infection, exercise, and allergen co-exposure. GI symptoms and anaphylaxis are the most common reason for participants withdrawing from OIT trials, and EoE has occasionally been documented. Sustained protection against an allergic reaction, independent of ongoing allergen exposure (sustained unresponsiveness [SU]), has not been adequately measured; only a minority of individuals achieved SU in the few studies measuring this outcome.[25] Further work is needed to determine which patients are most likely to develop SU, who will tolerate OIT with few dose-limiting adverse events, and the mechanisms underlying the development of desensitization and SU.

The initial immune response detected in desensitization includes an increase in food-specific IgG4, decreased basophil and mast cell responsiveness, and an initial increase in allergen-specific IgE. Allergen-specific IgE then decreases gradually over time. After 6 to 12 months of therapy there appears to be a shift away from Th2 cytokine production in response to allergen toward a Th1 profile. Treg upregulation occurs later in the course of OIT, with studies showing increased function of antigen-specific $CD4^+CD25^+$ $FOXP3^+$ Tregs. Epitope mapping typically changes over time indicating different antigen-specific responsiveness. Unfortunately, there are no biomarkers that consistently predict successful desensitization or SU.

Sublingual Immunotherapy

Sublingual immunotherapy (SLIT) utilizes a food protein dissolved in a liquid medium and delivered beneath the tongue. The oral mucosa contains tolerogenic APCs: SLIT is thought to rely on these cells to induce a desensitized state. SLIT dosing utilizes microgram to milligram quantities of protein, whereas OIT protocols utilize gram quantities of protein. Increasing the amount of allergen given is limited by the concentration of available extracts and the volume of liquid that can be held sublingually.

Peanut allergy with SLIT has been more closely studied than SLIT for other food allergies. Peanut SLIT treatment in both children and adults has shown an ability to raise the dose-triggering threshold while on therapy for the majority of participants.[25] The most commonly reported complaint is transient oropharyngeal itching and systemic reactions are rare.

Few studies have compared SLIT with OIT; current evidence indicates that SLIT has fewer side effects compared with OIT but SLIT does not appear to induce a similar level of desensitization or achieve SU as often as OIT; however, SLIT has not been as closely studied as OIT with many fewer participants receiving SLIT in randomized, controlled trials. Additional studies are needed to reveal whether adjuvant therapy with SLIT increases efficacy and to understand whether SLIT could be combined with OIT to improve its safety.

THERAPEUTIC PRINCIPLES

Avoidance

- Read ingredient labels closely. The eight most common food allergens are required to be disclosed on ingredient labels of foods manufactured and sold in the United States.
- Minimize cross-contact with food allergens during meal preparation.
  - Use utensils, cutting boards, and pans that have been thoroughly washed with soap and water.
  - If you are preparing several foods, make the allergy-safe food first.
  - Wash hands with soap and water before touching anything else if you have handled a food allergen.
  - Wash counters and table with soap and water after making meals.
- When eating at a restaurant, inform the waiter and cooking staff about food allergens.
- Avoid buffets.

Treatment

- Advise patients at risk of anaphylaxis to carry two auto-injectable epinephrine devices at all times.
- Recommend a medical identification bracelet.
- Provide an anaphylaxis emergency plan, and review indications for administration of auto-injectable epinephrine.
- Demonstrate the appropriate use of auto-injectable epinephrine with a trainer device at the physician's clinic visits.
- FDA-approval of a regulated peanut oral immunotherapy product may provide a therapeutic option for some patients.
- Actively engage patients and their families when discussing the pros and cons of interventional therapy versus continued avoidance.

Epicutaneous Immunotherapy

Epicutaneous immunotherapy (EPIT) delivers allergen to the skin through daily application of an allergen-containing patch. Langerhans cells in the skin are activated and effector cell responses are downregulated. A phase 3 trial for peanut EPIT reported 35.3% of peanut EPIT subjects responded to 12 months of treatment compared to 13.6% of subjects receiving placebo.[25] Despite the statistically significant difference in groups, the difference in responder rates did not meet a prespecified confidence interval boundary for the study to be considered positive. The most common side effect of EPIT a is an eczematous response at the site of patch application. Eight subjects in the phase 3 trial reported anaphylaxis from the peanut patch. SU has not been closely studied with EPIT use. Surrogate biomarkers do not predict response to therapy and at the present time an OFC is necessary to gauge treatment response.

Diagnosis of food allergy is reliant on an understanding of its pathogenesis and proper application of available diagnostic tools. Management of food allergy requires education about avoidance of the allergenic food(s), dietary supplementation of missing nutrients, and recognition and early treatment of any allergic reactions. Although food allergy has increased in prevalence over the past two decades, preventive strategies, including

ON THE HORIZON

- Oral immunotherapy (OIT) exposes the allergic patient to progressively larger doses of ingested allergen in an effort to induce a desensitized state. Gram quantities of allergen are typically administered. Side-effects, though typically mild, are common and significant treatment burden may prevent wide-spread adaptation.
- Epicutaneous immunotherapy (EPIT) applies microgram amounts of allergen directly to the allergic patient's skin, resulting in an effort to increase the threshold of reactivity. EPIT patch is typically kept on skin for up to 24 h at a time, and a new patch is applied daily. Therapeutic benefit is not as consistent as the benefit reported with OIT; however, the therapy is safe with few severe adverse effects reported.
- Sublingual immunotherapy (SLIT) involves sublingual administration of milligram quantities of allergen solubilized in a liquid formulation. Systemic reactions are rare. However, few studies have rigorously investigated SLIT for treatment of food allergy.

early introduction of allergenic solids, may help curb this as yet unexplained epidemic. As our knowledge of food allergy grows, we can expect significant changes in our approach to the treatment of those affected in the years to come.

REFERENCES

1. Gupta R, Holdford D, Bilaver L, et al. The economic impact of childhood food allergy in the United States. *JAMA Pediatr*. 2013;167:1026–1031.
2. Boyce JA, Assa'ad A, Burks AW, et al. Guidelines for the diagnosis and management of food allergy in the United States: summary of the NIAID-sponsored expert panel report. *J Allergy Clin Immunol*. 2010;126:1105–1118.
3. Warren CM, Jian J, Gupta RS. Epidemiology and burden of food allergy. *Curr Allergy Asthma Rep*. 2020;20:6.
4. Koplin JJ, Mills EN, Allen KJ. Epidemiology of food allergy and food-induced anaphylaxis: is there really a Western world epidemic? *Curr Opin Allergy Clin Immunol*. 2015;15:409–416.
5. Sampson HA, Aceves S, Bock SA, et al. Food allergy: a practice parameter update—2014. *J Allergy Clin Immunol*. 2014;134:1016–1025. e1043.
6. Bird JA, Lack G, Perry TT. Clinical management of food allergy. *J Allergy Clin Immunol Pract*. 2015;3:1–11. quiz 12.
7. Fiocchi A, Brozek J, Schunemann H, et al. World Allergy Organization (WAO) Diagnosis and Rationale for Action against Cow's Milk Allergy (DRACMA) guidelines. *Pediatr Allergy Immunol*. 2010;21:1–125.
8. Sicherer SH. Clinical implications of cross-reactive food allergens. *J Allergy Clin Immunol*. 2001;108:881–890.
9. Dellon ES, Liacouras CA. Advances in clinical management of eosinophilic esophagitis. *Gastroenterology*. 2014;147:1238–1254.
10. Nowak-Wegrzyn A. Food protein-induced enterocolitis syndrome and allergic proctocolitis. *Allergy Asthma Proc*. 2015;36:172–184.
11. Rubio-Tapia A, Hill ID, Kelly CP, et al. ACG clinical guidelines: diagnosis and management of celiac disease. *Am J Gastroenterol*. 2013;108:656–676. quiz 677.
12. Vickery BP, Chin S, Burks AW. Pathophysiology of food allergy. *Pediatr Clin North Am*. 2011;58:363–376. ix–x.
13. Savage J, Sicherer S, Wood R. The natural history of food allergy. *J Allergy Clin Immunol Pract*. 2016;4:196–203. quiz 204.
14. Kim JS, Nowak-Wegrzyn A, Sicherer SH, et al. Dietary baked milk accelerates the resolution of cow's milk allergy in children. *J Allergy Clin Immunol*. 2011;128:125–131. e122.
15. Leonard SA, Caubet JC, Kim JS, et al. Baked milk- and egg-containing diet in the management of milk and egg allergy. *J Allergy Clin Immunol Pract*. 2015;3:13–23. quiz 24.
16. Leonard SA, Sampson HA, Sicherer SH, et al. Dietary baked egg accelerates resolution of egg allergy in children. *J Allergy Clin Immunol*. 2012;130:473–480. e471.
17. Ho MH, Wong WH, Heine RG, et al. Early clinical predictors of remission of peanut allergy in children. *J Allergy Clin Immunol*. 2008;121:731–736.
18. Grabenhenrich L, Lange L, Hartl M, et al. The component-specific to total IgE ratios do not improve peanut and hazelnut allergy diagnoses. *J Allergy Clin Immunol*. 2016;137:1751–1760. e1758.
19. Santos AF, Lack G. Basophil activation test: food challenge in a test tube or specialist research tool? *Clin Transl Allergy*. 2016;6:10.
20. Bird JA, Leonard S, Groetch M, et al. Conducting an oral food challenge: an update to the 2009 Adverse Reactions to Foods Committee Work Group Report. *J Allergy Clin Immunol Pract*. 2020;8:75–90.
21. du Toit G, Tsakok T, Lack S, et al. Prevention of food allergy. *J Allergy Clin Immunol*. 2016;137:998–1010.
22. Du Toit G, Katz Y, Sasieni P, et al. Early consumption of peanuts in infancy is associated with a low prevalence of peanut allergy. *J Allergy Clin Immunol*. 2008;122:984–991.
23. Bird JA, Parrish C, Patel K, et al. Prevention of food allergy: beyond Peanut. *J Allergy Clin Immunol*. 2019;143:545–547.
24. Bunyavanich S, Berin MC. Food allergy and the microbiome: current undesrtandings and future directions. *J Allergy Clin Immunol*. 2019;144:1468–1477.
25. Kim EH, Patel C, Burks AW. Immunotherapy approaches for peanut allergy. *Expert Rev Clin Immunol*. 2020;16:167–174.

50

Drug Hypersensitivity

Shyam R. Joshi, Whitney Salinas and David A. Khan

Drug therapy requires an understanding of the fine line between the beneficial and harmful effects of the drug. While the majority of adverse drug reactions (ADRs) are predictable (type A reactions or "on target"), drug allergies are difficult to foresee, hence their designation as type B (*bizarre* or "off target") ADRs. Any drug is assumed to be capable of eliciting these types of reactions; however, the frequency differs widely, with antibiotics being the most common offender. Multiple factors play a role in the risk and severity of reactions, including the class of drug, dose, administration route, frequency, and duration of exposure, and the genetic predisposition of the subject, particularly with human leukocyte antigen B (HLA-B) alleles.

Drug hypersensitivity reactions can be a significant source of morbidity and mortality in clinical practice. Beyond a thorough clinical history, the tools available for identifying and diagnosing hypersensitivities are currently limited; however, accurate diagnoses are still possible and important to help protect patients from re-exposure to the culprit medication. Drug allergy labels can also often prohibit patients from receiving first-line therapies, and, therefore, previously diagnosed drug allergies should always be questioned.

EPIDEMIOLOGY

ADRs are common and occur in approximately 15% to 25% of patients.[1] This includes predictable pharmacologic side effects and account for 3% to 6% of all hospital admissions.[2] Hypersensitivity reactions, however, are considerably less common and account for only 5% to 10% of all ADRs. The true overall incidence of drug allergy remains unknown, as the available epidemiologic studies typically focus on select populations or specific subtypes of drug allergies. In addition, the heavy reliance on historical information and lack of standardized clinical questionnaires or confirmatory diagnostic testing complicates the interpretation of these studies.

ADRs occur more frequently in females with a 2:1 predominance, although there has been a higher prevalence of acute interstitial nephritis and fixed drug eruptions in males.[3] Overall, ADRs are most common in white races, although there have been select racial associations to certain ADRs: Black people have a higher prevalence of angiotensin converting-enzyme inhibitors (ACE-I) induced angioedema, while Asians have a higher prevalence of fixed drug reactions and severe cutaneous drug reactions (SCARs).[3] In regard to age groups, the incidence is less understood but seems to increase with age, likely due to increased drug exposure; however, elderly hospitalized patients appear to have a lower incidence of anaphylaxis and SCARs.[4]

Approximately half of reactions occur immediately (within 6 hours of last exposure) with the most common reported symptom being urticaria followed by itching and angioedema. Rashes are also the most frequently encountered delayed (>6 hours after exposure) hypersensitivity symptom. Of the delayed reactions, SCARs are very rare, affecting approximately 0.4% of the population, but do account for a significant portion of mortality associated with drug allergies.[3]

While any drug can elicit hypersensitivity reactions, the most frequently implicated are antibiotics, nonsteroidal anti-inflammatory drugs (NSAIDs), antiepileptics, chemotherapeutics, and radiocontrast media (RCM). Penicillin is the most commonly reported of all drug allergies at approximately 10% of the general population.[5]

IMMUNE SYSTEM RECOGNITION OF DRUGS

Drug allergies can present in a multitude of ways, which will be discussed later in this chapter. The initial recognition of a drug as a foreign antigen by the immune system remains an important step in this process. This is dependent on the structure, shape, and multimeric/multivalent presentation of the drug. Four hypotheses have been proposed to explain how drugs interact with HLA and T-cell receptors (TCRs) in the development of hypersensitivity reactions. These include (1) the hapten theory, (2) the direct pharmacologic interaction with immune receptor (p-I) concept, (3) the altered peptide repertoire model, and (4) the altered TCR repertoire model (Fig. 50.1).[6] It should be noted that these models are not mutually exclusive and can occur simultaneously during a drug reaction.

RISK FACTORS FOR DRUG ALLERGY

Drug-Related Factors

The immunogenicity of a drug is based on several factors, with its ability to be recognized by the immune system likely being the most important. The manufacturing process of the drug and the metabolic processes that occur within the body after administration can lead to undesired byproducts that can be highly immunogenic. Drug-specific risk factors for hypersensitivity reactions include higher doses, parenteral administration, intermittent and repeated doses compared to uninterrupted treatment, and concurrent illness.

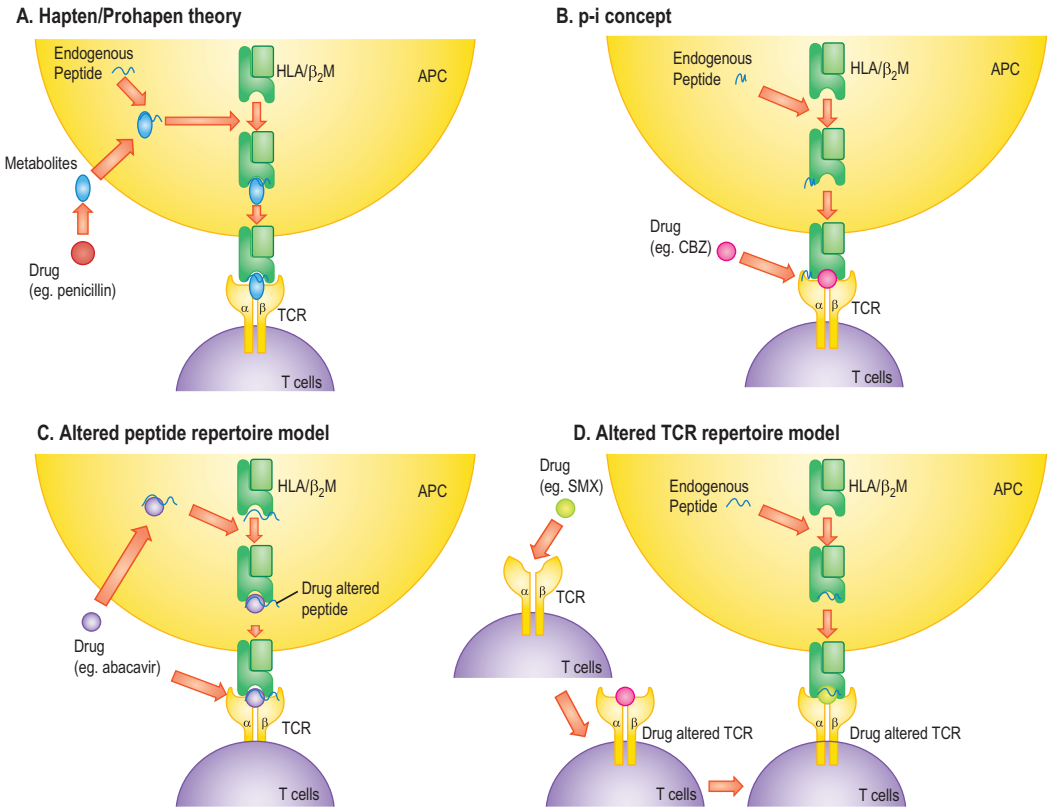


FIG. 50.1 Four proposed hypotheses on how drugs interact with human leukocyte antigen *(HLA)* and T-cell receptor *(TCR)* in the development of hypersensitivity reactions. (A) The hapten theory: drugs or their metabolites bind to endogenous proteins as haptens. This forms the HLA-peptide-drug complex in the antigen-presenting cell (APC) which is presented to and recognized by TCR leading to drug-specific T-cell activation. (B) The p-I concept: drugs directly bind to the HLA-peptide complex or TCR without intracellular processing in the APC. (C) The altered peptide repertoire model: small molecules directly bind to the MHC peptide-binding groove, altering the specificity of the binding cleft and allowing a novel presentation of the peptide. (D) Altered TCR repertoire model: drugs can bind directly to the TCR, allowing it to recognize HLA-self peptide complexes. (Adapted from Chung W, Wang C, Dao R. Severe cutaneous adverse drug reactions. *J Dermatol.* 2016;43:758–766.)

Host Related Factors

Women have consistently been shown to report drug allergies more frequently than men, but this propensity has not been thoroughly investigated.[3] The risk of drug allergy is lower in children, but this may be due to their infrequent exposure to drugs. Conversely, children with chronic medical conditions, such as cystic fibrosis, have a significantly higher rate of reported drug allergies to antibiotics, which has been suggested to be caused by their repeated exposure to these medications.[7–12]

Underlying disease states may increase the risk of drug hypersensitivity reactions with human immunodeficiency virus (HIV) being a prominent example. These patients are at particularly high risk for reactions to sulfonamide antibiotics, abacavir, and nevirapine, which appears to be related to the degree of immunodeficiency.[2, 13]

About 25% of women who are treated with carboplatin for gynecological cancers become sensitized after multiple exposures and then develop type I immunoglobulin E (IgE)–mediated reactions, most of which are anaphylactic. Since women with BRCA1 and BRCA2 mutations seem more prone to carboplatin reactions after fewer exposures, it is possible that these genetic mutations facilitate T-cell presentation.[14]

Heterologous responses to viral infections, in the context of specific HLA alleles, have been implicated in severe skin reactions. The induction of this type of drug hypersensitivity requires an HLA risk allele, a primary infection with human (HHV) or other viruses in the context of that HLA risk allele, polyclonal expansion of CD8 T cells, and the induction of memory T cells. Upon exposure to the drug, the interaction with HLA induces neoantigens, and the peptide–MHC complex is recognized by the TCR, triggering activation of T cells (Fig. 50.2).

Atopy does not appear to be a major risk factor for most drug allergies. The presence of autoimmunity as a risk factor remains unclear and additional studies are needed to further explore this association.

Genetics of Drug Allergy

Genetic studies have mainly focused on the area of HLA genotypes and their association with severe drug hypersensitivity. HLA molecules function as antigen presenters to the TCR and are categorized as HLA Class I (HLA A, HLA B, HLA C) or HLA Class II (HLA DP, HLA DQ, HLA DR). Specific HLA types have been associated with SCARs, which are outlined in Table 50.1.

The best examples include abacavir hypersensitivity syndrome and its association with HLA-B*57:01, carbamazepine-induced SJS/TEN associated with HLA-B*15:02, and allopurinol-induced DRESS/HSS/SJS/TEN associated with HLA-B*58:01. Routine HLA B*57:01 screening is now performed prior to initiation of abacavir therapy.[15] Recently, HLA-A*32:01 has been identified as a risk allele for vancomycin-induced DRESS.[16] HLA screening

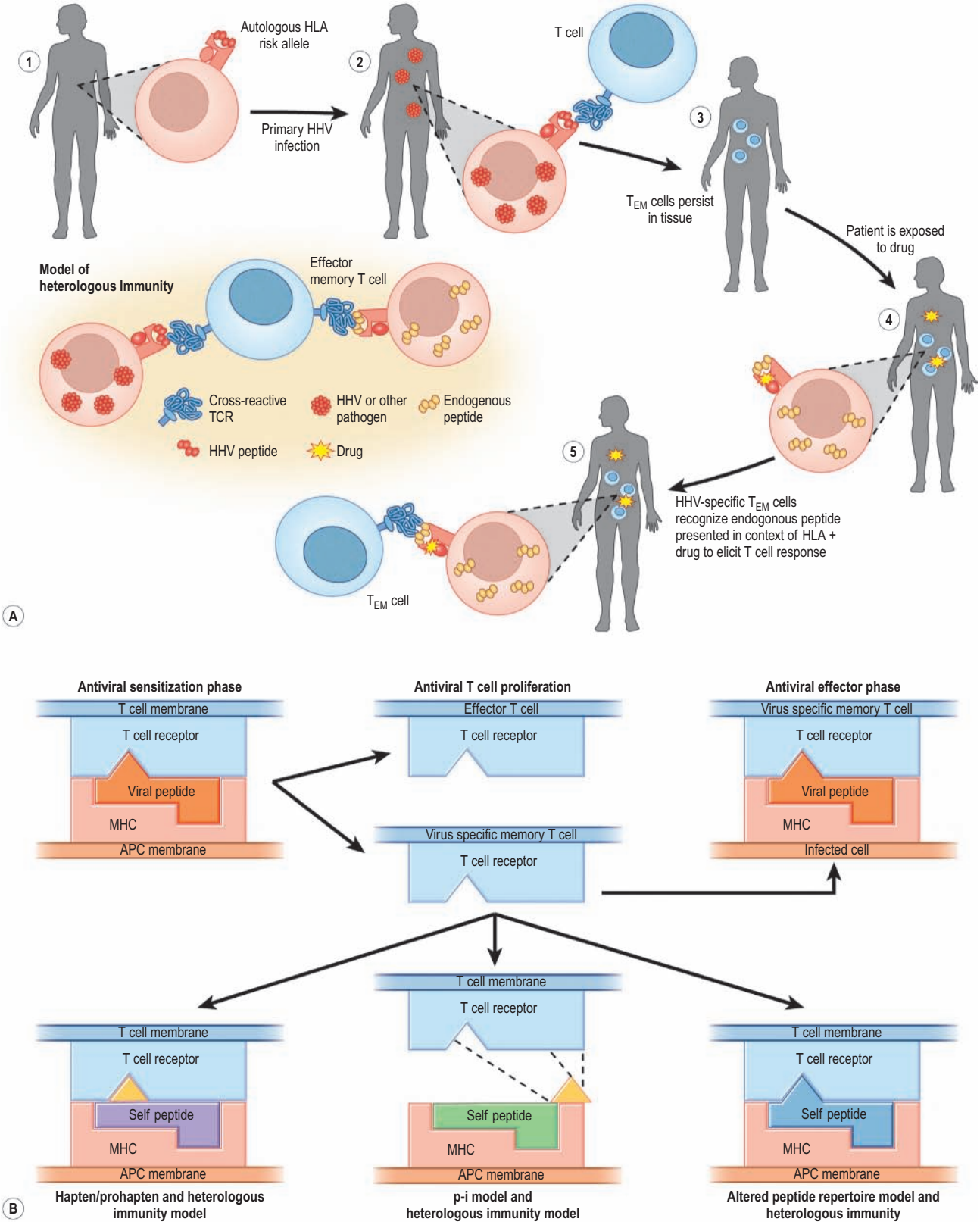


FIG. 50.2 Generation of heterologous immune responses that contribute to the pathogenesis of T-cell mediated adverse drug reactions (ADRs). (A) Timeline of the generation of ADRs. (B) Integration of the models of T-cell activation by small molecules and heterologous immunity. *APC*, Antigen-presenting cell; *HHV*, human herpesvirus; *HLA*, human leukocyte antigen; *MHC*, major histocompatibility complex; *TCR*, T-cell receptor. (From White KD, Chung W-H, Hung S-I, *et al*. Evolving models of the immunopathogenesis of T-cell mediated drug allergy: the role of host, pathogens, and drug response. *J Allergy Clin Immunol*. 2015;136:219–234.)

TABLE 50.1 HLA-Associated Drug Hypersensitivity Reactions

Drug	Adverse Drug Reaction	Associated HLA Alleles	Positive Predictive Value	Negative Predictive Value	Populations
Abacavir	HSS	B*57:01	55%	100%	European, African
Carbamazepoine	SJS/TEN	B*15:02	3%	100%	Han Chinese, Thai, Malaysian, Indian
					Korean, Japanese
					Japanese, Northern European, Korean
	DRESS	8.1 AH (HLA A*01:01, Cw*07:01, B*08:01, DRB1*03:01, DQA1*05:01, DQB1*02:01)			White
		A*31:01	0.89%	99.98%	European
		A*31:01	0.59%	99.97%	Chinese
		A*31:01			Northern European, Japanese, Korean
		A*11 and B*51 (weak)			Japanese
	MPE	A*31:01	34.9%	96.7%	
Allopurinol	SJS/TEN, DRESS	B*58:01 (or B*58 haplotype)	3%	100% in Han Chinese	Han Chinese, Thai, European, Italian, Korean
Oxcarbazepine	SJS/TEN	B*15:02 and B*15:18			Han Chinese, Taiwanese
Lamotrigine	SJS/TEN	B*15:02 (positive), B*15:02 (no association)			Han Chinese
Phenytoin	SJS/TEN	B*15:02 (weak), Cw*08:01, DRB1 * 16:02, CYP2C9*3			Han Chinese
	DRESS/MPE	B*13:01 (weak), B*5101 (weak) CYP2C9*3			Han Chinese
Nevirapine	SJS/TEN	C*04:01			Malawian
	DRESS	DRB*01:01 and DRB*01:02 (hepatitis and low CD41)	18%	96%	Australian, European, and South African
		Cw*8 or Cw*8-B*14 haplotype Cw*4			Italian and Japanese
					Black, Asian, White, and Han Chinese
		B*35, B*35:01, B*35:05	16%	97%	Asian
	Delayed rash	DRB1*01			French
		Cw*04			African, Asian, European, and Thai
		B*35:05, rs1576*G			Thai
		CCHCR1 status			
Dapsone	HSS	B*13:01	7.8%	99.8%	
Efavirenz	Delayed rash	DRB1*01			French
Sulfamethoxazole	SJS/TEN	B*38			European
Amoxicillin-clavulanate	DILI	DRB1*15:01, DRB107 (protective), A*02:01, DQB1*06:02, and rs3135388, a tag SNP of DRB1*15:01-DQB1*06:02			European
Lumiracoxib	DILI	DRB1*15:01-DQB1*06:02-DRB5*01:01-DQA1*01:02 haplotype			International, multicenter
Ximelagatran	DILI	DRB1*07 and DQA1*02			Swedish
Diclofenac	DILI	HLA-B11, C-24T, UGT2B7*2, IL-4 C-590-A			European
Flucloxacilin	DILI	B*57:01, DRB1*01:07-DQB1*01:03	0.12%	99.99%	European
Lapatinib	DILI	DRB1*07:01-DQA2*02:01-DQA1*02:01			International, multicenter

DILI, Drug-induced liver injury; *DRESS*, drug-induced reactions with eosinophilia and systemic symptoms; *HSS*, hypersensitivity syndrome; *MPE*, maculopapular eruption; *SJS*, Stevens-Johnson syndrome; *SNP*, single nucleotide polymorphism; *TEN*, toxic epidermal necrolysis.
From White KD, Chung W-H, Hung S-I, *et al.* Evolving models of the immunopathogenesis of T cell-mediated drug allergy: the role of host, pathogens, and drug response. *J Allergy Clin Immunol.* 2015; 136: 219–234.

will likely become more common for the prevention of hypersensitivity reactions in at-risk populations.

There are also non-HLA-related genetic associations that have been studied including single-nucleotide polymorphisms (SNPs) in IL-13 and the α chain of IL-4 that has been linked to immediate-type hypersensitivity reaction to β-lactams.[17] Other drugs associated with pharmacogenetic associations for immediate reactions include aspirin, NSAIDs, asparaginase, and infliximab.

HYPERSENSITIVITY REACTIONS

Historically, drug reactions were classified based on their immunologic mechanism, as defined by Gell and Coombs in 1968, into types I-IV, the last of which has been further subdivided into type IV a–d (Table 50.2). This classification system is useful; but unfortunately, not all drug reactions can be neatly categorized into a specific type. Examples of well-categorized hypersensitivities are demonstrated in Fig. 50.3.

Type I IgE-Mediated Hypersensitivity

Type I hypersensitivity reactions are classically referred to as IgE-mediated reactions; however, non-IgE-mediated mechanisms have also been described such as IgG-related activation of mast cells and basophils as well as direct mast cell activation, the latter of which is discussed later in this chapter.

TABLE 50.2 Immunologic mechanisms for drug hypersensitivity reactions type I-IV

Classification	Type I	Type II	Type III	Type IVa	Type IVb	Type IVc	Type IVd
Common name	IgE mediated	Antibody dependent	Immune-Complex	Delayed-type hypersensitivity, cell-mediated, antibody-independent			
Onset time	Minutes	Hours to days	Hours to days	Days to weeks			
Immune reactant	IgE	IgG	IgG	CD4 Th1	CD4 Th2	CTL	CD4 Th17
Effector	Mast cell/basophil activation	Phagocytes and NK cells	FcR^+ cells complement	Macrophage activation	Eosinophils	T cells	Neutrophils
Picture							
Example	Anaphylaxis, urticaria	Hemolytic anemia, thrombocytopenia	Serum sickness, arthus reaction, Drug fever, vasculitis	Tuberculin reaction, sarcoidosis	DRESS, maculopapular exanthema	SJS/TEN	AGEP

AGEP, Acute generalized exanthematous pustulosis; *IgE*, Immunoglobulin E; *NK*, natural killer; *SJS*, Stevens-Johnson syndrome; *TEN*, toxic epidermal necrolysis.

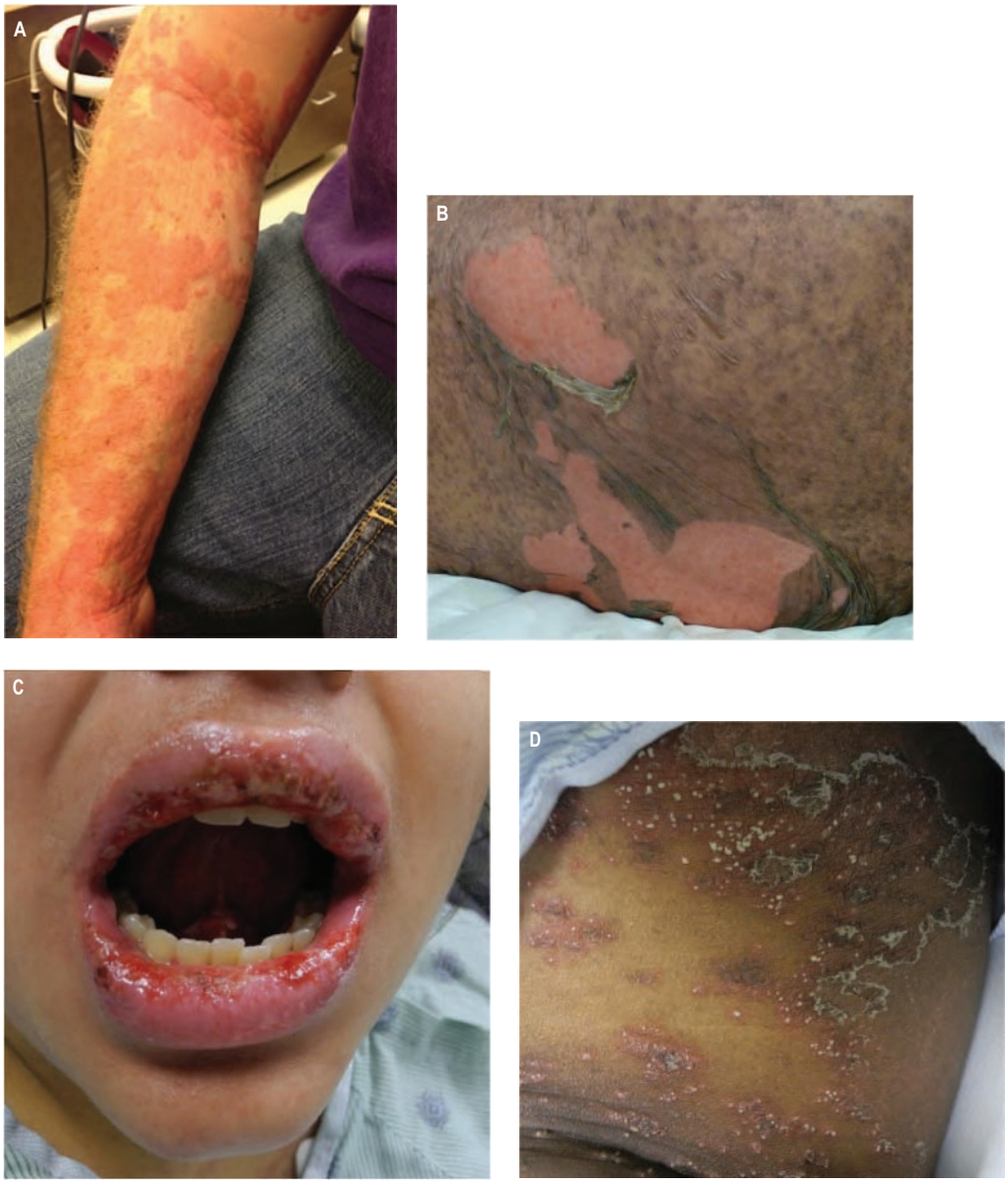


FIG. 50.3 Examples of hypersensitivity reactions: (A) Urticaria seen with type I hypersensitivity. (B) Epidermal detachment in toxic epidermal necrolysis (TEN) seen with type IVc hypersensitivity. (C) Mucousal involvement in Stevens-Johnson syndrome/TEN seen with type IVc hypersensitivity. (D) Acute generalized exanthematous pustulosis seen in type IVd hypersensitivity.

For IgE-mediated reactions, the sensitization phase is thought to occur mainly through the hapten mechanism, resulting proliferation and differentiation of drug-specific B cells into plasma cells, and ultimately drug-specific IgE production. The sensitization process takes several days, which is why patients typically do not react to drugs on initial exposure. Upon re-exposure to the offending drug, polyvalent haptenated allergens crosslink the high affinity IgE receptor, FcεRIα, on mucosal and connective tissue mast cells and blood basophils that have drug-specific IgE bound on their surface. This results in cell activation and degranulation with the release of preformed mediators such as histamine, tumor necrosis factor-α (TNF-α), and tryptase. Within minutes, membrane arachidonic acid releases leukotrienes (LTs) and prostaglandins (PGs), and within hours, newly generated cytokines such as IL-4 and IL-13 are also released to augment the Th2 response.

Symptoms of type I mediated reactions, such as urticaria, cardiovascular collapse, and bronchospasm, are directly correlated to the release of these vasoactive mediators. Reactions can be limited to one system, such as the skin, or can involve multiple systems. Despite a lack of consensus for the diagnostic criteria for anaphylaxis, the most commonly used definition is involvement of two or more organ systems. The World Allergy Organization's grading system for systemic adverse reactions highlights common symptoms seen in type I hypersensitivity reactions. The presence of urticaria is useful in detecting IgE-mediated hypersensitivity, as the wheal and flare are a direct result of mast cell degranulation; although, the lack of urticaria does not rule out an IgE-mediated allergy.

KEY CONCEPT

World Allergy Organization Systemic Allergy Reaction Grading System for Anaphylactic Reactions.

Grade	Description
1	Symptoms of 1 organ system involved: • Cutaneous: pruritus, urticaria, tingling/itching of lips, angioedema (not laryngeal) OR • Upper respiratory: nasal symptoms, throat clearing, cough (not bronchospasm)
2	Symptoms from ≥2 organ systems listed in grade 1
3	Symptoms involving ≥2 organ systems including: • Lower airway: mild bronchospasm (cough, wheeze, or shortness of breath that responds to treatment) OR • Gastrointestinal/Genitourinary: abdominal/uterine cramps, vomiting, or diarrhea • Any symptoms from grade 1
4	Symptoms involving ≥2 organ systems including: • Lower airway: Severe bronchospasm (not responsive to treatment) • Upper airway: laryngeal edema with stridor • Any symptoms from grade 1 or 3
5	Symptoms involving ≥2 organ systems including: • Lower or upper airway: respiratory failure • Cardiovascular: hypotension, loss of consciousness (vasovagal excluded) • Any symptoms from grades 1, 3, or 4

From Cox LS, Sanchez-Borges M, Lockey RF. World Allergy organization systemic allergic reaction grading system: is a modification needed?. *J Allergy Clin Immunol Pract.* 2017;5(1):58–62.

Type I hypersensitivity is generally considered an immediate-type reaction with symptoms often occurring within seconds of parenteral exposure or within minutes of oral administration. Symptoms generally occur upon the first dose after sensitization, and not days into the course, unless there are breaks in the treatment course. Symptoms occurring ≥6 hours after drug exposure are likely not IgE-mediated, although very rarely IgE-mediated reactions can occur up to 24 hours after exposure.[18] Elevation of serum inflammatory markers and fever are not commonly seen in IgE-mediated reactions and alternative causes should be considered.[19] Common implicated drugs include β-lactam antibiotics, neuromuscular blocking agents (NMBAs), platinum based chemotherapeutics, and chimeric and non-chimeric monoclonal antibodies (mAbs). The diagnosis and management of type I reactions are discussed later in this chapter.

Type II Antibody-Dependent Hypersensitivity

Type II responses require the development of specific IgG (specifically IgG1 and IgG3 subclasses) or IgM in response to a drug hapten-carrier complex. These antibodies bind directly to Fcγ receptors on macrophages, natural killer (NK) cells, platelets, or granulocytes. The antibody specificity is directed against antigens on the cell membranes of erythrocytes, leukocytes, platelets, or other cell membranes or tissues that are either destroyed or sequestered within the liver and spleen via complement fixation. Typically type II, in addition to type III, reactions are in the setting of high-dose and prolonged drug administration (Table 50.3).

Heparin-induced thrombocytopenia (HIT) is a prime example of a type II reaction and occurs when IgG and IgM immune complexes form with heparin and platelet factor 4 (PF4), leading to platelet activation through the FcγRIIa receptor. Activated platelets then release PF4 that causes platelet destruction and thrombocytopenia. This typically occurs 5-14 days into the treatment course and is more frequently encountered with unfractionated heparin compared to low molecular weight heparin.

Type III Immune-Complex Hypersensitivity

Immune complexes are created in the presence of drugs that form hapten-carrier complexes, which then bind to endothelial cells, leading to complement activation in small blood vessels, joints, or renal glomeruli. Typically, these complexes appear 4 to 10 days after initial exposure and interact with drug antigens,

TABLE 50.3 Type II Hypersensitivity Reaction Presentation and Common Culprits

Condition	Presentation	Common Drug Culprits
Drug-induced hemolytic anemia	Pallor, fatigue, dyspnea, dark urine, splenomegaly, signs of hyperdynamic state	Penicillins, cephalosporins, NSAIDs, quinine-quinidine
Drug-induced thrombocytopenia	Petechial rash, epistaxis, bleeding gums, hepatosplenomegaly	Heparin, abciximab, quinine-quinidine, sulfonamides, vancomycin, gold compounds, penicillins, cephalosporins, carbamazepinee, NSAIDs
Drug-induced neutropenia	Symptoms of infection including fever, sepsis, pneumonia, pharyngitis, stomatitis	Propylthiouracil, flecainide

NSAIDs, Nonsteroidal anti-inflammatory drugs.

TABLE 50.4 Type III Hypersensitivity Reaction Presentation and Common Culprits

Condition	Presentation	Common Culprits
Serum sickness	Fever, urticarial or pruritic rash, arthralgias, lymphadenopathy, and/or acute glomerulonephritis.	Penicillins, cefaclor, trimethoprim-sulfamethoxazole
Vasculitis	Fever, petechiae/palpable purpura, myalgias, arthralgias, lymphadenopathy. GI tract and kidneys can be involved.	Penicillins, cephalosporins, sulfonamides including diuretics, phenytoin, allopurinol
Drug-induced lupus erythematosus	Fever, myalgias, arthralgias, rash, serositis. Liver and kidney can be involved.	Procainamide, hydralazine, minocyline
Arthus reaction	Painful local swelling and erythema to site of injection	Tetanus, diptheria, hepatitis B vaccines

forming circulating immune complexes. Symptoms typically develop one to two weeks after drug introduction. The clinical consequences of immune complex hypersensitivity can affect the GI tract, kidneys, joints, and skin.

Serum sickness (SS) was first described in the early 1900s when heterologous serum was used in the treatment of diphtheria and scarlet fever as an anti-toxin (Table 50.4).

Antigen-antibody complexes form and, if not cleared efficiently, deposit most commonly in parenchymal tissues and synovial joint fluid. This process activates the classical complement pathway, triggering histamine release and an increase in vascular permeability. Inflammatory cells infiltrate the tissue causing various types of rashes (sparing the mucous membranes), arthralgias, arthritis, lymphadenopathy, fevers, and nephropathy. Presently, SS is most often seen in transplant patients receiving anti-thymocyte globulin or in individuals receiving mAb therapy such as rituximab.

Serum sickness-like reactions (SSLR) are much more common than SS. Antibiotics are a leading cause of SSLR, but many other drugs can cause these reactions, including several biologics. While often characterized as type III reactions due to the clinical similarities with SS, immune complexes have not been identified in SSLR. The triad of fever, rash, and arthralgias is common in SSLR as it is in SS; however, renal involvement and hypocomplementemia are typically absent.

Type IV Delayed-Type Hypersensitivity

Unlike type I-III reactions that are antibody-mediated, type IV reactions are dependent on the activation and expansion of T cells. These reactions are subdivided into subtypes a-d based on the cytokines produced and effector cells involved. Given that these reactions are cell-mediated, the timing of symptom onset can be days to weeks after drug exposure. If a patient is re-exposed after an initial reaction, symptoms may appear quickly, possibly within 24 hours. The skin houses a large number of T cells that are primed memory-effector cells. These cells respond rapidly to immunogenic agents, resulting in skin rashes, which is the most commonly encountered manifestation for this type of hypersensitivity reaction.[3]

Type IVa Reactions

Type IVa reactions are characterized by T helper 1 (Th1) cells stimulating macrophages through the release of interferon-γ and TNF-α. Activated macrophages can cause local cutaneous-only or systemic inflammatory responses. Examples of type IVa reactions include tuberculin reaction, allergic contact dermatitis, and sarcoidosis.

Type IVb Reactions

Type IVb reactions, on the other hand, are associated with type 2 inflammation. Eosinophils, mast cells, and IgE- and IgG4-producing B cells are promoted through Th2 release of IL-4, IL-5, IL-13, and eotaxin.

Maculopapular exanthema is the most common delayed-type hypersensitivity reaction causing a benign moribilliform rash occurring 4 to 9 days after drug exposure.

Another example of a type IVb reaction is DRESS, which is a rare, life-threatening drug reaction typically involving a rash, hematologic abnormalities such as leukocytosis with eosinophilia and/or thrombocytopenia, internal organ involvement (most often liver), and lymphadenopathy. Frequently implicated drugs include allopurinol, carbamazepine, lamotrigine, phenytoin, sulfasalazine, vancomycin, minocycline, dapsone, and sulfamethoxazole with a latency period of 2 to 8 weeks. DRESS remains a clinical diagnosis which is supported by laboratory data and histologic findings. The European Registry of Severe Cutaneous Adverse Reactions (RegiSCAR) has developed a scoring system to help assist with the probability of diagnosis based on clinical criteria.

KEY CONCEPT

RegiSCAR criteria for diagnosis of DRESS

Primary Criteria

Must have 3 of the following:

- Fever >38°C
- Enlarged lymph nodes at a minimum of 2 sites
- Involvement of at least 1 internal organ
- Blood count abnormalities

Secondary Criteria

- Hospitalization
- Reaction is suspected to be drug related
- Rash
- Lymphocytes above or below normal range
- Peripheral eosinophilia
- Thrombocytopenia

Type IVc Reactions

Type IVc reactions occur due to the migration of activated drug-specific $CD8^{St}$ cytotoxic T cells to various tissues, leading to cellular apoptosis or death secondary to the release of granzyme and perforin or through the Fas-ligand dependent pathway. This results in the recruitment of additional inflammatory cells such as neutrophils, eosinophils, and monocytes. Examples of type IVc hypersensitivity reactions include Stevens-Johnson syndrome (SJS) and toxic epidermal necrolysis (TEN).

SJS and TEN are on a spectrum of bullous diseases and are distinguished from one another by disease severity and body surface involvement (Table 50.5). SJS/TEN is characterized by skin necrosis and detachment of the epidermis, usually involving ocular, oral, or genital mucous membranes. Commonly implicated drugs include allopurinol, aromatic antiepileptics, lamotrigine, antibiotics (sulfonamides, β-lactams, and fluoroquinolones), nevirapine, and NSAIDs with a latency

TABLE 50.5 Spectrum of Stevens Johnson Syndrome/Toxic Epidermal Necrolysis

Characteristic	SJS	SJS/TEN Overlap	TEN
% BSA	<10	10–30	>30
Mucosal involvement	Yes	Yes	Yes
Primary lesions	Dusky red lesions Flat atypical targets	Dusky red lesions Flat atypical targets	Poorly delineated erythematous plaques Epidermal detachment Flat atypical targets
Distribution	Face and trunk	Confluent on face and trunk	Confluent throughout the body
Systemic involvement	Usually	Always	Always
Mortality	10%	30%	50%

BSA, body surface area; *SJS*, Stevens Johnson syndrome; *TEN*, toxic epidermal necrolysis.

period of 4 days to 4 weeks. Complications seen with SJS/TEN are secondary to the loss of the protective skin barrier, which leads to massive fluid losses, infections, electrolyte imbalance, hypovolemic shock, and multiple organ dysfunction. There are many potential long-term sequelae of SJS/TEN, including ocular complications that can be sight-threatening, and oral, gastrointestinal, genitourinary, pulmonary, autoimmune, and psychiatric complications.

Type IVd Reactions

Type IVd reactions are characterized by drug-specific T cells producing IL-8 and granulocyte macrophage-colony-stimulating factor (GM-CSF), which induces neutrophil chemotaxis and inhibits cellular apoptosis, ultimately creating sterile neutrophilic inflammation. Acute generalized exanthematous pustulosis (AGEP) is an illustrative example, which is diagnosed clinically through the appearance of numerous sterile, pin-sized, non-follicular pustules with surrounding edematous erythema. It is often associated with fever and neutrophilic leukocytosis, and the onset of symptoms typically occurs between 24 hours and 10 days after drug exposure. Commonly implicated drugs include antibiotics, antimycotics, protease inhibitors and other antimicrobials such as hydroxychloroquine, nifuroxazide, and pyrimethamine.

Direct Mast Cell Activation

Another type of hypersensitivity reaction occurs via innate immune activation of mast cells and are referred to as pseudoallergic reactions. These are important to recognize as diagnosis and management may be drastically different.

Mas-Related G Protein-Coupled Receptor-X2 (MRGPRX2)

First recognized in 2015 by McNeil *et al.*, Mas-Related G Protein-Coupled Receptor-X2 (MRGPRX2) was found to be one of many mast cell receptors capable of recognizing endogenous and exogenous stimuli provoking degranulation.[20] MRGPRX2 is expressed in sensory neurons, mast cells (particularly MC_{TC} found in the skin), and keratinocytes. MRGPRX2-mediated responses seem to be more rapid but transient in comparison to IgE-triggered events. Several medications have been identified that activate mast cells through MRGPRX2 including NMBAs (except for succinylcholine), fluoroquinolones, vancomycin (Red Man syndrome), icatibant, leuprolide, and morphine. Single nucleotide polymorphisms may be linked with MRGPRX2 variants that could predispose individuals to hyperactivation of this receptor by changing its structure, although additional studies are needed to determine their exact significance in clinical practice.[21] Fig. 50.4 demonstrates the main receptor systems and examples of ligands involved in mast cell activation.

Cytokine Release Syndrome

Cytokine release syndrome (CRS) is most commonly associated with biologic and cancer therapies. Patients can develop acute symptoms that may be mistaken in some cases for anaphylaxis. The clinical presentation can range from mild symptoms, such as fever, nausea, vomiting, back pain, and/or erythema, to more severe manifestations including hypoxia, hypotension, organ failure, and even death. Proinflammatory cytokines such as IL-6, IL-10 and IFN-γ are often elevated and targeted therapies against IL-6 or its receptor have been shown to be effective, especially in association with CAR T-cell therapy.

DIAGNOSIS

A detailed history is the most important element in the diagnosis of all drug reactions. It is necessary to determine the type of reaction, the test(s) required to confirm or refute the diagnosis, and to establish a management plan. Diagnostic tools including skin prick testing, intradermal skin testing, patch testing, and *in vitro* testing have been utilized, although most are not standardized or validated, making their interpretation challenging. The gold standard remains an observed drug challenge, which should be considered only if there is uncertainty in the diagnosis after obtaining a detailed history and completing appropriate testing.

Tryptase

Activation of mast cells leads to degranulation and the release of histamine, tryptase, and several other important mediators. This process can be triggered by IgE-dependent or independent pathways and results in symptoms typical of a type I hypersensitivity reaction. After secretion from tissue mast cells, tryptase requires 30 minutes to reach peripheral blood. Its level peaks 1.5–4 hours after symptom onset, but in some patients with anaphylaxis, the level may remain elevated for >24 hours. If the tryptase level is found to be elevated during an event, a baseline tryptase level at least 2 weeks after the episode should be drawn when the patient has returned to a normal state of health. This is helpful in evaluating patients for underlying mast cell disorders in which patients have chronically elevated tryptase levels.

Drug Challenges

Drug challenges (also known as provocation testing or *test dosing*) remains the gold standard for the evaluation of drug allergies. The objective of a challenge is to cautiously introduce a drug in patients who are unlikely to be allergic to it. Drug challenges can be performed as part of the initial diagnostic evaluation if the risk of a reaction is low or after negative skin and/or *in vitro* testing. Challenges can trigger mild to severe, acute and/or delayed reactions, but are considered

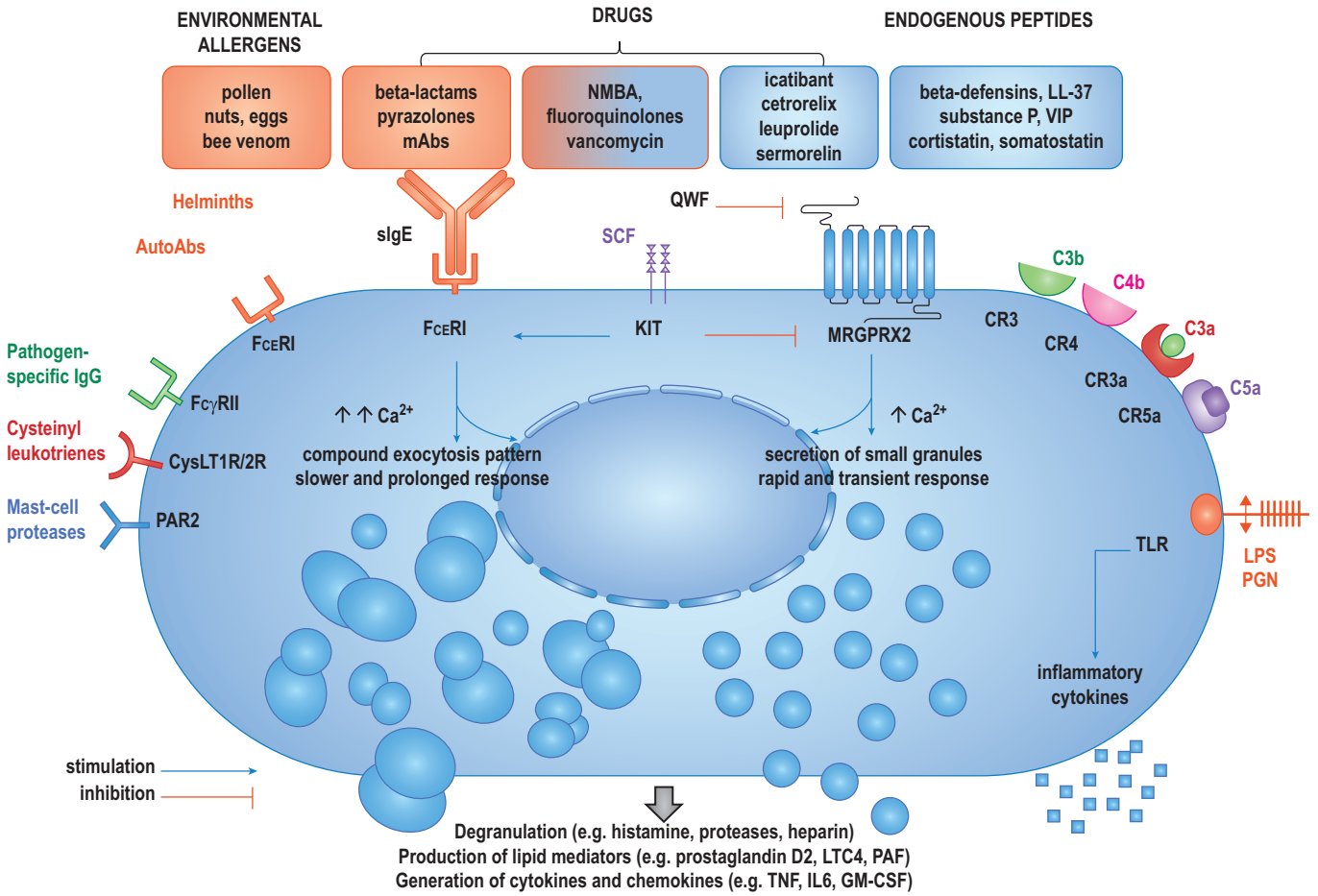


FIG. 50.4 Main receptor systems and examples of ligands involved in mast cell activation. Drugs can activate mast cells through both sIgE-dependent mechanism and the MRGPRX2 receptor. These activation routes are independent and inversely regulated by SCF (51). QWF inhibits activation of human MRGPRX2 by a number of basic secretagogues and medications. Granule processing and response program after the MRGPRX2 engagement differ from FcεRI-mediated response. Several other representative MC receptor systems and ligands are shown. *CR*, Complement receptor for corresponding complement components; *CysLT1R/2R* receptors, cysteinyl leukotrienes; *FcγRII*, a low-affinity receptor for IgG; *FcεRI*, the high affinity IgE receptor; *GM-CSF*, granulocyte/macrophage colony-stimulating factor; *IL*, interleukin; *KIT*, mast/stem cell growth factor receptor (CD117); *LPS*, lipopolysaccharide; *LTC4*, leukotriene C4; *Abs*, antibodies; *MRGPRX2*, Mas-Related G Protein-Coupled Receptor-X2; *NMBA*, neuromuscular blocking agents; *PAR2*, protease-activated receptor 2; *PGN*, peptidoglycan; *SCF*, stem cell factor; *sIgE*, specific IgE; *TLR*, Toll-like receptor; *TNF*, tumor-necrosis factor; *QWF*-tripeptide (the glutaminyl-D-tryptophylphenylalanine); *VIP*, vasoactive intestinal peptide. The schematic drawings were generated by modifying images obtained from Motifolio (Motifolio Inc., Elliocott City, MD).

extremely safe if performed appropriately. Important considerations to determine if a challenge is indicated include the risk of a severe reaction, patient comorbidities, comfort level of the provider in managing reactions, rescue protocol/medication availability, and the patient's needs for the specific drug in question. This procedure should almost never be performed if the reaction history is consistent with a severe non-IgE-mediated reaction, such as SJS/TEN, DRESS, or non-cutaneous organ-specific reaction (*e.g.*, hemolytic anemia).

Challenges can be a single, full-dose administration, or graded, with a small test dose followed by incrementally larger doses. Graded challenges are useful in higher-risk patients, as the initial smaller dose could potentially trigger a less severe reaction compared to a full-dose, single-step challenge.

If there is a concern for patient anxiety interfering with a challenge or actually causing the reaction symptoms, a placebo challenge should be incorporated. This could entail multiple doses of placebo given interspersed with actual drug, blinding the patient so symptoms reported can be objectively assessed without bias.

Skin Prick and Intradermal Testing

Skin prick and intradermal testing can be a valuable tool for evaluating IgE-mediated hypersensitivity reactions. The negative predictive value for penicillin skin testing is >97%. However, the sensitivity and specificity of skin prick testing for most drugs is unknown, which limits their utility in many circumstances. At higher concentrations, drugs can be very irritating to the skin, leading to false positive results. Examples of non-irritating concentrations for common antibiotics are listed in Table 50.6. Delayed intradermal readings (24 to 96 hours after injection) have been suggested as a possible tool for the diagnosis of type IV hypersensitivity reactions, but additional studies are needed to standardize this approach.

Patch Testing

Drug patch testing can be used for the evaluation of T-cell-mediated hypersensitivity reactions including maculopapular exanthem, fixed drug eruptions (FDE), SJS/TEN, DRESS, and AGEP. Current literature supports patch testing for select drugs, including antiepileptics and certain antibiotics; however, it cannot be used reliably for many medications (Table 50.7). Guidelines recommend that testing should occur between 6 weeks and 6 months after resolution of the initial reaction and at least 1 month after discontinuation of systemic steroids or immunosuppressants. Patch testing is typically placed on the upper back, except for FDEs, for which it needs to be performed on previously affected skin. If confirmatory testing is indicated for a T-cell-mediated hypersensitivity reaction and is not apparent based on history alone, delayed intradermal testing or patch testing should be considered.

TABLE 50.6 Nonirritating Concentrations for 15 Commonly Used Antibiotics

Antimicrobial Drug	Full-Strength Concentration (mg/mL)	NIC (as Dilution From Full-Strength Concentration)	No. of Patients Tested
Cefotaxime	100	10^{-1}	25
Cefuroxime	100	10^{-1}	25
Cefazolin	330	10^{-1}	25
Ceftazidime	100	10^{-1}	25
Ceftriaxone	100	10^{-1}	30
Tobramycin	40	10^{-1}	25
Ticarcillin	200	10^{-1}	25
Clindamycin	150	10^{-1}	25
Gentamicin	40	10^{-1}	30
Cotrimoxazole	80	10^{-2}	25
Levofloxacin	25	10^{-3}	25
Erythromycin	50	10^{-3}	25
Azithromycin	100	10^{-4}	30
Nafcillin	250	10^{-4}	25

NIC, Nonirritating concentration.
From Empedrad R, Darter AL, Earlet HS, et al. Nonirritating intradermal skin test concentrations for commonly prescribed antibiotics. J *Allergy Clin Immunol.* 2003;112:629–630.

In Vitro Testing: Specific IgE and Basophil Activation Test

In vitro specific IgE testing to drugs has limited utility at this time due to its lower sensitivity than skin testing. Basophil activation testing (BAT), an *in vitro* method to evaluate for immediate-type reactions, has shown some promising results in the evaluation of NMBAs as a cause of perioperative anaphylaxis when skin testing has been unrevealing.[22] BAT may also be helpful in the evaluation of IgE-mediated reactions to chlorhexidine and latex, but additional studies are needed. There are significant limitations to BAT, including up to 17% of patients being basophil non-responders and the need for fresh blood (<24 hours since blood collection).[23] Commercially available BAT tests generally have not been validated.

MANAGEMENT

Acute Treatment

Initial management is cessation of the suspected culprit drug. After discontinuation of the medication, the symptoms may resolve rapidly, such as in type I hypersensitivity reactions, or could persist for weeks, which can occur with FDEs or DRESS. Additional treatment is imperative for certain types of reactions;

TABLE 50.7 Patch Testing by Type of Drug Reaction

Drug Reaction	Grade of Recommendation[a]	Quality of Evidence[b]	Summary of Published Studies[c]
MPE	2A	B	Overall, 10.8%–38.4% of patients with MPE test positively to the implicated medication on patch testing. Radio contrast media, antiepileptics, penicillins, clindamycin, pristinamycin, and metamizole are more likely to have positive results compared with macrolides and sulfonamides, which are usually negative on patch testing for MPE.
FDE	2A	B	Patch testing is most likely to confirm culprit medication when NSAIDs and sulfa-based antimicrobials are tested. Note: patch testing must be performed on previously affected skin in FDE, with appropriate vehicle (*i.e.*, patch testing of cotrimoxazole is usually negative in petrolatum vehicle and more likely positive in DMSO).
AGEP	2A	B	Limited data suggest patch testing is often positive in AGEP (approximately 58% of cases). Most common medications that illicit positive patch tests include pristinamycin, β-lactams, antiepileptics, and diltiazem.
SCD	2B	B	Limited data suggest patch testing is diagnostic in 100% of cases of SCD with known prior sensitization, and 50% of cases without known prior sensitization.
DRESS	2A	B	Patch testing can confirm the culprit medication in 32%–64% of cases of DRESS. Patch testing is most helpful for DRESS caused by antibiotics (β-lactams and vancomycin), proton-pump inhibitors, and antiepileptics.
EM and SJS/TEN	2B	B	Data on patch testing for EM are limited. Data on SJS/TEN are limited and suggests no utility of patch testing, except possibly for carbamazepine induced SJS/TEN.

[a]According to criteria by Robinson et al.[1] 1 indicates a strong recommendation with high-quality, patient-oriented evidence; 2A indicates a weak recommendation with limited quality, patient-oriented evidence; and 2B indicates a weak recommendation with low-quality evidence.
[b]B indicates a systemic review or meta-analysis of low-quality clinical trials or studies with limitations and inconsistent findings, low-quality clinical trials, cohort studies, and case-control studies.
[c]Complete reference list available from authors.
AGEP, Acute generalized exanthematous pustulosis; *DMSO*, dimethyl sulfoxide; *DRESS*, drug rash with eosinophilia and systemic symptoms; *EM*, erythema multiforme; *FDE*, fixed drug eruption; *MPE*, maculopapular exanthem; *NSAID*, nonsteroidal anti-inflammatory drug; *SCD*, systemic contact dermatitis; *SJS/TEN*, Stevens-Johnson syndrome/toxic epidermal necrolysis.
From Zinn Z, Gayam S, Chelliah MP, *et al.* Patch testing for nonimmediate cutaneous adverse drug reactions. *J Am Acad Dermatol.* 2018;78(2):421–423.

for instance, the early use of epinephrine with anaphylactic reactions can be life-saving. In severe drug-induced immunopathology, such as SJS/TEN and DRESS, there are no specific therapies outside of removal of the offending drug. However, corticosteroids and/or high-dose immunoglobulin therapy are frequently used to assist in decreasing inflammation and accelerating recovery.

In certain clinical scenarios, removal of the culprit drug is not feasible. Depending on the reaction to the treatment medication, continuation of the drug ("treating through") can be considered if the reaction is mild, such as itching or a maculopapular rash. These symptoms may respond favorably to medications such as antihistamines or topical steroids. This approach should be taken with caution with frequent monitoring for progression to more a serious reaction such as DRESS.

Alternative Medications

After an ADR, there are three options for alternative treatment: administering a different medication class, giving a drug within the same or similar class with potential cross-reactivity, or re-administering the same medication. The safest and most common approach is choosing a medication outside of the drug class that caused the reaction. This approach should be taken with caution as second-line therapies may not be as effective in treating the underlying infection and could actually increase risk for multi-drug resistant infections, increased length of hospitalizations, and greater overall healthcare cost, as seen with use of broad-spectrum antibiotic alternatives in patients with reported penicillin allergy.[5]

The second option is choosing a drug that can be potentially cross-reactive, either within the same drug class or if the drug has a similar chemical structure. Commonly implicated drugs with cross-reactivity are penicillins with cephalosporins and carbapenems; sulfonamide-containing drugs; and aromatic anticonvulsants, which will be discussed in more detail later in this chapter.

The third option is re-administration of the same drug. Depending on the type of reaction and duration from the reaction, consideration can be taken to perform a challenge versus desensitization. This is discussed in more detail in the following sections.

Desensitization

For patients who have experienced a type I (either IgE- or non-IgE-mediated) reaction but still require the culprit drug for first-line therapy, desensitization is an option and has been successful with numerous types of drugs, including antibiotics, mAbs, insulin, and chemotherapy. Desensitization is a procedure of incremental administration of increasing sub-optimal doses of the drug to reach a target dose with the goal of remaining beneath the threshold for a hypersensitivity reaction. The underlying mechanism is not completely understood. When drugs are administered at typical doses, antigen-induced membrane changes occur with internalization of the antigen, IgE, and FcεRI leading to activation of mast cells or basophils. Decreasing the amount of drug administered to sub-optimal doses allows the antigen to bind to the membrane but cross-linking of IgE does not occur. Therefore, internalization of the surface molecules does not occur, ultimately preventing actin rearrangement, calcium entry, and the release of inflammatory mediators.

Desensitization protocols can be performed with oral or intravenous medications, with safety in favor of oral administration. There have been numerous effective protocols published with most protocols beginning in the range of 1:100, 000. Typically, the dose is incrementally doubled every 15 to 20 minutes until the target cumulative dose is reached. For parenteral administration of medications, a multi-bag approach was previously the standard of care, but recent single-bag protocols have been developed which demonstrate equivalent safety and efficacy.[24] An example of a parenteral paclitaxel protocol is shown in Table 50.8.

Desensitization lasts as long as the drug is present. If there are lapses, missed doses, or delayed doses, the procedure will need to be repeated. Depending on the half-life of the medication, the desensitized state will dissipate in days to weeks.

Drug Challenges

A drug challenge to a medication that had previously caused an ADR is a common approach taken to assess an allergy. Examples of situations when this practice is utilized includes: if the offending agent was not thought to be the cause of the reported symptoms, if the reaction was reported many years after occurrence, or if the medication is required and symptoms reported were mild. Details of drug challenges are described earlier in the Diagnosis section.

Premedication Protocols

Premedications can be useful in pseudoallergic reactions such as with vancomycin-induced Red Man Syndrome or prior to desensitization protocols for nonallergic reactions such as with taxanes. Premedications given prior to administration of the drug include antihistamines, topical or systemic steroids, leukotriene receptor antagonists, or NSAIDs. Premedicating is not recommended for drug challenges if there is concern for an IgE-mediated reaction, as premedications may mask symptoms and lead to false negative results. In this situation, subsequent administration of the drug could lead to a severe reaction even if the same pretreatment regimen is used.

Delabeling Allergy After Negative Testing

If the drug allergy in question has been confirmed negative, it is important to delabel (remove the label) from the medical record(s) of the patient. It is important to review this with the patient to assure optimal treatment in the future is utilized. Commonly, patients are provided with wallet cards that highlight all of their current allergies, in addition to negative testing of previously reported drug allergies; such cards can be helpful if the patient is seen by different providers in various care settings.

SPECIFIC DRUG HYPERSENSITIVITY

Antibiotics

Penicillins

It has been reported that all hypersensitivity reactions can be caused by penicillins (types I–IV), with the highest rate of type I-IgE mediated hypersensitivity reactions. Type I, II, and III hypersensitivity reactions occur through the development of hapten-carrier complexes which is discussed previously in this chapter and demonstrated in Fig. 50.5. While penicillin allergy is the most commonly reported drug allergy, comprising approximately 10% of the U.S. population, over 90% of patients can actually tolerate re-exposure.[5] This is in part due to the loss of sensitization overtime, as approximately 80% to 90% of allergic patients will lose their sensitivity with 10 years of avoidance.

TABLE 50.8 Parenteral Paclitaxel Desensitization Protocol

A. Three-Bag, 12-Step Protocol

Step	Conc (mg/mL)	Rate (mL/h)	Time (min)	Volume per Step (mL)	Dose per Step (mg)	Cumulative Dose (mg)
1	0.008	2.5	15	0.625	0.005	0.005
2	0.008	5	15	1.25	0.01	0.015
3	0.008	10	15	2.5	0.02	0.035
4	0.008	20	15	5	0.04	0.075
5	0.08	5	15	1.25	0.1	0.175
6	0.08	10	15	2.5	0.2	0.375
7	0.08	20	15	5	0.4	0.775
8	0.08	40	15	10	0.8	1.575
9	0.8	10	15	2.5	2	3.575
10	0.8	20	15	5	4	7.575
11	0.8	40	15	10	8	15.575
12	0.8	80	172.9	230.5	184.425	200

B. Nondilution 1-Bag, 13-Step Protocol

Step	Conc (mg/mL)	Rate (mL/h)	Time (min)	Volume per Step (mL)	Dose per Step (mg)	Cumulative Dose (mg)	Estimated Conc with Side Steam[a] (mg/mL)
1	0.8	0.1	15	0.03	0.02	0.02	0.0079
2	0.8	0.2	15	0.05	0.04	0.06	0.0157
3	0.8	0.4	15	0.1	0.08	0.14	0.0308
4	0.8	0.8	15	0.2	0.16	0.3	0.0593
5	0.8	1.5	15	0.38	0.3	0.6	0.1043
6	0.8	3	15	0.8	0.6	1.2	0.1846
7	0.8	6	15	1.5	1.2	2.4	0.3000
8	0.8	12.5	15	3.1	2.5	4.9	0.4444
9	0.8	25	15	6.3	5	9.9	0.5714
10	0.8	50	15	12.5	10	19.9	0.6667
11	0.8	100	15	25	20	39.9	0.7273
12	0.8	200	15	50	40	79.9	0.7619
13	0.8	350	25.7	150.1	120.1	200	0.7869

[a]Calculated by taking the rate of side stream 10 m:/h of 5% dextrose in water into account with the infusion rates of main steam 0.8 mg/mL of paclitaxel solution (200 mg in 250 mL of 5% dextrose water) at each step.

Conc, Concentration.

From Lee JH, Moon M, Kim YC, *et al.* A one-bag rapid desensitization protocol for paclitaxel hypersensitivity: A noninferior alternative to a multi-bag rapid desensitization protocol. *J Allergy Clin Immunol Pract.* 2020;8(2):696–703.

Drug allergy evaluation should be considered for nearly all patients with a history of penicillin allergy as recent studies have shown carrying a penicillin allergy label can have significant clinical consequences. These include the more frequent use of second-line antibiotics, increased treatment failure rates, longer hospital stays, increased risk for the development of resistant organisms, *Clostridium difficile* infections, and higher rates of surgical site infections perioperatively.[5]

For assessment of a type I IgE-mediated allergy, skin testing and/or direct challenges have been used effectively. Multiple protocols have been published to assist in risk-stratifying patients to determine if skin testing is necessary or if the patient can go directly to an observed challenge. For skin testing, typically both major and at least one minor determinant (usually benzylpenicillin) are tested, though there is lack of uniformity of which minor determinants should be used. There is significant cross-reactivity between penicillins, although in Europe, there are higher reports of isolated allergy to amoxicillin, which is not observed in the United States, possibly making more extensive skin testing to include amoxicillin a minor determinant useful for select high-risk patients. Currently, there are no commercially available skin test reagents for amoxicillin and all of the minor determinants, so skin testing with benzylpenicilloyl polylysine and benzylpenicillin followed by amoxicillin oral challenge is the most commonly utilized method for penicillin skin testing in the United States.

Assessment of type IV hypersensitivity reactions is less standardized. Delayed skin testing and patch testing have been attempted but results have been inconsistent. It is not currently recommended for clinical practice, therefore, oral challenges are the suggested method of assessment. Careful history taking remains paramount as challenges are relatively contraindicated for reactions consistent with drug-induced cytopenias, vasculitis, and SCARs. If there is a concern for a benign, delayed-type reaction, administration of a single-dose amoxicillin 250 mg challenge with at home assessment over 1 week for a potential delayed exanthem is a reasonable option.

In penicillin allergy, the degree of similarity of the R-group side chains determines the amount of cross-reactivity. In a recent meta-analysis of patients with proven penicillin allergy (mostly aminopenicillins), identical R-groups had a 16.45% risk of cross-reactivity with only 2.11% in dissimilar R-groups. Carbapenem cross-reactivity to penicillins was even lower at 0.87%.[25] The rates of cross-reactivity are likely much lower in patients with self-reported penicillin allergy since >95% of such patients when tested are not truly allergic.

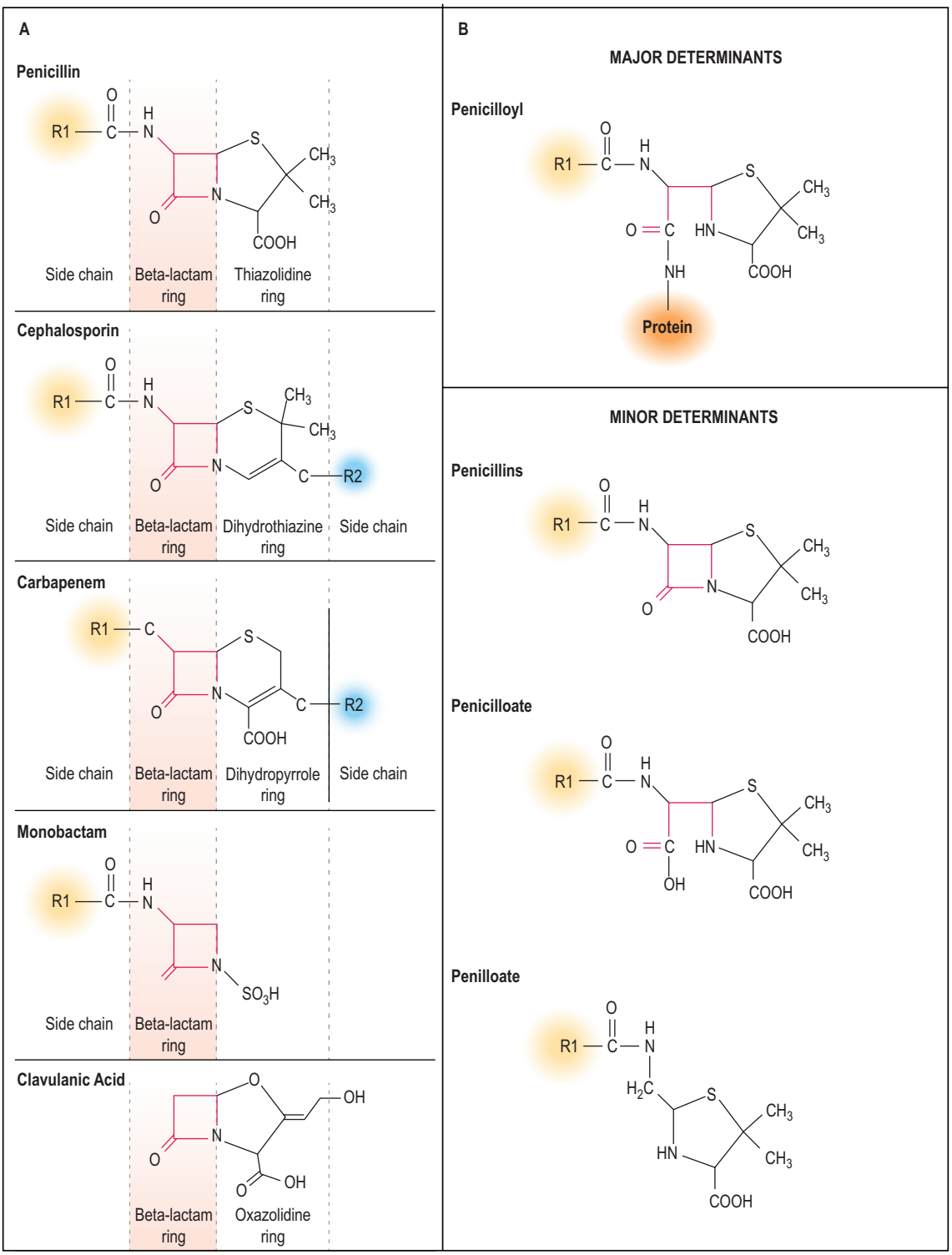


FIG. 50.5 Penicillin and beta-lactam structure and major and minor penicillin determinants. (From Castells M, Khan DA, Philips EJ. Penicillin allergy. *N Engl J Med.* 2019;381[24]:2338–2351.)

Cephalosporins

Like penicillins, cephalosporins have been associated with all types of hypersensitivity reactions. Skin testing to penicillins is not an appropriate method of evaluating cephalosporin allergy. In patients reporting an allergy to specific cephalosporins, skin testing can be considered, although the exact sensitivity and specificity for each cephalosporin is not currently known. They should, however, be able to tolerate other cephalosporins that do not share a common side chain, as demonstrated in Fig. 50.6. For patients with anaphylactic reactions to cephalosporins, skin testing to an alternative cephalosporin with a disparate R1 side chain is recommended prior to administration.

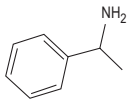
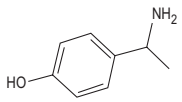
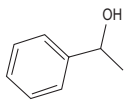
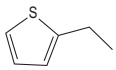
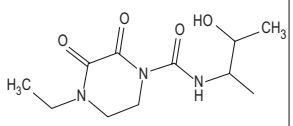
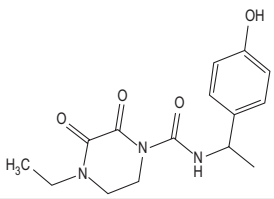
Cephalosporin	R_1 Structure	Penicillin Identical R_1	Penicillin Similar R_1
Cephalexin Cephaloglycin Cefaclor Loracarbef		Ampicillin Pivampicillin*+ Bacampicillin*+ Talampicillin*+	Mezlocillin+ Piperacillin Azlocillin+
Cefadroxil Cefatrizine Cefprozil		Amoxicillin	Mezlocillin Piperacillin Azlocillin
Cefamandole Cefonicid			Ampicillin Amoxicillin Pivampicillin Bacampicillin Talampicillin
Cefoxitin Cephaloridine Cephalothin			Ticarcillin+ Temocillin+
Cefbuperazone			Piperacillin
Cefbuperazone			Piperacillin

FIG. 50.6 Cephalosporin drugs with similar R1 side-chain structures. (From Zagursky RJ and Pichichero ME. Cross-reactivity in β-lactam allergy. *J Allergy Clin Immunol Pract*. 2018;6:72–81.)

Sulfonamide Antibiotics

Sulfonamide antibiotics are characterized by a five or six membered ring with sulfonamide (SO2-NH2) directly attached to the ring and an unsubstituted amine (-NH2) at position N4. There are non-antibiotic sulfonamide drugs that do not share these same structures, therefore, they are tolerated safely by patients with sulfonamide antibiotic allergy. This group of antibiotics is most commonly associated with delayed hypersensitivities, which includes FDEs, maculopapular rash, and SCARs, although all types of reactions have been observed. In comparison to other antibiotics, sulfonamide antibiotics have a higher risk of SJS/TEN. Patients with untreated HIV or with low $CD4^+$ T-cell counts have been found to have a higher risk for sulfonamide antibiotic allergy.[2] Most patients with histories of benign reactions to sulfonamide antibiotics will tolerate challenges to trimethoprim-sulfamethoxazole.[26]

Other Antibiotics

Intolerance prevalence and immunogenicity of common antibiotics are highlighted in Table 50.9.

Aspirin and Nonsteroidal Anti-Inflammatory Drugs

After antibiotics, nonsteroidal anti-inflammatory drugs (NSAIDs) are the most commonly reported drug allergy by patients with a variety of different mechanisms, including immunologic and non-immunologic processes. NSAID reactions can be classified into one of five subtypes, which are highlighted in Table 50.10. The diagnosis can often be made on clinical history alone, while some patients require a drug challenge. Skin and *in vitro* testing have not been shown to be reliable for NSAID allergies.

NSAID-Exacerbated Respiratory Disease

NSAID-exacerbated respiratory disease (N-ERD), also known as aspirin-exacerbated respiratory disease (AERD) in the United States, is a chronic inflammatory disorder of the upper and lower respiratory tract. This eosinophilic-predominant condition was previously referred to as Samter's triad and involves asthma and/or chronic rhinosinusitis with nasal polyps along with NSAID sensitivity, including aspirin. NSAID-related reactions typically occur within 30 to 180 minutes after exposure and can include nasal congestion, rhinorrhea, wheezing, coughing, and shortness of breath. A subset of patients may have cutaneous or gastrointestinal symptoms as well. The provocation dose is typically between 30 and 300 mg of aspirin with most reacting to doses of 60 mg or less.

The underlying pathophysiology of N-ERD has been hypothesized to be related to blocking cyclooxygenase-1 (COX-1), and not COX-2, causing inhibition of prostaglandin E_2 (PGE_2). This decrease in PGE_2 biosynthesis can trigger innate activation of inflammatory cells, specifically mast cells, basophils, eosinophils and possibly platelets. The eventual release of cysteinyl leukotrienes, PGD_2, histamine, tryptase, and other mediators could explain the clinical phenotype. Desensitization protocols

TABLE 50.9 Intolerance Prevalence and Immunogenicity of Common Antibiotics

Antibiotic or Antibiotic Family	Intolerance Prevalence (%)	IgE-Medicated Allergy	T-cell-Mediated Delayed Hypersensitivity	Intrafamily Immunologic Cross-Reactivity
Penicillins	7.9	Possible	Possible	Common
Sulfonamides	4.3	Unlikely	Possible	Unlikely
Macrolides	1.2	Unlikely	Unlikely	Unknown
Cephalosporins	1.1	Possible	Possible	Unlikely
Tetracyclines	0.70	Unlikely	Unlikely	Unknown
Quinolones	0.46	Possible	Unknown	Common
Nitrofurantoin	0.24	Unlikely	Unlikely	NA
Clindamycin	0.20	Unlikely	Possible	NA
Metronidazole	0.15	Unlikely	Possible	NA

IgE, Immunoglobulin E; *NA*, not applicable/available.

From Macy E, Romano A, Khan D. Practical management of antibiotic hypersensitivity in 2017. *J Allergy Clin Immunol Pract.* 2017;5:577–586.

TABLE 50.10 Classification of NSAID-Induced Hypersensitivity Reactions

Timing of Reaction	Clinical Symptoms	Cross-reactivity with NSAID Class	Presence of Underlying Disease	Putative Mechanism
AERD				
Acute	Rhinitis, nasal congestion, bronchoconstriction, asthma exacerbation	Cross-reactive	Asthma/rhinosinusitis/nasal polyps	COX-1 inhibition
Multiple NSAID-Exacerbated Urticaria/Angioedema in Patients With Underlying Cutaneous Disease				
Acute	Urticaria/angioedema	Cross-reactive	Chronic urticarial	COX-1 inhibition
Multiple NSAID-Induced Uricaria/Angioedema in Otherwise Asymptomatic Patients				
Acute	Urticaria/angioedema	Cross-reactive	None	Likely COX-1 inhibition
Single NSAID-Induced Anaphylactic Reactions				
Acute	Anaphylaxis, urticarial/angioedema	Single drug-induced	Atopy is common	IgE-mediated
Delayed Reactions to NSAIDs				
Delayed	Varied: Fixed drug eruptions, severe bullous skin reactions, maculopapular drug eruptions	Can be single drug-induced or cross-reactive	None	Varied: T-cell-mediated, cytotoxic T cells, natural killer cells, other

From Laidlaw TM, Cahill KN. Current knowledge and management of hypersensitivity to aspirin and NSAIDs. *J Allergy Clin Immunol Pract.* 2017;5(3):537–545.

have been successful in N-ERD patients to decrease nasal polyp burden, improve asthma control, and improve hyposmia/anosmia.

Radiocontrast Media

Severe reactions to iodinated and noniodinated contrast media are relatively rare (0.02% to 0.03%), and their mechanisms are not completely understood. In the past, high-osmolar iodinated contrast media (ICM) were thought to cause reactions through IgE-independent pseudoallergic mechanisms, and premedication regimens were considered effective. Use of low or iso-osmolar ICM has further reduced the rate of severe hypersensitivity reactions. There is accumulating evidence that reactions to low or iso-osmolar ICM may be IgE-mediated. A recent consensus document suggested that skin testing may help identify safe alternatives, but prospective studies are needed to determine the optimal testing strategy.[27] Skin testing therefore remains controversial as a means of predicting subsequent reactions, and the decision to perform testing remains with the clinician.[28, 29] The most recent update to the anaphylaxis parameters, based on a systematic review, recommended against routinely administering glucocorticoids and/or antihistamines to prevent anaphylaxis in patients with prior radiocontrast hypersensitivity reactions when re-administration of a low- or iso-osmolar, nonionic ICM agent is required.[30] The parameter notes that premedication may be considered in clinical circumstances associated with greater anaphylaxis fatality risk (*e.g.*, underlying cardiovascular disease, prior severe anaphylaxis). Numerous pretreatment regimens have been proposed with the most widely used protocols involving systemic steroids and antihistamines at various intervals prior to contrast exposure.

Gadolinium-based contrast agents (GBCA) do not cross-react with ICM and risk of hypersensitivity reactions is even lower. Rates of immediate reactions to different GBCAs vary with higher hypersensitivity reactions associated with protein binding, macrocyclic structure, and ionicity.[31] Reactions to GBCA may also be IgE-mediated, and skin testing has been utilized to identify alternative GBCAs.[32]

Angiotensin-Converting Enzyme Inhibitors

ACE-I competitively inhibit angiotensin-converting enzyme, blocking the synthesis of angiotensin-II. In addition to the desired effect of decreasing cardiac preload and afterload, ACE-I can reduce the breakdown of bradykinin and substance P, resulting in bradykinin-mediated angioedema. A higher risk of ACE-I induced angioedema has been seen in Black people, women, smokers, and those with underlying hereditary angioedema. After the first episode of angioedema, ACE-I should be permanently discontinued, but many patients can continue to have episodes of angioedema for several months after ACE-I

cessation.[33] Angiotensin receptor blockers (ARBs) can safely be used in these patients.

Perioperative Anaphylaxis

The incidence of perioperative anaphylaxis is reported to be between 1:1, 250 and 1:20, 000 with a 3% to 9% mortality rate.[34] Evaluation of these events can often be complex due to multiple drugs used in a short period of time; furthermore, early signs of a hypersensitivity reaction are often missed due to the setting, as patients are intubated, sedated, and draped. While nearly all medications and substances used in the perioperative period have been implicated as causes of anaphylaxis, the most common include neuromuscular blocking agents (NMBAs), antibiotics (particularly cefazolin), disinfectants, and latex. While skin and *in vitro* specific IgE testing can be helpful, the causative agent in 30% to 50% of cases remains unknown. Common skin testing concentrations can be found in a review by Volcheck and Hepner.[34] Other diagnostic modalities that are less frequently utilized include direct provocation challenges and BAT.

Local Anesthetics

Systemic hypersensitivity reactions to local anesthetics are exceedingly rare, and most are not IgE-mediated. While reactions are often reported by patients, the largest cohort study to date showed that >85% of episodes were due to psychosomatic or vasovagal reactions.[35] Only 0.5% of the reported cases were attributed to the local anesthetic. The categorization of local anesthetics into benzoic ester and amide drugs is not particularly meaningful in the evaluation of systemic reactions. The original publications used this classification only for allergic contact dermatitis reactions, and its relevance for other types of hypersensitivity reactions is unknown.

Chemotherapy

Anti-cancer drug therapies are defined based on their target and include chemotherapy, hormonal therapy, and immunotherapy. Chemotherapeutics are then sub-grouped based on their structure and mechanism of action. Chemotherapeutics commonly associated with adverse reactions include platinum salts and taxanes.

Platinum salts have been implicated in both IgE- and non-IgE-mediated reactions with an incidence of 12% to 17% respectively.[36] The most common implicated platinum salt causing reactions is carboplatin followed by oxaliplatin and cisplatin.[36] Cross-reactivity is much higher between carboplatin and oxaliplatin and is less common with carboplatin.[37] Risk factors for platinum salt reactions are female gender, platinum-free interval greater than 1 year, and BRCA-positive genetic mutations. Hypersensitivity reactions generally occur between the 4th and 10th dose, while non-IgE-mediated reactions will occur on the 1st or 2nd dose. Skin testing to platinum salts has been demonstrated to be useful given most reactions are IgE mediated, and it is often used to predict the need to desensitize, although this process is not currently standardized.[38]

Taxane reactions are most often caused by non-IgE mediated hypersensitivity, with reactions typically occurring on the 1st or 2nd exposure. The reactions are thought to be secondary to formulation vehicles, such as polyethoxylated castor oil and polysorbate 80, which are used to solubilize taxanes and are capable of causing mast cell activation. Typically, these reactions respond favorably to premedication regimens. Taxanes have also been implicated in IgE-mediated hypersensitivity, although less commonly than platinum salts. The use of skin testing is less established.

Biologics and Monoclonal Antibodies

Biologic agents (BA) are becoming increasingly prevalent in the medical management of cancer and chronic inflammatory conditions. As BAs have been developed to mimic and alter the immune system, it is not surprising that new types of AEs have been noted. These are often due to overstimulation or suppression of the immune system leading to conditions such as CRS, infusion reactions, secondary immunodeficiency, and autoimmunity. mAbs, one type of BA, were originally created as murine analogs but have been generally replaced by chimeric, humanized, and fully human antibodies to improve efficacy and decrease allergenicity.

Reactions to mAbs can occur during the first exposure, as has been reported with cetuximab and trastuzumab or after multiple exposures, such as with rituximab. Infusion-related reactions to mAbs present as cytokine storm-like reactions with symptoms of nausea, chills, fever, and malaise. These are thought to be caused by release of proinflammatory cytokines

TABLE 50.11 Biologic Agents: Actions, Incidence, and Hypersensitivity Drug Reactions

Drug	Target	Overall Reactions	HSR
Rituximab (Rituxan) IV	CD20	77% (first infusion)	5%–10%
Ofatumumab (Arzerra) IV	CD20	44% (first infusion) 67% (combination therapy)	2%
Obinutuzumab (Gazyva) IV	CD20	66%	[a]
Trastuzumab (Herceptin) IV	HER-2	40% (mild; first infusion)	0.6%–5%[a]
Cetuximab (Erbitux) IV	EGFR	15%–21%	1.1%–5% 14%–27% (southern USA)[a]
Tocilizumab (Actemra) IV	IL-6 receptor	7%–8%	0.1%–0.7%[a]
Infliximab (Remicade) IV	TNF-α	5%–18%	1%[a]
Etanercept (Enbrel) SC	TNF-α	15%–37%	<2%[a]
Adalimumab (Humira) SC	TNF-α	20%	1%[a]
Golimumab (Simponi) SC	TNF-α	4%–20%	Not reported
Certolizumab (Cimzia) SC	TNF-α	0.8%–4.5%	Not reported
Brentuximab (Adcetris) IV	CD30	12%	[a]
Bevacizumab (Avastin) IV	VEGF-A	<3%	[a]
Omalizumab (Xolair) SC	IgE	45%	0.09%–0.2%[a]

[a]Case reports of anaphylaxis.
HSR, Hypersensitivity reactions; *IL*, interleukin; *TNF*, tumor necrosis factor.
Modified from Galvao VR, Castells MC. Hypersensitivity to biologic agents-updated diagnosis, management, and treatment. *J Allergy Clin Immunol Pract.* 2015;3(2):175–185; quiz 186.

(*e.g.*, IL-1, IL-6, and TNF-α) and respond to NSAIDs and systemic steroids.

Hypersensitivity reactions to cetuximab and several other biologics on first exposure have been attributed to sensitization to the galactose-α-1, 3-galactose (alpha-gal) epitope caused by a bite from the lone star tick (*Amblyomma americanum*), which has a high prevalence in Southeastern United States.[39] The carbohydrate alpha-gal is expressed in the major blood group of non-primate mammalian proteins and also on cetuximab, abciximab, basiliximab, canakinumab, infliximab, golimumab, and ustekinumab.

mAbs given subcutaneously can elicit injection-site reactions, with symptoms including local redness, warmth, burning, stinging, itching, urticaria, pain, and induration. Such reactions can start within one hour of the injection and usually resolve over a few days, but large and persistent reactions can lead to discontinuation of the mAb; desensitization protocols have been developed for patients who have no alternative medication options (Table 50.11).

CONCLUSION

The understanding of immunologic and non-immunologic ADRs remains incomplete and is further compounded by the limited number of reliable diagnostic tests. A detailed history remains the key to distinguishing a true drug allergy from a non-immunologic or anxiety reaction, which can further delineate the underlying mechanism and diagnostic approach that should be taken. Given the high morbidity and mortality associated with drug hypersensitivity reactions as well as the increased failure rate with second- or third-line therapies, appropriate evaluation by Allergy & Immunology specialists is imperative.

KEY CONCEPT

General Evaluation of Drug Allergy

General Evaluation of Drug Allergy

- History: Indication for drug use, association with viral/bacterial infection
- Physical examination
- Blood count/differential, Liver function tests
- Serum tryptase
- Skin testing
- Patch testing/delayed reading of skin test
- Specific IgE, basophil activation test
- Genotyping

Specific Evaluation of Drug Allergy

- Name of the drug, ingredients, preservatives
- First exposure
- How long ago did the reaction occur?
- Re-exposure: Has the patient been exposed to the drug or a related drug?
- Other drugs administered at the same time
- Associated narcotics
- Symptoms and signs of the reaction
- Timing of symptoms relative to the drug exposure

Underlying Condition for which the Drug was Prescribed

- Similar symptoms unrelated to the drug exposure (urticaria)
- Treatment and response to treatment (epinephrine)
- Timing of resolution

REFERENCES

1. Warrington R, Silviu-Dan F, Wong T. Drug allergy. *Allergy Asthma Clin Immunol*. 2018;14(suppl 2):60.
2. Thong BY, Tan TC. Epidemiology and risk factors for drug allergy. *Br J Clin Pharmacol*. 2011;71(5):684–700.
3. Wong A, Seger DL, Lai KH, et al. Drug hypersensitivity reactions documented in electronic health records within a large health system. *J Allergy Clin Immunol Pract*. 2019;7(4):1253–1260.
4. Demoly P, Mertes M, Moneret-Vautrin AD, et al. Contribution of epidemiology to the prevention of drug allergies. *Bull Acad Natl Med*. 2011;195(6):1335–1342. discussion 42–44.
5. Castells M, Khan DA, Phillips EJ. Penicillin allergy. *N Engl J Med*. 2019;381(24):2338–2351.
6. Chen CB, Abe R, Pan RY, et al. An updated review of the molecular mechanisms in drug hypersensitivity. *J Immunol Res*. 2018;2018:6431694.
7. Solensky R, Jacobs J, Lester M, et al. Penicillin allergy evaluation: a prospective, multicenter, open-label evaluation of a comprehensive penicillin skin test kit. *J Allergy Clin Immunol Pract*. 2019;7(6):1876–1885.
8. Wei CY, Chung WH, Huang HW, et al. Direct interaction between HLA-B and carbamazepine activates T cells in patients with Stevens-Johnson syndrome. *J Allergy Clin Immunol*. 2012;129(6):1562–1569. e5.
9. Yun J, Marcaida MJ, Eriksson KK, et al. Oxypurinol directly and immediately activates the drug-specific T cells via the preferential use of HLA-B*58:01. *J Immunol*. 2014;192(7):2984–2993.
10. Illing PT, Vivian JP, Dudek NL, et al. Immune self-reactivity triggered by drug-modified HLA-peptide repertoire. *Nature*. 2012;486(7404):554–558.
11. Watkins S, Pichler WJ. Sulfamethoxazole induces a switch mechanism in T cell receptors containing TCRVbeta20-1, altering pHLA recognition. *PLoS One*. 2013;8(10):e76211.
12. Matar R, Le Bourgeois M, Scheinmann P, de Blic J, Ponvert C. Beta-lactam hypersensitivity in children with cystic fibrosis: a study in a specialized pediatric center for cystic fibrosis and drug allergy. *Pediatr Allergy Immunol*. 2014;25(1):88–93.
13. Carr A, Cooper DA, Penny R. Allergic manifestations of human immunodeficiency virus (HIV) infection. *J Clin Immunol*. 1991;11(2):55–64.
14. Moon DH, Lee JM, Noonan AM, et al. Deleterious BRCA1/2 mutation is an independent risk factor for carboplatin hypersensitivity reactions. *Br J Cancer*. 2013;109(4):1072–1078.
15. Mallal S, Phillips E, Carosi G, et al. HLA-B*5701 screening for hypersensitivity to abacavir. *N Engl J Med*. 2008;358(6):568–579.
16. Konvinse KC, Trubiano JA, Pavlos R, et al. HLA-A*32:01 is strongly associated with vancomycin-induced drug reaction with eosinophilia and systemic symptoms. *J Allergy Clin Immunol*. 2019
17. Jurado-Escobar R, Perkins JR, Garcia-Martin E, et al. Update on the genetic basis of drug hypersensitivity reactions. *J Investig Allergol Clin Immunol*. 2017;27(6):336–345.
18. Lieberman P, Nicklas RA, Randolph C, et al. Anaphylaxis—a practice parameter update 2015. *Ann Allergy Asthma Immunol*. 2015;115(5):341–384.
19. Anno T, Kaneto H, Kawasaki F, et al. Drug fever and acute inflammation from hypercytokinemia triggered by dipeptidyl peptidase-4 inhibitor vildagliptin. *J Diabetes Investig*. 2019;10(1):182–185.
20. McNeil BD, Pundir P, Meeker S, et al. Identification of a mast-cell-specific receptor crucial for pseudo-allergic drug reactions. *Nature*. 2015;519(7542):237–241.
21. Porebski G, Kwiecien K, Pawica M, et al. Mas-related G protein-coupled receptor-X2 (MRGPRX2) in drug hypersensitivity reactions. *Front Immunol*. 2018;9:3027.
22. Li J, Best OG, Rose MA, et al. Assessing cross-reactivity to neuromuscular blocking agents by skin and basophil activation tests in patients with neuromuscular blocking agent anaphylaxis. *Br J Anaesth*. 2019;123(1):144–150.
23. Santos AF, Shreffler WG. Road map for the clinical application of the basophil activation test in food allergy. *Clin Exp Allergy*. 2017;47(9):1115–1124.
24. Lee JH, Moon M, Kim YC, et al. A one-bag rapid desensitization protocol for paclitaxel hypersensitivity: a noninferior alternative to a multi-Bag rapid desensitization protocol. *J Allergy Clin Immunol Pract*. 2020;8(2):696–703.

25. Picard M, Robitaille G, Karam F, et al. Cross-reactivity to cephalosporins and carbapenems in penicillin-allergic patients: two systematic reviews and meta-analyses. *J Allergy Clin Immunol Pract*. 2019;7(8):2722–2738.
26. Krantz MS, Stone Jr CA, Abreo A, et al. Oral challenge with trimethoprim-sulfamethoxazole in patients with "sulfa" antibiotic allergy. *J Allergy Clin Immunol Pract*. 2020;8(2):757–760.
27. Sanchez-Borges M, Aberer W, Brockow K, et al. Controversies in drug allergy: radiographic contrast media. *J Allergy Clin Immunol Pract*. 2019;7(1):61–65.
28. Sese L, Gaouar H, Autegarden JE, et al. Immediate hypersensitivity to iodinated contrast media: diagnostic accuracy of skin tests and intravenous provocation test with low dose. *Clin Exp Allergy*. 2016;46(3):472–478.
29. Schrijvers R, Breynaert C, Ahmedali Y, et al. Skin testing for suspected iodinated contrast media hypersensitivity. *J Allergy Clin Immunol Pract*. 2018;6(4):1246–1254.
30. Shaker MS, Wallace DV, Golden DBK, et al. Anaphylaxis-a 2020 practice parameter update, systematic review, and grading of recommendations, assessment, development and evaluation (GRADE) analysis. *J Allergy Clin Immunol*. 2020;145(4):1082–1123.
31. Behzadi AH, Zhao Y, Farooq Z, et al. Immediate allergic reactions to gadolinium-based contrast agents: a systematic review and meta-analysis. *Radiology*. 2018;286(2):471–482.
32. Bianchi L, Hansel K, Marietti R, et al. Anaphylaxis after first exposure to gadoterate meglumine: a case report and literature review. *J Allergy Clin Immunol Pract*. 2018;6(6):2124–2126.
33. Beltrami L, Zanichelli A, Zingale L, et al. Long-term follow-up of 111 patients with angiotensin-converting enzyme inhibitor-related angioedema. *J Hypertens*. 2011;29(11):2273–2277.
34. Volcheck GW, Hepner DL. Identification and management of perioperative anaphylaxis. *J Allergy Clin Immunol Pract*. 2019;7(7):2134–2142.
35. Trautmann A, Goebeler M, Stoevesandt J. Twenty years' experience with anaphylaxis-like reactions to local anesthetics: genuine allergy is rare. *J Allergy Clin Immunol Pract*. 2018;6(6):2051–2058.
36. Pradelli J, Verdoire P, Boutros J, et al. Allergy Evaluation of hypersensitivity to platinum salts and taxanes: a six-year experience. *J Allergy Clin Immunol Pract*. 2020;8(5):1658–1664.
37. Pasteur J, Favier L, Pernot C, et al. Low cross-reactivity between cisplatin and other platinum salts. *J Allergy Clin Immunol Pract*. 2019;7(6):1894–1900.
38. Markman M, Zanotti K, Peterson G, et al. Expanded experience with an intradermal skin test to predict for the presence or absence of carboplatin hypersensitivity. *J Clin Oncol*. 2003;21(24):4611–4614.
39. Steinke JW, Platts-Mills TA, Commins SP. The alpha-gal story: lessons learned from connecting the dots. *J Allergy Clin Immunol*. 2015;135(3):589–596. quiz 97.

Mechanisms of Autoimmunity

Tory P. Johnson, Brendan Antiochos, and Antony Rosen

Human autoimmune diseases affect >5% of the population worldwide and impose a significant burden of morbidity and mortality. Autoimmune diseases are defined as diseases in which immune responses to specific self-antigens contribute to ongoing tissue damage. Both the specificity of the immune response and its role in tissue damage are central componnts of the definition. Autoimmune diseases can be either tissue-specific (e.g., thyroid), where unique tissue-specific antigens are targeted, or they can be systemic, in which multiple tissues are affected and a variety of ubiquitously expressed autoantigens are targeted.[1] Although the definition appears simple in concept, the enormous complexity of this spectrum of disorders has challenged clarification of simple shared mechanisms. This complexity affects almost every domain, including genetics, phenotypic expression, and kinetics. For example, there is frequently a prolonged period between initial onset of symptoms and development of the diagnostic phenotype, and the expression of disease may vary within the same individual over time. However, despite this enormous complexity, there is a striking association of the clinical phenotype with the targets of the autoimmune response. Indeed, this association is so strong that autoantibodies have been used for diagnosis and prognosis in human autoimmune diseases.[1] For example, autoantibodies recognizing thyroid peroxidase are found in patients with autoimmune thyroiditis, and autoantibodies to the Sm splicing ribonucleoprotein (RNP) complex are diagnostic of systemic lupus erythematosus (SLE). The immune response in autoimmune diseases has features of an adaptive immune response—usually directed against exogenous antigens—but its targets are autoantigens. Since the adaptive immune response is initiated when suprathreshold concentrations of molecules with structures not previously tolerized by the host are encountered in a proimmune context, the association of specific autoantibodies with distinct clinical phenotypes provides critical clues to understanding the initiation and propagation of autoimmune diseases.

KEY CONCEPTS

Autoantibodies in Autoimmune Diseases

- Autoantibodies may precede the development of any symptoms by years (e.g., antinuclear antibodies and antiphospholipid antibodies in systemic lupus erythematosus [SLE], anti-cyclic citrullinated peptide [CCP] in rheumatoid arthritis [RA])
- Some autoantibodies only occur at the onset of disease manifestations (e.g., anti-Sm and anti-ribonucleic protein [RNP] in SLE).
- There is a striking association of specific autoantibodies with distinct clinical phenotypes (e.g., antitopoisomerase-1 with diffuse scleroderma and interstitial lung disease).

This chapter highlights some of the mechanistic principles that underlie autoimmune diseases. The extraordinary breadth and complexity of this disease spectrum precludes encompassing everything relevant.

THE DISTINCT PHASES IN THE DEVELOPMENT OF AUTOIMMUNITY

A major barrier to understanding mechanisms of autoimmunity comes from difficulty in defining early events in these diseases. Since diseases are only recognizable after development of the diagnostic phenotype, there has been a tendency to interpret findings at the time of diagnosis with disease initiation. However, significant recent data from longitudinal studies have demonstrated that the onset of autoimmune responses and the development of clinical symptoms are separated in time.[1,2] For example, development of islet cell autoantibodies frequently precedes diabetes and autoantibodies recognizing citrullinated proteins (rheumatoid arthritis [RA]–specific autoantibodies) precedes the development of RA.[2] These findings indicate that either a threshold needs to be exceeded in terms of tissue damage before symptoms manifest, or that there are two distinct phases in disease development—one marked by production of a group of autoantibodies and the second by autoamplifying tissue damage. In a landmark study in SLE, Arbuckle et al. have provided important insights into this issue.[3] They analyzed sera collected from patients from the US military, who subsequently developed SLE. Interestingly, autoantibodies in SLE could be divided into two groups: *(i)* those that precede the diagnosis of SLE by several years—these included antinuclear and antiphospholipid antibodies; and *(ii)* those that occurred around the time of onset of symptoms—these included anti-Sm, anti-RNP, and to a lesser extent anti-DNA. The observation that one group of autoantibodies precedes symptoms in SLE and that another group appears coincident with the phenotype strongly suggests that the groups mark distinct events in the development of autoimmune disease. Members of the first group are likely markers of disease initiation; members of the second group are likely markers of disease propagation. The antigens targeted by the immune system in this latter phase (i.e., associated with clinical disease) are more likely to have some function in disease propagation, possibly through their possession of proinflammatory or adjuvant functions (see below).[1]

KEY CONCEPTS

Barriers to Defining Mechanisms of Human Autoimmune Disease

- Genetic and phenotypic complexity.
- Interval between initiating events and development of diagnostic phenotype.
- Challenges in quantifying human immune responses.

Development of autoimmune diseases can therefore be examined in four phases (Fig. 51.1):

1. *Susceptibility phase*—before disease, but where one or several preconditions for later initiation are satisfied. This would include impaired tolerance induction or altered immune signaling thresholds. The susceptibility phase could either be inherited or acquired and permanent or transient.
2. *Initiation phase*—before onset of clinical disease but marked by the presence of an autoimmune response (e.g., in the case of SLE—antiphospholipid antibodies).
3. *Propagation phase*—this corresponds with the onset of clinical disease, marked by propagation-specific immune responses (e.g., in the case of SLE—anti-Sm antibodies).
4. *Regulation/resolution phase*—it should also be noted that in many cases during disease propagation, immunoregulatory pathways are also activated, which may result in natural inhibition of clinical disease over time. In rare cases, these inhibitory pathways can lead to permanent resolution. This resolution phase will not be discussed further here, but its existence provides important evidence that homeostasis can be reestablished even after the amplified phenotype develops.

PHASE 1: SUSCEPTIBILITY

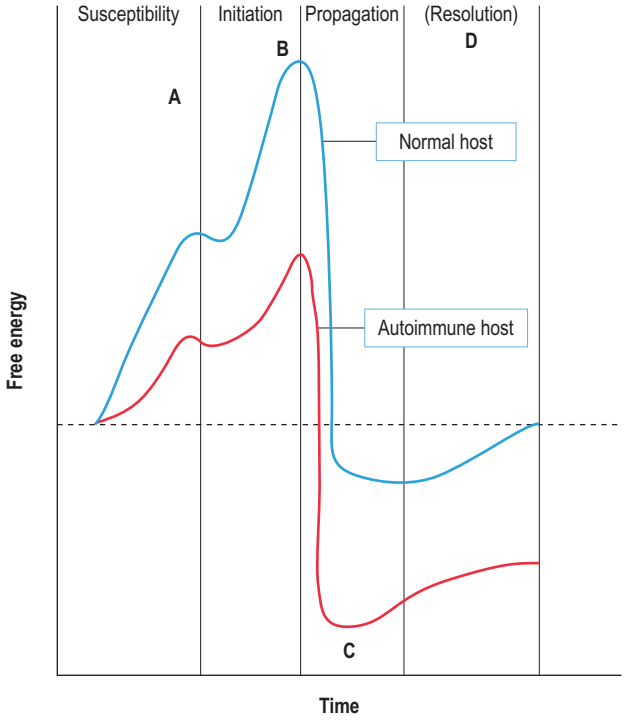
Although autoimmune diseases in humans are genetically complex, decreasing costs and expanding capacity of sequencing studies have led to the generation of large datasets from patients with a wide repertoire of autoimmune diseases. The strength of these sequencing studies has allowed for unbiased genome-wide association studies (GWAS) that have revealed several genetic risk factors for autoimmune diseases, many of which are associated with multiple diseases[4] GWAS studies have also allowed for the development of disease prediction models based on genetic susceptibility.[5]

Important advances in the genetics of autoimmunity also come from the study of autoimmunity with mendelian patterns of inheritance (e.g., autoimmune polyendocrinopathy with candidiasis and ectodermal dysplasia [APECED], immune dysfunction/polyendocrinopathy/enteropathy/X-linked [IPEX] syndrome, C1q deficiency) and from several mouse models. Collectively, these studies highlight a critical role for pathways of tolerance induction, immunoregulation, and setpoints/thresholds for immune signaling in avoiding emergence of autoimmunity.

KEY CONCEPTS

Mechanisms Underlying Susceptibility to Autoimmunity

- Incomplete induction of tolerance in the thymus to peripherally expressed autoantigens (autoimmune regulator [AIRE] deficiency causing autoimmune polyendocrinopathy with candidiasis and ectodermal dysplasia [APECED]).
- Impaired clearance and tolerance induction by apoptotic cells (e.g., deficiency of C1q, C4, milk fat globule–epidermal growth factor [EGF] 8 [MFG-E8], Mer).
- Defective production of regulatory T cells (FOXP3 deficiency causing immune dysfunction/polyendocrinopathy/enteropathy/X-linked [IPEX] syndrome).
- Altered immune signaling thresholds (e.g., cytotoxic T lymphocyte antigen-4 [CTLA-4] polymorphisms, protein tyrosine phosphatase, non–receptor type 22 [PTPN22] polymorphisms).
- Infection and dysbiosis.



A	**Susceptibility**
• Impaired tolerance induction • Impaired production of regulatory T cells • Altered immune signaling thresholds	
B	**Initiation**
• Suprathreshold concentration of autoantigens • Non-tolerized structure • Pro-immune context — infection, malignancy, exposure to adjuvants	
C	**Propagation**
• Acquisition of adjuvant properties by disease-specific autoantigens • Increased autoantigen expression in the target tissue • Immune effector pathways generate/expose autoantigen, which further drives the immune response	

FIG. 51.1 Mechanisms of Autoimmunity. Autoimmune diseases result from a complex interplay of pathways and events that allow autoreactivity to manifest and cause self-sustaining tissue damage. Mechanistically, it is useful to divide the process into three phases: (A) susceptibility phase—this is present before disease and is the phase in which one or several preconditions for later initiation are satisfied; (B) initiation phase—this is marked by the presence of autoimmunity, but precedes the diagnostic clinical phenotype; (C) propagation phase—this is marked by autoimmunity and tissue damage, in which immune effector pathways cause damage and provide antigen to drive the ongoing immune response.

Incomplete Thymic Tolerance Induction Predisposes to Autoimmunity

Insights into basic mechanisms can derive from the study of rare human phenotypes. This has been true for autoimmunity, where several monogenic disorders have defined important pathogenic principles. Autoimmune polyendocrinopathy syndrome type 1 (APS-1; also called APECED) is a rare disease in which patients develop multiple autoimmune diseases, often beginning in childhood. Although candidiasis and ectodermal dystrophy (including involvement of enamel and nails, as well as keratopathy) are features of the disease, the syndrome is characterized by striking autoimmunity directed against multiple different target tissues. Autoimmune processes include autoimmune hypoparathyroidism, Addison disease, autoimmune gastritis with pernicious anemia, type I DM, thyroid disease, autoimmune hepatitis, celiac disease, and gonadal failure. Numerous autoantigens have been defined as targets of autoimmunity in APS-1, including enzymes specifically expressed in various endocrine tissues. The genetic basis of APS-1 was mapped to a gene on chromosome 21q22.3, subsequently termed *AIRE* (for autoimmune regulator). *AIRE* expression is highest in the thymus, where it is expressed in medullary thymic epithelial cells. Several predicted structural features of the AIRE protein and its localization in nuclear dots suggested that the protein might be a transcriptional regulator, and significant evidence for this proposal was obtained in vitro. Several *AIRE*-deficient mouse models were subsequently generated, which allowed for the definition of important pathogenic pathways in APS-1 that may be broadly relevant to the mechanisms of autoimmunity in general. Mice deficient in *AIRE* developed various autoimmune phenotypes, resembling those found in human APS-1. These included multiorgan lymphocytic infiltration and autoantibodies, as well as autoimmune eye disease. In an elegant series of experiments, Anderson et al. demonstrated that AIRE regulates expression in thymic epithelial cells of various peripheral autoantigens normally expressed exclusively in endocrine target tissues.[6] Thus, AIRE appears to regulate the ectopic expression in the thymus of tissue-restricted autoantigens and to provide an antigen source against which to establish central tolerance.[6] In a recent study, phage display immunoprecipitation and sequencing (PhIP-seq) was used to define novel autoantigens in patients with APS-1. Interestingly, the defined autoantigens had tissue-specific expression, and autoantibodies were associated with tissue-specific autoimmune phenotypes in APS1, providing novel diagnostic tools and evidence that the proposed mechanism is operative in the human syndrome.[7]

Impaired Clearance and Tolerance Induction: Susceptibility Defect in Systemic Autoimmunity

Although little is known in humans about the cell biology of thymic pathways of tolerance induction to ubiquitously expressed autoantigens, there is accumulating evidence to suggest that in the periphery, apoptotic cells play an important role in providing a source of autoantigens against which the organism becomes tolerant.[1] Apoptotic cells are generally very efficiently cleared by phagocytic cells; these events are normally associated with the production of anti-inflammatory cytokines and result in tolerance induction.[1] Interestingly, early components of the classical complement pathway (e.g., C1q and C4) and cross-reactive protein (CRP) are required for efficient apoptotic cell clearance. Therefore it is of particular note that homozygous C1q deficiency is associated with a striking susceptibility to SLE in humans, and this suggests that efficient, tolerance-inducing clearance of apoptotic cells may play a role in preventing subsequent emergence of autoimmunity to ubiquitously expressed autoantigens.[8] Additional support for this model comes from recent studies of milk fat globule–epidermal growth factor (EGF) 8 (MFG-E8), a glycoprotein secreted from macrophages that is required for the efficient attachment and clearance of apoptotic cells by macrophages and immature dendritic cells (DCs). MFG-E8 is expressed in tingible-body macrophages in the germinal centers of secondary lymphoid tissues. Interestingly, many unengulfed apoptotic cells are present in the germinal centers of the spleen in *MFG-E8*–deficient mice, which develop a striking lupus-like phenotype (reviewed in Rai and Wakeland[9]).

Netosis—the process of neutrophil extracellular traps (NET) release—is another form of cellular death that results in antigen presentation. In this process, neutrophils release extracellular fibers, composed of DNA and proteins, that trap and prevent the spread of extracellular pathogens (reviewed in Papayannopoulos[10]). In this manner, neutrophils prevent pathogen spread in a proinflammatory context, which recruits further immune responses to the site of infection. Thus, the formation of NETs, although only recently recognized, is an important component of the innate immune response. However, during netosis, self-antigens are also presented. Known targets of the immune response in SLE, RA, and vasculitis have all been identified in NETs from patients with these conditions. Additionally, elevated NETs or fragments of NETs have been observed in the serum of patients with rheumatic diseases. Therefore, factors that contribute to enhanced netosis may also impair peripheral tolerance by allowing for autoantigen presentation in a highly proinflammatory setting. Collectively, these data strongly suggest that efficient, anti-inflammatory clearance of NETs and apoptotic cells plays a central role in tolerance induction and prevention of autoimmunity.

Defective Production of Regulatory T Cells

Although there are pathways that (i) regulate autoantigen expression at sites of tolerance induction and (ii) guide autoantigens toward tolerance-inducing outcomes, these pathways alone are clearly insufficient to prevent the emergence of autoimmune disease. This fact is highlighted by the emergence of autoimmunity when regulatory T-cell (Treg) differentiation is abnormal in humans with IPEX syndrome, the human equivalent of the *scurfy* mouse. IPEX is a rare, X-linked recessive disorder characterized by type I DM, thyroiditis, atopic dermatitis, and inflammatory bowel disease (IBD) and is caused by mutations in the *FOXP3* gene.[11] FOXP3 is a member of the forkhead family of transcription factors and is essential for the development of $CD4^+$ Tregs, which have been shown to regulate the activation and differentiation of effector T cells.

Decreases in numbers and functionality of Tregs have been detected in some autoimmune conditions. These cells likely play important roles in regulating disease onset and amplitude, and therapeutic interventions focusing on Tregs are under investigation in numerous clinical trials (reviewed in Romano et al.[12]).

Signaling Thresholds and Susceptibility to Autoimmunity

Several modulators of T-cell signaling, including CTLA-4 and PD-1, have been defined as important susceptibility determinants in autoimmunity.[9] For example, cytotoxic T-lymphocyte antigen-4 (CTLA-4) polymorphisms are associated with increased risk of a variety of autoimmune diseases, including type I DM, Graves disease, and RA. Similarly, a functional polymorphism in protein tyrosine phosphatase—non–receptor type 22 (PTPN22)—has been identified as a major risk factor for several human autoimmune diseases, including SLE, RA, and type I DM. Although the exact mechanisms underlying susceptibility to autoimmunity remain unclear, these polymorphisms appear to regulate the balance of stimulatory and inhibitory signaling in effector and Tregs, favoring effector T-cell activation. These findings have recently been underscored by the introduction of checkpoint inhibitor therapies.

In 2013, "Cancer Immunotherapy" was named the Science Breakthrough of the Year. Since then, multiple checkpoint inhibitor therapies have been developed for treatment of cancer. By inhibiting the negative regulation of immune cells, checkpoint inhibitor therapy alters signaling thresholds, leading to the activation of effector cells and restricting the function of Tregs (Chapter 80). Thus, checkpoint inhibitor therapy has harnessed the power of the immune system to kill cancer cells that had previously evaded the immune system through tumor- and microenvironment-mediated tolerance. This major contribution to the field of medicine has also emphasized the importance of signaling thresholds in the development of autoimmunity. Indeed, one of the major complications of the use of checkpoint inhibitor therapies is the concurrent development of various autoimmune diseases collectively called immune-related adverse events (irAEs), which could inform on autoimmune diseases (recently reviewed in Ramos-Casals et al.[13]). For example, in patients treated with ipilimumab—a CTLA4 inhibitor—increasing numbers of effector T cells as compared to Tregs are observed as well as an increase in the number of interleukin-17 (IL-17)–expressing cells, potentially driving autoimmunity. Inhibition of the PD-1 pathway also skews the immune response to a more aggressive effector response, often directed at neoepitopes generated by somatic mutation, releasing a prominent anti-cancer effect and proinflammatory side effects.[14] Interestingly, pre-existing autoimmunity does not preclude successful treatment with checkpoint inhibitor therapy and some irAEs resolve after cessation of checkpoint inhibitor therapy. Systemic autoimmune disease in patients treated with checkpoint inhibitors indicates that immune recognition of autoantigens is the rule and not the exception and that antigen recognition is only one aspect of turning protective into pathogenic immune responses.

Recent genetic studies have also suggested a potential role for innate immune sensors in autoimmunity, focusing attention on the critical balance between activation of the immune response to mitigate infectious damage and limiting the magnitude of the response to avoid immunopathology. Gain-of-function (GOF) variants of *IFIH1* (encoding melanoma differentiation-associated protein 5 [MDA5]) are associated with susceptibility to SLE.[15] This RNA helicase is essential for the detection of cytoplasmic viral RNA and activation of type I interferon (IFN) secretion by infected cells. Increased susceptibility to SLE with increased activity of this antiviral pathway suggests that excessive IFN signaling may facilitate the development of autoimmunity. Key immune signaling pathways that protect the host from deleterious infectious and malignant challenges but also enhance tissue damage in the process may, therefore, be important susceptibility factors in autoimmune diseases.

Infections and the Microbiome

Association studies have suggested that both the microbiome and infections (reviewed in Christen[16]) can influence the development of autoimmune diseases. Some infections are associated with specific autoimmune diseases, e.g., herpes simplex virus and NMDAR-encephalitis, *Helicobacter pylori* (HP) and SLE, H1N1 influenza and narcolepsy, *Campylobacter jejuni* and Guillain-Barré syndrome (GBS), and periodontal infections and RA. Dysbiosis may also influence autoimmunity, through several proposed mechanisms, and remains an area of active investigation (reviewed in Christen[16]). Yet, many of the microbes associated with autoimmunity are typically harmless or have limited disease courses and infect a great number of people without giving rise to autoimmunity. For example, 46% of the adult population in the United States has periodontitis, with 8.9% having severe disease, yet only approximately 0.6% of the population develops RA.[17] Although data indicate that, in some people, infections may increase susceptibility to autoimmunity, it remains unclear as to whom or how. Until recently, few studies have highlighted the molecular mechanisms potentially driving these relationships.

Recent work identified a role for microbial translocation in the development of autoimmune disease. Here, the authors demonstrated that translocation of the intestinal bacteria, *Enterococcus gallinarum*, could induce autoimmune responses in a murine model of SLE.[18] Importantly, antibiotic treatment that targeted *E. gallinarum* or vaccination against these bacteria decreased mortality and suppressed the development of autoantibodies and autoreactive T cells. This study suggests that intestinal infection alone was insufficient to increase the risk of autoimmunity, but infection of internal organs after bacterial translocation was. These findings may help explain some of the disparity between infection rates and autoimmune disease occurrence.

Microbes may also directly influence the generation of autoantigens. Evidence for this comes from a study investigating microbial factors that contribute to the development of RA. Konig et al. showed that a pore-forming protein produced by *Aggregatibacter actinomycetemcomitans*— one strain of bacteria that was found in the examined periodontal infections—induced protein citrullination in neutrophils.[19] This is a critical observation, as hypercitrullination of autoantigens in the RA joint by a similar mechanism has been described.[1] Further, the authors demonstrated that infection by this microbe was associated with the presence of known autoantigens in RA.[19] Collectively, these studies provide mechanistic links between infection and autoimmune disease and implicate these bacteria not only in increasing susceptibility to autoimmunity but also in disease initiation, discussed in the following section.

PHASE 2: INITIATION

Initiation of an adaptive immune response requires presentation to T cells of suprathreshold concentrations of molecules with structures not previously tolerized by the host. The immunodominance of T-cell epitopes is one model proposed that explains the existence of potentially autoreactive T cells. This model provides major insights into the pathogenesis of autoimmunity.[9,20]

KEY CONCEPTS

Potential Mechanisms That Can Alter Antigen Processing to Reveal Potentially Cryptic Epitopes

- Modification of autoantigen processing through high-affinity binding to ligands or antibodies.
- Distinct proteolytic machinery in the thymus and periphery—or differential modification of proteolytic activity.
- Modification of autoantigen structure that modifies its processing by endogenous antigen-presenting cell (APC) machinery, generally through posttranslational modifications.
- Novel proteolytic events not present in the normal APC pathways (e.g., novel cleavage during cell death or damage or inflammation).
- Novel forms of autoantigens generated by mutation, truncation, or splicing.

Dominance and Cryptticity

Studies by Sercarz et al.[20] have stressed that, although antigens contain numerous potential determinants that could be presented on major histocompatibility complex (MHC) class II molecules during antigen processing, not all determinants are equally likely to be presented. Those determinants that are most efficiently presented are termed *dominant*; those that are not recognized to a significant degree are termed *cryptic*. For self-antigens, it is likely that a constant set of dominant determinants are generated during antigen processing under most circumstances, with similar outcomes in the thymus and periphery. Antigens processed by the "standard" pathway are, therefore, fully tolerized, with the T-cell repertoire purged of reactivity to the dominant self. However, the balance of dominant and cryptic epitopes presented during antigen processing is influenced significantly by changes of protein structure, which occur during various relevant physiological states.[21] Several potential mechanisms that may alter antigen processing to reveal potentially cryptic epitopes are summarized below.

High-Affinity Binding of Antigen to Ligands or Antibodies

Several studies have demonstrated that antigen processing can be dramatically altered when the antigen binds with high affinity to a ligand or antibody. The study by Simitsek and colleagues (reviewed in Lanzavecchia[21]) demonstrated that presentation of T-cell determinants in tetanus toxin can be either enhanced or suppressed as a direct consequence of antibody modulation of antigen processing in human B-lymphoblastoid cells. Remarkably, a single bound antibody can simultaneously enhance the presentation of one T-cell determinant by more than 10-fold while strongly suppressing the presentation of a different T-cell determinant. Biochemical analyses have shown that both the suppressed and boosted determinants fall within an extended domain of antigen stabilized by this antibody during proteolysis. Thus, ligand-induced changes in processing can destroy dominant determinants or reveal cryptic self determinants. Similar observations have also been made with numerous other antigen–antibody partners.[21]

Tissue-Specific Protease Expression

The study by Watts and associates (reviewed in Rosen and Casciola-Rosen[1]) showed that a principal human leukocyte antigen–D related type 2 (HLA-DR2)–restricted epitope in myelin basic protein amino acids 85–99 (MBP85–99) contains a processing site for asparagine endopeptidase (AEP), with cleavage by AEP abolishing the epitope. AEP, which is abundantly expressed in the thymus, is therefore a critical factor in presentation of this epitope.[1] In human antigen-presenting cells (APCs), presentation of MBP85–99 is inversely proportional to the amount of cellular AEP activity, and inhibition of AEP greatly enhances presentation of the MBP85–99 epitope. Thymus-specific serine protease (TSSP) is also expressed in the thymus and also limits the expression of self-antigens, thereby reducing negative selection. In murine models of MS, the elimination of TSSP resulted in a decrease in autoreactive T cells and disease severity.[22] Collectively, these data suggest that major epitopes in neurological autoimmunity may not be presented under normal circumstances in the thymus as a result of destruction by proteases, therefore raising the potential for later presentation in the periphery in the setting of decreased enzymatic activity.

Posttranslational Modification of Autoantigens

Autoantigens undergo a variety of posttranslational modifications, including phosphorylation, proteolytic cleavage, ubiquitination, transglutamination, citrullination, and isoaspartyl modification.[23] In several cases, autoantibodies recognize exclusively the modified form of the antigen (e.g., citrullinated vimentin), indicating that the modified forms of the molecules are important in driving the immune response. In addition, Doyle and Mamula[23] have demonstrated that posttranslational modification of autoantigen structure may be more broadly relevant than can be appreciated by studying autoantibody specificity alone. They showed that, although mouse immunization with a murine cytochrome c peptide (amino acids 90–104) resulted in no T- or B-cell response, immunization with the isoaspartyl form of this peptide resulted in strong T- and B-cell responses. The autoantibodies that were elicited recognized both the modified and the native forms of the antigen, but T cells only recognized the isoaspartyl form. Similar observations have also been made for several SLE autoantigens. The difficulty detecting and quantifying antigen-specific T cells in various autoimmune diseases may reflect their preferential recognition of subtly modified forms of autoantigen.

Novel Antigen Cleavage During Cell Damage, Cell Death, or Inflammation

Recent studies have provided evidence that single proteolytic events early in antigen processing can play critical roles in defining the epitopes generated. For example, Watts and colleagues (reviewed in Rosen and Casciola-Rosen[1]) have demonstrated that early cleavage by AEP determines subsequent proteolytic events. Antigen modifications that affect this early cleavage dramatically change the epitopes loaded onto MHC class II molecules.[1]

Inflammatory microenvironments can create significant potential to load distinct epitopes because unique proteolytic activities are present. Activated inflammatory cells constitute a major source of proteases, including various cytotoxic lymphocyte granule proteases (granzymes) as well as numerous neutrophil and monocyte granule proteases. It is of interest that numerous autoantigens targeted in systemic autoimmune diseases are substrates for these inflammatory proteases and that unique autoantigen fragments are generated through the activity of granzyme B and potentially other similar proteases.[1] This was recently demonstrated within the context of RA. In this study, Darrah and colleagues demonstrated that proteolysis of the autoantigen PAD4 by granzyme B induced conformational

changes in PAD4, which altered the peptides presented. Importantly, these peptides were recognized by PAD4-specific T cells in patients with RA.[24] Such autoantigen forms are not generated during other forms of cell damage or death, and similar activity is not observed against non-autoantigens. Thus, novel proteolytic cleavage of intracellular autoantigens during activity of cytotoxic immune effector pathways may provide a source of cryptic epitopes not generated during homeostatic "tolerance-inducing" tissue turnover.

Autoantigen Alteration Caused by Mutation, Truncation, or Splicing

Since the final epitopes generated and loaded onto MHC class II molecules can be profoundly influenced by single, early cleavage events during antigen processing, relatively minor but critically placed changes in the primary structure of autoantigens can have the capacity to influence peptide selection. A study of the melanoma and vitiligo-associated autoantigens, tyrosinase-related proteins (TRPs) 1 and 2 has demonstrated that this mechanism may play an important role in generating immune responses to self- and tumor antigens (reviewed in Rosen and Casciola-Rosen[1]). This study examined whether mutated self gene products are more likely to initiate immunity and used a systematic approach to define some of the principles that determine this. The investigators generated complementary DNA (cDNA) libraries encoding large numbers of random mutations in syngeneic TRPs. They then used an approach of DNA immunization of black mice to test the immunogenicity of the altered proteins encoded by the pools of mutated cDNA. Immunization with nonmutated proteins induced no detectable immune responses, consistent with establishment of tolerance to the full-length molecules. In contrast, the mutated cDNA pools elicited both autoimmune depigmentation and the ability to reject melanoma tumors. Additional analysis showed that autoimmunity resulted from mutations that altered autoantigen cell biology, particularly degradation rates and pathways. Mutations also created new T-helper (Th) cell epitopes and induced recognition of nonmutated but previously cryptic epitopes. Interestingly, mutations themselves did not form part of CD8 epitopes that drive anti-self and antitumor immune responses. Mutated molecules that were immunogenic were frequently truncated, leading the authors to propose that inappropriately truncated self proteins can provoke autoimmunity when present in a proinflammatory environment. This study has provided an important mechanistic underpinning for the proposal that accumulated mutations have a role in the initiation of autoimmunity and that "autoimmunity" might target the cancer mutanome. As further evidence for this hypothesis, investigations into the immunity induced by checkpoint blockade demonstrate a preferential targeting of the mutanome.[25,26]

Recent work in the autoimmune rheumatic diseases (scleroderma and dermatomyositis) highlights the relationship between cancer and the autoimmune process. In these diseases, patients demonstrate an increased risk of cancer and a temporal clustering of cancer diagnosis around the time of dermatomyositis or scleroderma onset (reviewed in Rosen and Casciola-Rosen[1]). Similarly, there is evidence for an association between SLE and cancer, particularly lymphoma, clustered within the first 2 years of SLE diagnosis. These associations—both with timing of diagnosis and preferentially with specific tumor types—are strongly indicative of a nonrandom clustering of autoimmune processes and cancer, which is likely of mechanistic significance.

A mechanistic link between scleroderma and cancer was confirmed in a study that tested the hypothesis that an anticancer immune response may target a mutated autoantigen in the patient's cancer, which spreads to the wild-type version of the antigen to trigger a self-sustaining autoimmune disease (reviewed in Rosen and Casciola-Rosen[1]). In this study of patients with scleroderma and cancer, genetic changes (either mutations or loss of heterozygosity) were identified in *POLR3A* (encoding RNA polymerase III large subunit) in the tumors of six of eight patients with anti-RNA polymerase III antibodies, whereas no such changes were identified in eight patients who lacked these antibodies. Anti-RNA polymerase III autoantibodies in patients with cancer demonstrated no specificity for the mutant form of the protein over the wild-type version, and analyses of peripheral blood lymphocytes identified T-cell reactivity against the mutant protein. These data suggest that somatic mutations arising in the context of cancer may prompt immune responses that mediate immunoediting as well as damage to nontumor host tissues.

Therefore, it is likely that somatic mutations acquired with age and their association with malignancy are important in the genesis of some forms of human autoimmunity. Additional studies to confirm this and elucidate the underlying mechanisms remain a high priority. However, the barriers to such studies in humans are very significant, as effective anticancer immunity may be phenotypically silent, and convenient technologies to quantify somatic mutation and specific immune responses in normal individuals will be needed to draw conclusions of causality.

Antigen Mimicry

Foreign antigens often differ structurally from their homologous self-antigens yet may bear significant similarity in focal regions. Initiation of an immune response to the foreign antigen may generate a cross-reactive antibody response that also recognizes the self protein. This process, known as antigen mimicry, has frequently been proposed as a potential initiator of autoimmune diseases.[27] When the antigen is a cell surface molecule, antibody-mediated effector pathways can lead to host tissue damage. Although the antibody response is cross-reactive with self molecules, the T cells that drive this response are generally directed at the foreign antigen. Diseases involving this sort of "antigen mimicry," therefore, tend to be self-limiting, although they can recur upon reexposure to the offending antigen. It is important to realize that antigen mimicry alone cannot explain self-sustaining autoimmune diseases, which are driven by self-antigens and autoreactive T cells. In these cases, there is a requirement for overcoming T-cell tolerance to the self protein. The central issues in this regard are *(i)* how T-cell tolerance to self-antigens might initially be broken, and *(ii)* once this has occurred, why these antigens continue to drive the immune response to self. Several studies have suggested that when a humoral response to a foreign protein is induced that cross-reacts with the self-antigen, a strong T-cell response specific for the self-antigen can occur. The simultaneous liberation of significant amounts of self-antigen in the setting of a cross-reactive antibody response may allow for effective presentation of cryptic epitopes in the self-antigen to autoreactive T cells by activated cross-reactive B cells. If continued release of self-antigen occurs, a specific, adaptive immune response to self will be sustained. Antigen release from tissues likely plays a critical role in driving this autoimmune process. Understanding the unique

pathways of cell injury and death that result in ongoing antigen release at sites of tissue damage in autoimmune disease is a high priority for future work.

The simultaneous confluence of susceptibility factors and initiation forces required to induce an autoimmune disease is a rare occurrence. In contrast, once activation of autoreactive T cells has occurred, the ability of the immune system to vigorously respond to low concentrations of antigen, to amplify responses to those antigens, and to spread the response to additional antigens in that tissue, greatly reduces the stringency that must be met to keep the process going.

PHASE 3: PROPAGATION

Principles of Amplification

One of the central features of human autoimmunity is the tendency of the process to amplify progressively with the accumulation of significant immune-mediated tissue damage. Furthermore, once such amplification begins, the process is very unlikely to resolve spontaneously. Autoantigens themselves can be very important in this phase, in terms of both acquisition of adjuvant properties and regulation of levels of expression. Amplification is a cyclical process in which antigen expression and adjuvant properties induce an immune response, which, in turn, induces increased antigen expression and tissue damage—and further drives the immune response. The importance of tissue-specific autoantigen expression in focusing such immune responses is only beginning to be recognized.

Acquisition of Adjuvant Properties by Disease-Specific Autoantigens

In spite of the fact that tens of thousands of molecules could be targeted by the immune system in autoimmunity, the number of molecules that is frequently targeted in the different phenotypes is markedly restricted—limited perhaps to 100 or so. This has led to the proposal that frequently targeted autoantigens may themselves have properties that make them proimmune. Work by Howard and associates (reviewed in Rosen and Casciola-Rosen[1]) provided important initial support for this proposal. They observed that the autoantigenic histidyl aminoacyl tRNA synthetase (HRS), which is targeted in autoimmune myositis (but not non–autoantigenic lysyl- and aspartyl-aminoacyl transfer RNA [tRNA] synthetases), is a chemoattractant to immature DCs and other leukocytes, which, in turn, may enhance its autoantigenicity. Additional support for this hypothesis comes from work demonstrating that antibodies against PAD2, an enzyme crucial in the formation of autoantigens in RA, are associated with a less severe phenotype.[28] These data suggest that a subset of the autoantigens targeted in autoimmune diseases may reflect an attempt by the immune system to absorb an excess of damaging, proinflammatory molecules.

Role of Innate Immune Receptors in Amplification

Innate immune receptors, including Toll-like receptors (TLRs), cyclic GMP-AMP synthase, and RIG-I, sense and transduce the signals from pathogen-associated molecular patterns (PAMPs). These receptors may also play roles in transducing the proinflammatory properties of autoantigens—the best understood are the TLRs.[1]

Ligands for TLRs are typically microbial. However, endogenous molecules, including complexes containing nucleic acids from stressed, injured, and dying cells (reviewed in Rosen and Casciola-Rosen[1]), are also able to activate TLRs with differences in adjuvant activity. For example, bacterial and viral DNA and oligonucleotides with CpG motifs have significant adjuvant activity, whereas mammalian genomic DNA, in which CpG is usually methylated, has very poor adjuvant activity. In contrast, human DNA within immune complexes in SLE serum effectively activate plasmacytoid dendritic cells (pDC) in a DNA-dependent way. Several potential explanations have been advanced to explain these observations. First, FcγR-mediated uptake effectively captures self DNA bound by anti-DNA antibodies and directs it to the correct endosomal compartment for TLR signaling. Second, co-ligation of TLR9 and either B-cell receptor or FcγR alters the signaling threshold of immune complexes. Last, the difference lies in the nucleic acid itself, with additional modifications of DNA and RNA structure occurring in cells under different physiological circumstances (e.g., cell death) and regulating nucleic acid binding to TLRs.[1]

The TLR–IFN interface has been recognized as critical in the propagation phase of systemic autoimmune diseases (reviewed in Hall and Rosen[29]). Type I IFNs have a broad set of functions that likely contribute to the propagation phase of systemic autoimmune diseases. For example, they (i) promote the differentiation of monocytes into mature DCs, which drive autoreactive T- and B-cell responses; (ii) increase target cell sensitivity to killing pathways; (iii) upregulate cytotoxic effector pathways; and (iv) upregulate expression of autoantigens.

This ability of autoantigens, particularly in the context of immune complexes, to stimulate secretion of IFN and other cytokines is likely an important principle in the initiation and propagation of autoimmunity. For example, Rönnblom and colleagues (reviewed in Hall and Rosen[29]) demonstrated that when added to material from apoptotic or necrotic cells, autoantibodies from patients with SLE and Sjögren syndrome (SS) with specificity for DNA or RNA autoantigens induce striking IFN secretion. Further, studies in mouse models have demonstrated that the TLR–IFN axis can alternately promote wound healing or chronic inflammation, depending on the predisposing genetic background. For example, nonspecific skin injury in wild-type mice (by tape stripping) leads to pDC recruitment, TLR7 and TLR9 recognition of nucleic acids, and transient expression of type I IFNs with subsequent wound healing. In contrast, the same skin damage in a lupus-prone mouse strain results in chronic inflammation mediated by sustained type I IFN expression, which can be ameliorated with pDC depletion or TLR7/9 inhibition (reviewed in Rosen and Casciola-Rosen[1]).

Enhanced Autoantigen Expression in the Target Tissue

The association of specific autoantibody responses with distinct phenotypes suggests that autoantigen expression or conformation in target tissues may play an important role in both focusing the immune response and generating tissue damage. For example, recent work demonstrated that within the labial salivary gland, the double-stranded DNA sensor IFN-inducible protein 16 (IFI16) was in an activated, filamentous form.[30] Importantly, antibodies from patients with SS preferentially recognized filamentous IFI16 as compared to inactivated IFI16. These results demonstrate that, within the target tissue, the autoantigen undertook a unique conformational form, which enhanced its immune recognition.

Likewise, enhanced autoantigen expression has been demonstrated in target tissues. For example, myositis-specific autoantigens are expressed at very low levels in control muscle

but at high levels in myositis tissue. Indeed, antigen expression is highest in regenerating muscle cells.[1] Further supporting this hypothesis, expression of 3-hydroxy-3-methylglutaryl-coenzyme A reductase (HMGCR)—an autoantigen in patients with statin-associated autoimmune myopathy—was found to be induced by statins in vitro and enhanced in regenerating muscle fibers of anti–HMGCR-positive patients (reviewed in Rosen and Casciola-Rosen[1]). A recent case control study has also shown that the autoantigen topoisomerase-1 has enhanced expression in the lung tissues of patients with autoimmune inflammatory interstitial lung disease. Collectively, these data suggest that enhanced autoantigen expression in the target tissue and during tissue repair may provide an ongoing antigen source to sustain and amplify tissue damage.

ON THE HORIZON

- Precise clinical and molecular phenotyping of patients at various disease stages is critical for improving diagnosis, monitoring, and treatment of autoimmune diseases.
- Understanding mechanisms of disease amplification, propagation, and regulation will enable the development of effective targeted therapies.
- Understanding the mechanisms of human autoimmunity will likely provide important insights into the normal functioning of the immune system, particularly with regard to natural cancer immunity.

TRANSLATIONAL RESEARCH

Critical areas of future investigation include clarification of the roles of genetic, epigenetic, and environmental factors in disease susceptibility and initiation, understanding the role of the target tissue to ongoing disease amplification, and elucidating mechanisms of regulation of disease amplitude or disease resolution. Collectively, these investigations may provide important opportunities for interventions, including antigen-specific therapies. Further investigations into the molecular underpinnings of the distinct phases in the development of autoimmune disease will provide additional opportunities to identify patients at risk of developing tissue damage and to interdict the process before the amplification cycle is established. Precise biomarker identifications of all events in the development of autoimmunity requires exceptional clinical phenotyping of patients early in the disease course, longitudinal analyses that follow large numbers of patients, and effective coupling to basic laboratory enterprise.

REFERENCES

1. Rosen A, Casciola-Rosen L. Autoantigens as partners in initiation and propagation of autoimmune rheumatic diseases. *Annu Rev Immunol*. 2016;34:395–420.
2. Dekkers J, Toes RE, Huizinga TW, van der Woude D. The role of anticitrullinated protein antibodies in the early stages of rheumatoid arthritis. *Curr Opin Rheumatol*. 2016;28(3):275–281.
3. Arbuckle MR, McClain MT, Rubertone MV, et al. Development of autoantibodies before the clinical onset of systemic lupus erythematosus. *N Engl J Med*. 2003;349(16):1526–1533.
4. Long T, Hicks M, Yu HC, et al. Whole-genome sequencing identifies common-to-rare variants associated with human blood metabolites. *Nat Genet*. 2017;49(4):568–578.
5. Marigorta UM, Denson LA, Hyams JS, et al. Transcriptional risk scores link GWAS to eQTLs and predict complications in Crohn's disease. *Nat Genet*. 2017;49(10):1517–1521.
6. Anderson MS, Venanzi ES, Klein L, et al. Projection of an immunological self shadow within the thymus by the aire protein. *Science*. 2002;298(5597):1395–1401.
7. Vazquez SE, Ferre EM, Scheel DW, et al. Identification of novel, clinically correlated autoantigens in the monogenic autoimmune syndrome APS1 by proteome-wide PhIP-Seq. *Elife*. 2020:9.
8. Macedo AC, Isaac L. Systemic lupus erythematosus and deficiencies of early components of the complement classical pathway. *Front Immunol*. 2016;7:55.
9. Rai E, Wakeland EK. Genetic predisposition to autoimmunity—what have we learned? *Semin Immunol*. 2011;23(2):67–83.
10. Papayannopoulos V. Neutrophil extracellular traps in immunity and disease. *Nat Rev Immunol*. 2018;18(2):134–147.
11. Bennett CL, Christie J, Ramsdell F, et al. The immune dysregulation, polyendocrinopathy, enteropathy, X-linked syndrome (IPEX) is caused by mutations of FOXP3. *Nat Genet*. 2001;27(1):20–21.
12. Romano M, Fanelli G, Albany CJ, et al. Past, Present, and future of regulatory T cell therapy in transplantation and autoimmunity. *Front Immunol*. 2019;10:43.
13. Ramos-Casals M, Brahmer JR, Callahan MK, et al. Immune-related adverse events of checkpoint inhibitors. *Nat Rev Dis Primers*. 2020;6(1):38.
14. Topalian SL, Drake CG, Pardoll DM. Immune checkpoint blockade: a common denominator approach to cancer therapy. *Cancer Cell*. 2015;27(4):450–461.
15. Oliveira L, Sinicato NA, Postal M, Appenzeller S, Niewold TB. Dysregulation of antiviral helicase pathways in systemic lupus erythematosus. *Front Genet*. 2014;5:418.
16. Christen U. Pathogen infection and autoimmune disease. *Clin Exp Immunol*. 2019;195(1):10–14.
17. Gabriel SE, Michaud K. Epidemiological studies in incidence, prevalence, mortality, and comorbidity of the rheumatic diseases. *Arthritis Res Ther*. 2009;11(3):229.
18. Manfredo Vieira S, Hiltensperger M, Kumar V, et al. Translocation of a gut pathobiont drives autoimmunity in mice and humans. *Science*. 2018;359(6380):1156–1161.
19. Konig MF, Abusleme L, Reinholdt J, et al. *Aggregatibacter actinomycetemcomitans*-induced hypercitrullination links periodontal infection to autoimmunity in rheumatoid arthritis. *Sci Transl Med*. 2016;8(369): 369ra176.
20. Sercarz EE, Lehmann PV, Ametani A, et al. Dominance and crypticity of T cell antigenic determinants. *Annu Rev Immunol*. 1993;11:729–766.
21. Lanzavecchia A. How can cryptic epitopes trigger autoimmunity? *J Exp Med*. 1995;181(6):1945–1948.
22. Serre L, Girard M, Ramadan A, et al. Thymic-specific serine protease limits central tolerance and exacerbates experimental autoimmune encephalomyelitis. *J Immunol*. 2017;199(11):3748–3756.
23. Doyle HA, Mamula MJ. Posttranslational modifications of self-antigens. *Ann N Y Acad Sci*. 2005;1050:1–9.
24. Darrah E, Kim A, Zhang X, et al. Proteolysis by granzyme B enhances presentation of autoantigenic peptidylarginine deiminase 4 epitopes in rheumatoid arthritis. *J Proteome Res*. 2017;16(1):355–365.
25. Le DT, Durham JN, Smith KN, et al. Mismatch repair deficiency predicts response of solid tumors to PD-1 blockade. *Science*. 2017;357(6349): 409–413.
26. Rizvi NA, Hellmann MD, Snyder A, et al. Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science*. 2015;348(6230):124–128.
27. Cusick MF, Libbey JE, Fujinami RS. Molecular mimicry as a mechanism of autoimmune disease. *Clin Rev Allergy Immunol*. 2012;42(1):102–111.
28. Darrah E, Giles JT, Davis RL, et al. Autoantibodies to peptidylarginine deiminase 2 are associated with less severe disease in rheumatoid arthritis. *Front Immunol*. 2018;9:2696.
29. Hall JC, Rosen A. Type I interferons: crucial participants in disease amplification in autoimmunity. *Nat Rev Rheumatol*. 2010;6(1):40–49.
30. Antiochos B, Matyszewski M, Sohn J, Casciola-Rosen L, Rosen A. IFI16 filament formation in salivary epithelial cells shapes the anti-IFI16 immune response in Sjogren's syndrome. *JCI Insight*. 2018;3(18).

Systemic Lupus Erythematosus

Cynthia Aranow, Betty Diamond, and Meggan Mackay

Systemic lupus erythematosus (SLE) is a systemic autoimmune disease characterized by the production of autoantibodies and a broad spectrum of clinical manifestations. It most commonly presents in women during their childbearing years. Although the etiology of SLE is unknown, both genetic and environmental factors contribute to loss of self-tolerance. Current therapeutic modalities are antiinflammatory and immunosuppressive.

EPIDEMIOLOGY

Originally developed in 1982 and revised in 1997, the American College of Rheumatology (ACR) classification criteria for SLE (Table 52.1) have been used for diagnosis. In accordance with improved understanding of disease heterogeneity, additional classification criteria have been proposed to improve sensitivity and specificity for the diagnosis. The 2012 Systemic Lupus International Collaborating Clinics (SLICC) criteria improved sensitivity at the expense of specificity; added new mucocutaneous and neuropsychiatric manifestations, new antiphospholipid (APL) antibodies, and low complement components; and proposed that biopsy-proven lupus nephritis with either positive antinuclear antibodies (ANA) or antibodies to double-stranded DNA (dsDNA) was sufficient for diagnosis.[1] Recently, a third criteria set, the European League Against Rheumatism/American College of Rheumatology (EULAR/ACR) classification criteria has been developed and validated for use in clinical and translational SLE studies. It includes an obligatory positive ANA test followed by additive weighted clinical and immunologic criteria.[2] An important aim of refining classification criteria is to increase the sensitivity, especially for patients with early disease, to allow these patients access to early therapeutic interventions, including clinical trials.

During childbearing years, the ratio of women to men with lupus is approximately 9:1. This ratio is less in younger and older populations, supporting a role for hormonal factors in disease induction. The majority of SLE presents during adulthood; approximately 20% of cases are in the pediatric population. Recent studies suggest that age at diagnosis may be increasing in some populations; mean ages at diagnosis reported since 2002 range from 31 years in Martinique and Brazil to 51.7 years in Wisconsin (United States) and 47 years in Sweden.

Lupus occurs globally, and susceptibility is linked to race, ethnicity, and environmental exposures. Several recent population surveys[3] report prevalence in the Unites States ranging from 71 to 200 cases per 100,000. The highest prevalence was among Native American Indians, followed by African American, Hispanic, Asian, and White populations. Reported prevalence around the world is also variable; 37 to 65/100,000 in Estonia, Sweden, and Denmark, 85 to 124/100,000 in the United Arab Emirates, and the highest prevalence of 399 to 661/100,000 in Afro-Caribbeans in the United Kingdom. Although epidemiologic data are scant, it appears that the incidence of SLE in Africa is lower. Data from animal models suggest that this may be a consequence of a protective effect of malaria infection. Clinical manifestations, disease activity, damage accrual, mortality, and response to therapy are also modulated by ethnicity, and all studies support observations of increased disease activity, damage, and mortality in non-White populations (Table 52.2).[4]

Mortality and Organ Damage

Increased awareness leading to earlier diagnoses, availability, and use of immunosuppressive agents and to improved treatment for comorbid diseases has contributed to a dramatic decrease in mortality over the past 70 years, with a 5-year survival rate of 50% in the 1950s to 96% to 99% currently.[5] However, all-cause mortality for SLE patients is approximately twofold to threefold greater than for non-SLE patients and has not changed in the past 2 decades. Like disease susceptibility, mortality is also associated with sociodemographic, epidemiologic, clinical, and genetic factors. In general, higher mortality rates are associated with non-White race, Hispanic ethnicity, low socioeconomic status, childhood-onset disease, and male sex. Although mortality causes vary across regions, significant risk is associated with increased disease activity (lupus nephritis in particular), cardiovascular disease, and infection.[3,5] A bimodal distribution of death is well recognized; early deaths often result from infections or active disease, whereas deaths occurring later in the course of disease are frequently attributed to end-stage organ damage and cardiovascular disease. Hydroxychloroquine has been associated with increased survival. Other interventions that decrease mortality include vaccinations and life-style/therapeutic strategies that reduce risk for cardiovascular disease.

With improved survival, the impact of comorbid conditions and medication toxicities has become increasingly important. The ACR/SLICC Damage Index (SDI) is a validated instrument to measure damage related to disease activity, comorbid illness, and/or toxicities of medications that have accrued since SLE diagnosis. Damage scores are associated with poorer quality of life and increased morbidity and mortality. Renal, musculoskeletal, and cardiac domains are consistently the major contributors to damage.

TABLE 52.1 American College of Rheumatology Criteria for Systemic Lupus Erythematosus

Criteria	Description
Malar rash	Fixed malar erythema, flat or raised
Discoid rash	Erythematous raised patches with adherent keratotic scaling and follicular plugging; atrophic scarring may occur in older lesions
Photosensitivity	Skin rash as an unusual reaction to sunlight, by patient history or physician observation
Oral ulcers	Oral and nasopharyngeal ulcers, usually painless, observed by physician
Arthritis	Nonerosive arthritis involving two or more peripheral joints, characterized by tenderness, swelling, or effusion
Serositis	Pleuritis (convincing history of pleuritic pain or rub heard by physician or evidence of pleural effusion) *or* Pericarditis (documented by electrocardiogram or rub or evidence of pericardial effusion)
Renal disorder	Persistent proteinuria >0.5 g per day or >3+ if quantification not performed *or* Cellular casts may be red cell, hemoglobin, granular, tubular, or mixed
Neurologic disorder	Seizures—in the absence of offending drugs or known metabolic derangements (*e.g.*, uremia, ketoacidosis, or electrolyte imbalance) *or* Psychosis—in the absence of offending drugs or known metabolic derangements (*e.g.*, uremia, ketoacidosis, or electrolyte imbalance)
Hematologic disorder	Hemolytic anemia—with reticulocytosis *or* Leukopenia (<4000/mm^3 total on two or more occasions) *or* Lymphopenia (<1500/mm^3 on two or more occasions) *or* Thrombocytopenia (<100,000/mm^3 in the absence of offending drugs)
Immunologic disorder	Anti–double-stranded DNA: antibody to native DNA in abnormal titer *or* Anti-Sm: presence of antibody to Sm nuclear antigen*or* Positive finding of antiphospholipid antibodies based on (1) an abnormal serum level of immunoglobulin G (IgG) or IgM anticardiolipin antibodies; (2) a positive test for lupus coagulant using a standard method; or (3) a false-positive serologic test for syphilis known to be positive for at least 6 months and confirmed by *Treponema pallidum* immobilization or fluorescent treponemal antibody absorption test
Antinuclear antibodies	An abnormal titer of antinuclear antibody by immunofluorescence or an equivalent assay at any point in time and in the absence of drugs known to be associated with "drug-induced lupus" syndrome

IMMUNOPATHOGENESIS

The immune system is designed to protect the host against foreign pathogens and to remove cellular debris without damaging self. With their universal production of immunoglobulin G (IgG) autoantibodies and characteristic pathologic findings of inflammation, vasculopathy, and immune complex deposition, SLE patients display a failure to maintain immune tolerance and immune homeostasis. The heterogeneity of disease manifestations reflects the multiplicity of genetic, hormonal, and immune abnormalities and the diversity of environmental triggers or modifiers contributing to clinical disease (Table 52.3). Progression from initial autoreactivity to clinical disease occurs over time (Fig. 52.1); the first detectable immune abnormalities are either the presence of serum autoantibodies or the increased expression of interferon (IFN)-inducible genes in blood cells (the IFN signature).

It is now recognized that there are two dominant pathways involved in disease. The innate immune system is perturbed in ways that lead to the immunogenicity of cellular debris and the production of high levels of type I IFN. The adaptive immune system exhibits hyperactivation of B cells. These two pathways intersect. High IFN promotes enhanced B-cell activation. The failure to clear apoptotic debris leads to reactivity to cellular and especially nuclear antigens. Reciprocally, the increased production of IgG ANA leads to nucleic acid–containing immune complexes which transport nucleic acids to endosomal Toll-like receptors (TLRs) through an Fc receptor–mediated pathway. Thus active SLE exhibits perturbations in both innate and adaptive immune responses. SLE risk alleles contribute to these perturbations. Genes conferring risk for SLE are most abundantly expressed in myeloid cells and B cells. Many lead directly to the major immune abnormalities in SLE (*e.g.*, the failure in clearance of apoptotic debris, the increase in IFN production, or B-cell hyperactivity). These observations suggest that B cells and myeloid cells may be important drivers of disease. Alterations in B-cell tolerance, alterations in antigen presentation to T-effector and T regulatory cells (Tregs), and alterations in cytokine production are now recognized as central to disease pathogenesis.

Autoantibodies

ANAs are present in more than 98% of patients diagnosed with SLE. Their presence is not specific to SLE because they are observed in patients with other autoimmune diseases, malignancies, and viral (hepatitis) and parasitic (malaria) infections, as well as in response to environmental triggers such as therapeutic agents (see section on Drug-Induced Lupus, later). Furthermore, ANAs are found in low titer in 5% of the general population, with prevalence increasing with age. Common ANA specificities found in lupus patients include dsDNA, single-stranded DNA (ssDNA), extractable nuclear antigens (Sm, RNP, Ro, and La), histones, and chromatin. Specific antibody specificities are associated with disease subsets such as anti-Ro antibodies with subacute cutaneous and neonatal SLE, anti-dsDNA, and anti-C1q antibodies with renal disease.[6] Most autoantibody titers do not correlate with disease activity; anti-dsDNA antibodies

TABLE 52.2 Prevalence (%) of American College of Rheumatology Criteria in Different Ethnic Cohorts

	†Hispanic PROFILE Cohort[a] (*n* = 78)	†Hispanic Puerto Rican[b] (*n* = 134)	‡Caucasian Spain[c] (*n* = 239)	‡Caucasian USA[d] (*n* = 46)	†Caucasian PROFILE Cohort[a] (*n* = 260)	†Caucasian Norwegian[e] (*n* = 346)	†Caucasian Danish[f] (*n* = 513)	†African American PROFILE Cohort[a] (*n* = 216)	‡Chinese[d] (*n* = 175)
Malar rash	64	72	NA	24	67	40	48	45	58
Discoid rash	6	10	27	24	12	13	14	33	6
Photosensitivity	59	77	29	46	72	41	43	46	31
Oral/nasal ulcers	58	30	18	7	57	1	11	46	15
Arthritis	91	67	71	54	87	83	67	89	54
Seizure/ psychosis	12	9	6	4	9	8	13	16	9
Renal	59	30	23	54	23	17	45	54	29
Serositis	64	28	33	26	42	34	39	60	11
Cytopenias	85	77	55	83	62	36	67	82	58
Antinuclear antibody	97	93	100	83	97	99	98	97	95
Immunologic	83	NA	NA	57	65	57	98	79	81

†Cumulative data.
‡Inception data.
[a]Alarcón GS, McGwin Jr G, Perti M, et al. Baseline characteristics of a multiethnic lupus cohort: PROFILE. *Lupus*. 2002;11:95–101.
[b]Vilá LM, Mayor AM, Valentín, AH, García-Soberal, M, Vilá, S. Clinical and immunological manifestations in 134 Puerto Rican patients with systemic lupus erythematosus. *Lupus*, 1999;8:279–286.
[c]Buján S, Ordi-Ros J, Paredes, J, et al. Contribution of the initial features of systemic lupus erythematosus to the clinical evolution and survival of a cohort of Mediterranean patients. *Ann Rheum Dis*. 2003;62:859–865.
[d]Thumboo J, Uramoto K, O'Fallon, WM, et al. A comparative study of the clinical manifestations of systemic lupus erythematosus in Caucasians in Rochester, Minnesota, and Chinese in Singapore, from 1980 to 1992. *Arthritis Rheum*. 2001;45:494–500.
[e]Gilboe IM, Husby G. Application of the 1982 revised criteria for the classification of systemic lupus erythematosus on a cohort of 346 Norwegian patients with connective tissue disease. *Scand J Rheumatol*. 1999;28:81–87.
[f]Jacobsen S, Petersen J, Ullman, S, et al. A multicentre study of 513 Danish patients with systemic lupus erythematosus. I. Disease manifestations and analyses of clinical subsets. *Clin Rheumatol*. 1998;17:468–477.
NA, not available.

TABLE 52.3 Factors Contributing to Autoimmunity

- Genetic factors
- Loss of peripheral tolerance
  - B-cell abnormalities
  - T-cell abnormalities
  - Dendritic cell abnormalities
- Cytokine milieu
- Hormonal influences
- Environmental triggers

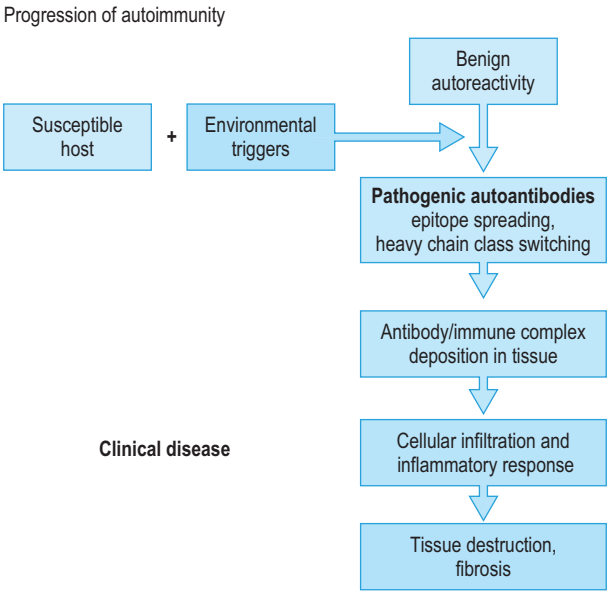


FIG. 52.1 Spectrum of Autoimmunity.

are a notable exception. Their fluctuation with disease activity suggests a pathogenic role for this autoantibody, and monitoring their titer helps to predict impending disease flare in some patients.

The Predisposed Host: Genetic Contributions

SLE is generally a multigenic disease. However, there are a few monogenic causes of SLE, such as C1q deficiency. Most disease-associated alleles are present in healthy individuals as well. Only when multiple alleles are present, along with an appropriate environmental trigger or other genes that are not identified as risk alleles, but function together with risk alleles to perturb the immune system, will a lupus-like phenotype arise. Familial disease clustering and a higher disease concordance in monozygotic than dizygotic twins suggest both an underlying genetic susceptibility and the importance of environmental or epigenetic factors.

Susceptibility genes affect lymphocyte activation, proliferation and apoptosis, cytokine production, antigen presentation, and clearance of apoptotic debris. Many of these genes are also implicated in susceptibility to other autoimmune diseases (*e.g., CTLA4* in Graves disease and type 1 diabetes, *PTPN22* polymorphisms in rheumatoid arthritis and type 1 diabetes).[7,8] Genome-wide association studies (GWASs) and genome-wide linkage analyses have utilized high-throughput techniques to study hundreds of thousands of single nucleotide polymorphisms (SNPs) in individual patients with SLE and have yielded approximately 100 SLE-associated susceptibility loci,[9] some of which are of relevance in particular ethnic and racial groups.

Antigen Presentation

Polymorphisms of major histocompatibility complex (MHC) genes determine the peptides of self and foreign antigens presented within MHC molecules that determine the naïve T-cell repertoire. Human leukocyte antigen (HLA)-DR2 haplotypes in African American, African, Taiwanese, and Korean populations and HLA-DR3 haplotypes in White populations have been associated with a twofold to threefold increased risk for developing SLE.[10] Associations between anti-Ro antibodies and HLA-DR3 and anti-La antibodies and HLA-DR25 are consistent with the concept of antigen-driven processes involving T-cell recognition. The association of HLA with SLE is not as dominant as with other autoimmune disease, suggesting that the T-cell response may dictate autospecificity (as earlier), but is less important in determining risk of disease.

Impaired Clearance of Apoptotic Debris

Patients with severe deficiencies of C2, C4, and C1q display disease risks of 10%, 75%, and 90%, respectively, for SLE.[11] Reduced uptake of apoptotic cells has been implicated in disease initiation in murine models of SLE and is seen histopathologically in lymph nodes of lupus patients.[12]

Polymorphisms of mannose-binding lectin (MBL) and C-reactive protein (CRP)—acute-phase reactants that facilitate opsonization and phagocytosis of immune complexes, apoptotic debris, and microbes—also associate with SLE susceptibility.

The Fcγ receptors—FcγR1 (CD 64), FcγRII (CD32), and FcγRIII (CD16)—have different binding affinities for IgG and immune complexes and different cell-specific expression and function. FcγRI, FcγRIIa, and FcγRIIIa and IIIb are all activating receptors. Cross-linking these receptors by immune complexes results in degranulation, phagocytosis, antibody-dependent cellular cytotoxicity, cytokine gene transcription, and release of inflammatory mediators by myeloid cells. Substitutions of one or more amino acids in the activating FcγR genes—arginine (R) for histidine (H) at position 131 in FcγRIIa and phenylalanine (F) for valine (V) at position 158 in FcγRIIIa—results in decreased affinity for IgG immune complexes. Because apoptotic debris is opsonized by antibody to promote clearance, such deficiencies may lead to immune activation by apoptotic debris. FcR polymorphisms may also predict a therapeutic response to immunobiologic agents such as rituximab.[13,14] In contrast to the activating Fc receptors, FcγRIIb is an inhibitory receptor. Engagement of FcγRIIb on macrophages and dendritic cells (DCs) delivers an inhibitory signal.

Lymphocyte Activation, Proliferation, and Function

Several SLE risk genes have been implicated in the regulation and activation of lymphocytes. *BLK*, *LYN*, and *BANK1* encode the tyrosine-kinase proteins Blk and Lyn and B-cell scaffold protein with ankarin repeats; all are associated with intracellular signaling pathways. *ETS1* and *IKZF1* encode transcription factors and are believed to play a role in B-cell differentiation and self-tolerance. Cytotoxic T lymphocyte antigen-4 (CTLA-4) is upregulated on T cells after activation and dampens inflammatory responses. It has a higher affinity than CD28 for B7.1 (CD80) and B7.2 (CD86), thereby competitively inhibiting engagement of CD28 and blocking the costimulatory signal

required for T-cell activation. CTLA-4 ligation of B7 also activates indoleamine dioxygenase (IDO) expression, an enzyme involved in tryptophan metabolism, and diminishes T-cell proliferation. Finally, CTLA-4 is critical for activation of regulatory T cells. In CTLA-4$^{-/-}$ mice, an uncontrolled, lethal inflammatory response occurs. CTLA-4 alleles with decreased production of soluble CTLA-4 are implicated in the pathogenesis of several autoimmune diseases, including Sjögren disease, ulcerative colitis, psoriasis, type 1 diabetes, and multiple sclerosis as well as SLE.[15]

The *PTPN22* gene encodes a tyrosine phosphatase that has been suggested to downregulate T-cell receptor (TCR) and B-cell receptor (BCR) activation, although the precise role of PTPN22 in disease pathogenesis remains unresolved. A *PTPN22* polymorphism resulting in diminished interaction with Csk has been reported in SLE and other autoimmune diseases, suggesting a common mechanism of immune dysregulation. Surprisingly, the PTPN22 risk allele has also been associated with reduced type I IFN production by myeloid cells after activation of TLR7.

Cross-linking of FcγRIIb and the BCR results in decreased intracellular calcium flux with decreased B-cell activation and proliferation. The FcγRIIb I232T allele leads to the inability of the receptor to enter lipid rafts and associate with the BCR, thereby diminishing its ability to perform an inhibitory function. Dysregulation of FcγRIIb expression on activated memory B cells has been reported in SLE patients.[14]

Cytokine and Chemokine Regulation

IFN regulatory factor 5 gene *(IRF5)* encodes a critical transcription factor in the type I IFN pathway; the *IRF5* locus has the strongest association with SLE outside of the MHC region. Four allelic variants have been identified in multiple ethnically diverse populations. These alleles seem to be most important for myeloid cell function, even though IRF5 is expressed in numerous cell lineages. Variants of the signal transducer and activator of transcription factor 4 protein *(STAT4)* gene have also been associated with SLE susceptibility. Interleukin (IL)-1 receptor-associated kinase 1 *(IRAK1)* and methyl-CPG-binding protein 2 *(MECP2)* genes are both found on the X chromosome. IRAK1 regulates multiple pathways in innate and adaptive immune responses, including the link between immune complexes, TLR signaling, and IFN production. Monocyte chemoattractant protein (MCP)-1 is a potent chemoattractant for monocytes, memory T cells, and natural killer T cells. MCP-1 expression is upregulated in renal tubular cells, and glomeruli and urine levels of MCP-1 are increased in patients with active lupus nephritis. An MCP-1 polymorphism resulting in increased MCP-1 production has been associated with nephritis.[16] Polymorphisms in the tyrosine kinase-2 *(TYK2)* gene are associated with increased expression of type I IFNs (IFN-α, IFN-β) in SLE. IFN-α–regulated genes are highly expressed in peripheral blood cells from SLE patients compared with healthy controls (the IFN signature).[17] IFN-α mediates maturation of DCs and monocytes, increasing the capacity for T-cell activation; promotes B-cell differentiation and immunoglobulin class switching; and skews to extrafollicular plasma cell differentiation. However, in two murine models of lupus, decreases in type I IFNs unexpectedly led to worsening disease, consistent with numerous studies showing that type I IFNs have both proinflammatory and antiinflammatory effects on innate and adaptive immune responses and suggesting that the effect of IFN-α on autoimmunity is complex.

TREX1 encodes a 3′ repair exonuclease that monitors DNA synthesis; TREX1 deficiency leads to accumulation of endogenous DNA and is associated with increased expression of IFN and autoimmunity.

Multiple polymorphisms in the *IL10* gene have been reported, with conflicting results with respect to SLE susceptibility. A meta-analysis of 15 studies concluded that some *IL10* polymorphisms do associate with SLE, but their importance is modulated by ethnic background.[18]

In the NZB/W murine model, a tumor necrosis factor (TNF) allele associated with low production is linked with disease, and treatment with TNF decreases autoantibody production. Consistent with this observation, TNF blockade for rheumatoid arthritis or inflammatory bowel disease can lead to autoantibody production and infrequently to frank lupus. Several polymorphisms for genes encoding TNF and lymphotoxin have been associated with SLE; these associations are also influenced by ethnicity.[19]

Cell Survival

Fas ligand (expressed on activated T cells) binding to Fas (CD95) stimulates a signaling pathway resulting in apoptotic death of the Fas-expressing cell. Fas-induced apoptosis of activated cells contributes to the elimination of autoreactive B and T lymphocytes. Lymphopenia in SLE has been associated with increased Fas expression on lymphocytes, and Fas and Fas ligand alleles have been linked to disease susceptibility.[20] Lack of Fas expression in mice leads to an SLE-like phenotype, whereas lack of Fas expression in humans is associated with a lymphoproliferative disease that does not share autoantigen specificities and target organs with SLE.

Bcl-2 family genes encode intracellular proteins that are either proapoptotic or antiapoptotic. Increased expression of Bcl-2, an antiapoptotic molecule, leads to a lupus-like serology and nephritis in mice with certain genetic backgrounds. Increased intracellular levels of Bcl-2 have been reported in SLE. The combination of a Bcl-2 susceptibility allele and IL-10 susceptibility allele confers a 40-fold increased risk of SLE, demonstrating that *infelicitous* combinations of risk alleles potentiate risk.[21]

Target Organ Damage

Only a few genes are known to regulate the vulnerability of target organs (*e.g.*, the kidney) to autoimmune attack. In mice, genes encoding kallikreins, which upregulate bradykinins, have SLE susceptibility alleles. In human studies a risk allele for ABIN1, which regulates nuclear factor (NF)-κB activation, can lead to greater kidney disease. Two risk alleles of APOL1 are particularly common in African Americans and are thought to contribute to the increased severity of renal disease in this population.

Epigenetic Contributions

Epigenetic regulation plays a determining role in gene activation. Major epigenetic influences in SLE involve DNA methylation at cytosine-guanine nucleotides (CpG methylation) and histone posttranslational modifications (lysine acetylation or methylation, phosphorylation of serine or threonine, arginine methylation).[22] SLE susceptibility and autoantibody production are associated with DNA hypomethylation.[23] Consistent with this observation is the fact that several drugs known to induce a lupus-like disease (procainamide, hydralazine) also cause decreased DNA methylation. A landmark study of high-throughput analysis of DNA methylation in discordant twins

demonstrated greater DNA hypomethylation in the affected siblings.[24] It is not surprising that genes encoding integrins, NGAL, CD40 ligand, IFN-γ receptor, and IL-6 are among the hypomethylated genes in SLE. Mechanisms for hypomethylation remain unclear; decreased efficacy of DNA methyltransferases (DNMTs) and overexpression of microRNAs (miRNA) that interfere with DNMT activity have been proposed.[25]

Vitamin D has been shown to contribute to the regulation of the epigenome. Vitamin D levels are low in many lupus patients, but the relationship of vitamin D levels to disease risk or disease severity remains controversial.[26]

Inflammation itself may alter the epigenome; some studies suggest that the metabolic modulator metformin can reverse some of the inflammation-induced alterations. Recent studies show that metformin can reduce murine lupus, although the mechanism for this therapeutic effect is not clear.

B Cells

The process of immunoglobulin variable region gene rearrangement produces large numbers of self-reactive B cells. Most B cells displaying self-reactive immunoglobulin are deleted centrally in the bone marrow and at subsequent checkpoints in the periphery (Fig. 52.2), so the frequency of autoreactive cells decreases from approximately 75% in immature B cells in the bone marrow to approximately 20% in the mature naïve B-cell population in healthy individuals. Many of the autoreactive B cells in the naïve population are normally suppressed by anergy induction. Several of these checkpoints may be deficient in SLE. The IgM autoantibodies made by these remaining autoreactive mature naïve B cells are thought to facilitate clearance of apoptotic cells in a non-immunogenic fashion. When IgM autoreactive antibodies are missing in mice, a lupus-like phenotype develops.

B cells differentiate to plasma cells through two distinct pathways, extrafollicularly or in the germinal center. Both these pathways have been implicated in SLE.

An important peripheral checkpoint for B cells maturing through a germinal center pathway is entry into the T-cell–dependent, long-lived memory compartment. B cells with the 9G4 idiotype express antibodies encoded by the *VH4-34* gene, reactive with *N*-acetyllactosamine (NAL) determinants of glycoproteins on blood group antigens targeted by cold agglutinins, gangliosides, gastrointestinal (GI) mucins, glycolipids, and CD45 on B lymphocytes.[27] 9G4 B cells are reported to be present in 5% to 10% of the naïve B-cell population in healthy donors as well as in the IgM memory compartment. However, 9G4 B cells are excluded from the T-cell–dependent IgG memory and plasma cell populations, suggesting that these autoreactive cells fail to cross a developmental checkpoint after activation in normal individuals. Evaluation of tonsillar biopsies and spleens from healthy donors shows that the frequency of germinal centers with 9G4-positive cells is less than 1%, implying that negative selection of autoreactive cells occurs at the transition of naïve to germinal center B cells. In contrast, tonsillar biopsies from SLE patients demonstrate that 15% to 20% of germinal centers are positive for autoreactive 9G4 B cells. This study and others suggest that checkpoints exist for entry into and egress from the germinal center response and that these may be altered in SLE.

It is clear that some pathogenic autoantibodies are not derived from natural autoantibodies. Back mutation of several anti-DNA antibodies to their germline-encoded precursors has identified non-autoreactive precursors. The failure of censoring mechanisms in germinal centers to prevent the maturation of autoreactive cells may reflect intrinsic B-cell abnormalities or abnormalities in costimulatory molecules, cytokines (*e.g.*, B cell–activating factor [BAFF]; see later), follicular DCs, or T-cell, B-cell interactions.

An alternative model recently put forth is that more B cells differentiate to IgG plasma cells in SLE patients, leading to greater antibody production rather than defects in specific tolerance checkpoints in SLE patients.[28]

B lymphocyte stimulator (BAFF; also known as BLyS) is a member of the TNF family and participates in B-cell maturation and survival. BAFF enhances survival of B cells through engagement of several receptors (BCMA, BAFF-R, and TACI). High levels of this cytokine allow survival of autoreactive B cells in mice, resulting in a lupus-like disease.[29] Evidence supporting a role for BAFF in autoreactive B-cell rescue and human SLE includes the elevated levels seen in lupus patients, associations of BAFF levels with autoantibody titers, and, in some reports, correlations with disease activity. Belimumab, a monoclonal antibody directed against soluble BAFF, has been shown to increase the percentage of anergic ANA-positive B cells in SLE patients.

Interactions between CD40 (B cell) and CD40 ligand (CD40L, T cell) are essential for B-cell proliferation, differentiation of memory cells into plasma cells, and germinal center formation. Immature autoreactive B cells can be rescued from antigen-induced apoptosis by engagement of CD40 or by IL-4. SLE T and B cells have upregulated CD40L, a critical molecule to mediate B-cell rescue. The combination of IL-17 and BAFF facilitates B-cell proliferation and maturation.[30] This combination can serve as an alternative stimulatory signal for B-cell activation and can replace CD40/CD40L interactions. Therefore in a permissible and proinflammatory cytokine milieu, B-cell

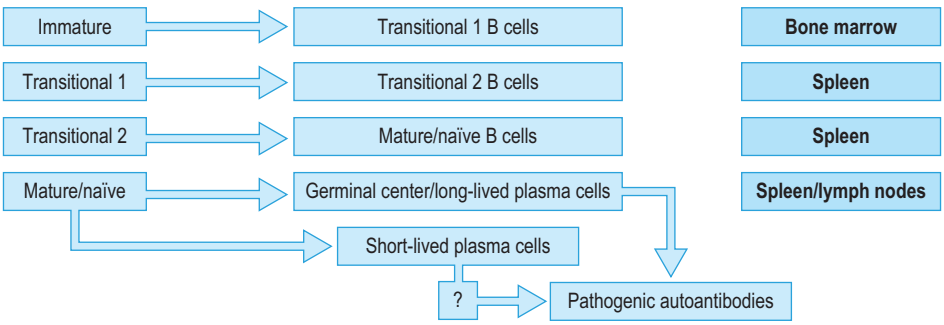


FIG. 52.2 Autoreactive B-Cell Checkpoints. There are tolerance checkpoints at every stage of B-cell activation and maturation. How many checkpoints need to be breached to achieve a pathogenic state and clinical disease is not known.

activation and autoantibody production may occur in the absence of cognate T-cell help.[31] These studies are consistent with the increased numbers of plasmablasts and class-switched CD27-negative B cells in blood of lupus patients.

B cells are efficient antigen-presenting cells; B cells with self-reactive specificity that have escaped tolerance are likely to present self-antigen to autoreactive T cells. This may explain, in part, the beneficial effects of treatment with rituximab, a B cell–depleting drug, in SLE even as autoantibody levels remain largely unaffected.

Hyperactive B-cell responses to immunologic stimulation are implicated in the production of pathogenic antibodies. SLE B cells have increased intracellular Ca^{2+} flux in response to BCR signaling,[31] partially due to FcγRIIb dysfunction (Ile 232 Thr).[32]

The intracellular protein tyrosine kinase Lyn has both positive and negative effects on BCR signaling. Decreased expression of Lyn results in increased intracellular Ca^{2+} flux and B-cell hyperactivity. Correspondingly, Lyn expression is decreased in resting and activated B cells in one-half to two-thirds of SLE patients.

A new B-cell subset has been identified, the so-called age or autoimmunity-associated B cells (ABCs). This subset is expanded in patients with SLE. It is dependent on TLR activation, and some studies suggest it is a major source of autoantibodies.[33]

Neutrophils

An important link exists between neutrophils and autoimmunity. Neutrophil extracellular traps (NETs) are chromatin filaments released from neutrophils to trap microbes. NETs also contain neutrophil peptides, including neutrophil-encoded antimicrobial peptide LL37 (cathelicidin). Anti-DNA antibodies bind to DNA in NETs and are potent TLR9 stimulants, resulting in production of IFN-α by plasmacytoid DC (pDC).[34] Moreover, anti-RNP antibodies induce "NETosis" in a process that is dependent on TLR7 and FcγRIIa signaling.[35] These studies confirm earlier work that described a "granulocyte signature" with significant upregulation of granulocyte-specific transcripts within peripheral blood mononuclear cells in pediatric SLE patients.[36]

Dendritic Cells

DCs recognize pathogens through membrane pattern recognition receptors (PRRs) and are a critical component of the immune system connecting innate and adaptive immune responses.[37] They can be tolerogenic or immunogenic due to their high expression of costimulatory molecules. DCs function normally as surveillance cells, phagocytizing cellular debris and determining whether there is cause for alarm. PRRs on DCs bind pathogen-associated molecular patterns (PAMPs) and damage-associated molecular patterns (DAMPs) that are present in sterile inflammation. After internalization of the PAMPs or DAMPs, endosomal TLRs are activated by nucleic acids and other ligands. Enhanced IFN-α in SLE pathogenesis is now well described. The primary cells responsible for IFN-α production are pDCs. TLR7 or TLR9 activation with RNA and DNA, respectively, induces immature DCs to differentiate to immunocompetent, IFN-α–producing DC that have wide-reaching effects on T cells, B cells, neutrophils, and monocytes. Of note, type I IFN increases plasmablast differentiation favoring short-lived plasma cells over germinal center–matured long-lived plasma cells. pDCs have been demonstrated in skin and renal lesions, and upregulation of the "IFN signature" is observed in many, but not all, SLE patients and has been associated with disease activity. A number of susceptibility genes identified through GWASs are associated with IFN pathways in SLE.

T Cells

T cells are critical in the abrogation of self-tolerance by providing help to autoreactive B cells and facilitating the production of somatically mutated, high-affinity, pathogenic autoantibodies. Lupus patients can exhibit T cell phenotypes consisting of increased numbers of $CD3^+CD4^-CD8^-$ T cells, increased Th17 cells, T-follicular helper (Tfh) cells, and decreased numbers or function of Tregs.[38] In addition, lupus T cells display increased expression of activation markers and abnormal TCR signaling responses. A substitution of TCR ζ-chain by the γ-chain of the Fc receptor results in increased intracellular influx of Ca^{2+} after TCR stimulation and a concomitant decrease in IL-2 production, disfavoring the induction of Tregs.

Self-tolerance is maintained in part by the suppressive actions of Tregs. "Natural" Tregs arise *de novo* in the thymus, whereas "induced" Tregs evolve from naïve T cells exposed to IL-2 and transforming growth factor (TGF)-β in the periphery. These cells are characterized by high surface expression of the IL-2 receptor α-chain, CD25, and high levels of FOXP3 intracellularly. Tregs act together with tolerogenic DC to maintain a steady state of immature DCs. In contrast, effector T cells can secrete IFN-γ and IL-17 that promote immature DC differentiation to immunogenic DCs capable of secreting IL-1, IL-6, IL-12, and TNF and activating autoreactive T cells, thereby establishing a feedback loop with impaired Tregs and activation of autoreactive T cells.[39] Studies demonstrate alterations in Tregs in patients with SLE. Some have demonstrated reduced numbers and altered function of peripheral $CD4^+CD25^+FOXP3^+$ cells in patients with active disease, whereas others fail to identify a defect.[40] SLE and some other autoimmune diseases are characterized by an expanded population of $CD4^-CD8^-$ (double negative) T cells that produce IL-10; some believe these are autoreactive T cells.[41]

HORMONAL INFLUENCES

The most compelling evidence for the role of sex hormones in SLE is the observation that lupus preferentially affects women of childbearing age. The female-to-male ratio is 2:1 before menarche, 8 to 9:1 in the fourth decade, and 2:1 after menopause. Numerous case reports and studies of disease flares correlating with pregnancy, menstruation, and use of oral contraceptives containing high doses of estrogen suggest a role for estrogen in disease activity. A significant correlation between plasma levels of estradiol, increased α-hydroxylation of estrogen in SLE, yielding the more active metabolite 16α-hydroxyestrone, and clinical disease activity is also reported.[42] No significant differences in levels of sex hormones (including estrogen, testosterone, prolactin) are noted in male SLE patients, suggesting that the development of SLE in females may be more closely related to sex hormones than in men. Randomized controlled studies of estrogen in SLE suggest that the use of exogenous estrogen in patients with stable disease may be safe; however, a subset of patients appears to have an estrogen-sensitive disease. Mild to moderate flare rates were significantly increased in postmenopausal women treated with hormone replacement therapy.

Most of what we understand about hormonal modulation of B-cell development comes from mouse studies.[43] Estrogen

treatment of NZB/W and MRL/*lpr* mice or castration of male lupus-prone mice exacerbates disease, whereas oophorectomy of female mice ameliorates disease. Treatment of lupus-prone mice with the selective estrogen receptor modulator tamoxifen also ameliorates disease.

Estrogen and prolactin promote the loss of B cell tolerance in nonautoimmune mice. Estrogen results in diminished B-cell responsiveness to BCR cross-linking and in less stringent negative selection. In addition, estrogen has also been reported to inhibit activation-induced T-cell death by downregulation of Fas ligand expression, thereby permitting increased numbers of autoreactive T cells.[44] Emerging data suggest that estrogen receptor signaling promotes differentiation of DC into immune-competent DC through increased expression of the transcription factor IRF4.[45]

Elevated prolactin levels are reported in 20% of patients with SLE, and increased prolactin exposure in lupus-prone mice exacerbates disease activity. Treatment of patients with bromocriptine yielded equivocal results, whereas treatment of NZB/W lupus mice with bromocriptine results in improved survival. Transmembrane prolactin receptors are present on a variety of cells, including T and B cells. Upregulation of both Bcl-2 and CD40 on B cells and CD40L on T cells occurs in response to prolactin, suggesting pathways that may be involved in the prolactin-mediated rescue of autoreactive B cells.

Data on the microbiome demonstrate that sex hormones can modulate its composition. Studies in murine lupus suggest that androgens generate a microbiome that prevents the development of lupus, whereas estrogen maintains a disease-permissive microbiome. Studies of the microbiome in human disease are just beginning.[46]

CLINICAL MANIFESTATIONS

CLINICAL PEARLS

- Systemic lupus erythematosus (SLE) is a systemic disease with the potential to affect any organ system.
- Not all symptoms experienced by a patient with SLE are from SLE-disease activity. Attribution is critical to determine before initiation of treatment. The differential diagnosis of a lupus flare mandates consideration of infection, drug toxicities, or other etiologies.
- Corticosteroid exposure should be minimized.
- Use of aggressive treatment must be balanced against associated toxicities.
- SLE patients accumulate damage from both repeated episodes of inflammatory disease and medication toxicities. Prompt recognition and appropriate treatment of disease flares should result in reduced exposure to corticosteroids and immunosuppressive agents.
- SLE patients are at increased risk of developing atherosclerotic disease, osteoporosis, malignancy, diabetes mellitus, and hypertension. It is essential to screen for and reduce modifiable risk factors.

The prevalence of the diverse clinical and laboratory features of lupus varies in published reports (see Table 52.2). Both genetic and environmental factors are likely to account for much of the variability between cohorts. Characteristics of published cohorts, including inception versus long-standing lupus cohorts, community versus academic settings, socioeconomic factors, and ascertainment differences are also likely to contribute to observed differences in the prevalence of clinical manifestations. Over time, the course of lupus is characterized by disease flares and remissions.

KEY CONCEPTS

- Continued heightened awareness of systemic lupus erythematosus to shorten the time between onset of symptoms and diagnosis is needed to improve outcomes.
- Lupus is a disease characterized by recurrent flares.
- Attentive monitoring, even during periods of disease remission, leads to early recognition of impending flare, better control, and better prognosis.
- Lupus is a chronic disease; the importance of a therapeutic partnership between physicians and patients, emotional/social support, and patient education cannot be overemphasized.
- Advances in understanding mechanisms of disease pathogenesis correspond to advances in treatment strategies and development of therapies with improved efficacy and safety profiles.

The most common features of lupus are constitutional and include fatigue, malaise, low-grade fever, anorexia, and lymphadenopathy. These symptoms are believed to result from elevated serum levels of inflammatory cytokines and may accompany other organ system manifestations of active disease, or may occur in isolation. Although frequent, these symptoms are nonspecific and do not aid in making the diagnosis of SLE. Symptoms of fatigue and malaise may also represent fibromyalgia, which can co-occur with SLE or confound the diagnosis.[47]

Musculoskeletal Involvement

The musculoskeletal system is the most common organ system affected in SLE; joint pain is the presenting symptom in approximately 60% to 70% of patients, and 85% have joint involvement after 5 years.[48]

Arthritis and Arthralgia

The pattern of joint involvement is usually symmetric, affecting the small joints of the hands, wrists, and knees. Large joint or monoarticular involvement is less typical. In contrast to rheumatoid arthritis, morning stiffness is typically limited to several minutes. Frequently, the subjective complaints of pain are greater than the objective findings of warmth, swelling, and erythema and must be distinguished from concomitant fibromyalgia. Lupus arthritis is characteristically nonerosive on x-ray and nondeforming. Anticyclic citrullinated peptide antibodies occur frequently in the rare patients with erosive arthritis. Some lupus patients develop a nonerosive hand deformity with hypermobile joints secondary to tendon and ligamentous laxity (Jaccoud arthritis) (Fig. 52.3). Proliferative tenosynovitis, synovitis, small erosions not detectable on plain radiographs, capsular swelling, and bone marrow edema are features of joint and soft-tissue involvement that may be seen on MRI. Although sensitive techniques, the role of ultrasound and MRI in the evaluation and management of musculoskeletal symptoms is not established.

Joint effusions, when they occur, are usually small. The fluid is clear yellow with normal viscosity and forms a mucin clot. It is typically noninflammatory with a normal glucose level and a white blood cell (WBC) count of less than 2000 cells/mL that is predominantly lymphocytic. ANA performed on the synovial fluid may be positive, and lupus erythematosus (LE) cells may be present. Synovial fluid complement levels can be normal or depressed. Synovial histology in lupus is not specific and shows synovial hyperplasia with fibrin deposition and microvascular changes that include perivascular infiltrates in the majority of cases.

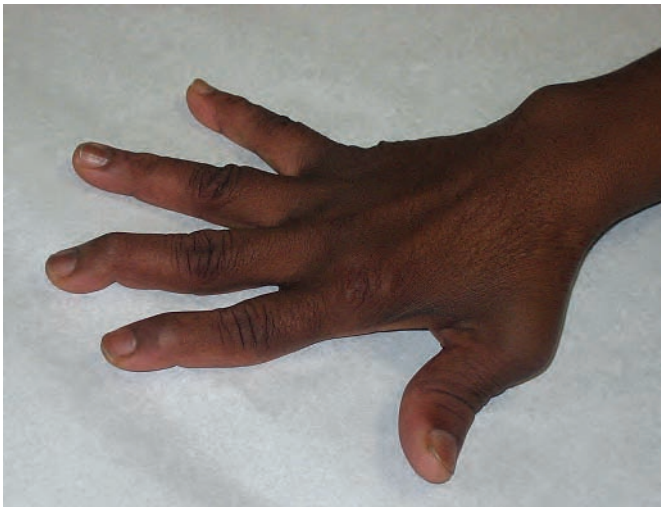


FIG. 52.3 Jaccoud Arthritis in Systemic Lupus Erythematosus.

Tendinitis

Tendinitis is not usually attributed to SLE unless associated with tendon rupture. When present, it is usually located in the Achilles tendon or the tendons around the knee. Tendon ruptures are more common in males and have been associated with trauma, steroid use, long disease duration, and Jaccoud arthropathy.[49] Biopsy shows a mononuclear infiltrate with tendon degeneration and neovascularization. The diagnosis can be easily made by ultrasound or magnetic resonance imaging (MRI).

Myositis/Myalgia

Generalized myalgia is extremely common in lupus. It frequently affects the deltoids and quadriceps and occurs during flares of active disease. Muscle disease secondary to treatment with corticosteroids, statins, and antimalarials or in association with hypothyroidism is also frequent and must be considered in the evaluation of a lupus patient with myalgia. Inflammatory muscle disease with weakness and an elevated creatine phosphokinase is less common, occurring in approximately 10% of lupus patients.[47] Electromyography can be normal or can be characteristic of the myositis observed in polymyositis or dermatomyositis. Muscle biopsy can also be normal or can show changes associated with dermatomyositis such as a perivascular or perifascicular infiltrate and immunoglobulin and complement deposition. Muscle atrophy, fiber necrosis, microtubular inclusions, and/or a mononuclear infiltrate have been documented. MRI findings are nonspecific.

Avascular Necrosis

Avascular necrosis (AVN) has been reported in up to 30% of lupus patients, is frequently asymptomatic, and is detected by MRI.[50] The most commonly affected site is the femoral head. Groin pain exacerbated with weight bearing is a common complaint. In addition to the hip, AVN can involve the knees, shoulders, and wrists. The majority of AVN is associated with previous administration of high doses of corticosteroids (>30 mg/day), vitamin D deficiency, minority ethnicity, hypertension, and renal disease. Disease activity, independent of corticosteroid exposure, is an additional risk factor in patients with early lupus.[51] Bone biopsy in lupus patients affected by AVN does not reveal unique findings.

Mucocutaneous Manifestations

Skin: Classification and Pathogenesis

Cutaneous lupus erythematosus (CLE) occurs commonly, up to 70% of the time, in the context of SLE. It consists of a wide variety of lesions from the characteristic malar erythema to severe bullous lupus and scarring discoid lesions. Although rarely life threatening, CLE can be disfiguring and contributes substantially to depressive symptoms and diminished life quality.[52,53] The 2012 SLICC classification criteria include 10 mucocutaneous lesions not included in the 1997 ACR criteria, making careful appraisal and attribution of these lesions critically important because they contribute substantially to SLE diagnosis. Skin lesions are generally categorized as acute cutaneous lupus erythematosus (ACLE), subacute cutaneous lupus erythematosus (SCLE), and chronic cutaneous lupus erythematosus (CCLE).[54] Genetic associations include HLA haplotypes and SNPs, most of which may contribute to CLE pathogenicity through dysregulated antigen presentation, IFN response, and apoptosis regulation.[55,56] Known triggers for CLE include ultraviolet (UV) light (UVA and UVB), infections, and drug reactions, the most common of which arise from angiotensin-converting enzyme inhibitors, calcium channel blockers, beta blockers, antifungal agents, and TNF inhibitors. UV light induces DNA strand breaks in keratinocytes, resulting in apoptotic cell death and providing a rich source of autoantigen (*e.g.*, Ro52 antigen). SLE patients have increased numbers of apoptotic keratinocytes after exposure to UV light, and these apoptotic cells have been identified in the basal layer of CCLE lesions. Particularly in a setting of impaired ability to clear apoptotic debris, the abundance of autoantigen, including endogenous RNA and DNA, provides stimuli for autoreactive T and B cells and recruitment of pDC with ensuing production of proinflammatory cytokines, chemokines, and TGF-β, which contributes to fibrosis and scarring.[55] Infiltrating pDCs have been identified in a majority of CLE lesions within perivascular inflammatory dermal nodules and at the dermal-epithelial junction in severely damaged skin in association with cytotoxic T cells.[57] Transcript analysis of lesional skin in CCLE demonstrates a paucity of Tregs with increased numbers of T helper 1 (Th1), IFN-γ–producing cells, and increased type I IFN signature.[58] Increased NETs extruded from dying neutrophils are also seen in CLE lesions. The DNA in NETs and UV-oxidized DNA is resistant to enzymatic degradation and may contribute to ongoing pDC activation and IFN-α production through TLR activation.

Recent molecular profiling of DLE and SCLE lesional biopsies has demonstrated significant similarity between both CLE subtypes in differentially expressed type I IFN pathway genes and in repression of epidermal growth factor receptor (EGFR) pathways, suggesting underlying pathobiologic similarities between phenotypically different clinical subtypes.[58]

Autoantibodies are frequently identified in the dermal-epidermal junction and may facilitate antibody-dependent cell-mediated cytotoxicity. However, their contribution to CLE pathogenesis is unclear except for the significant association between anti-Ro/SSA antibodies and SCLE.[59] The lupus band test (LBT) refers to the deposition of immunoglobulin (IgG, IgM, and/or IgA) and/or C3 along the dermoepidermal junction. Approximately 25% of normal individuals display weak IgM staining at the dermoepidermal junction, whereas 70% to 80% of SLE patients have a positive LBT in sun-exposed, nonlesional skin. Examination of non-lesional skin in SLE has also

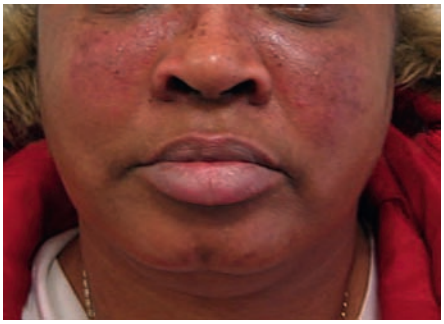


FIG. 52.4 Malar Rash in a Systemic Lupus Erythematosus Patient.

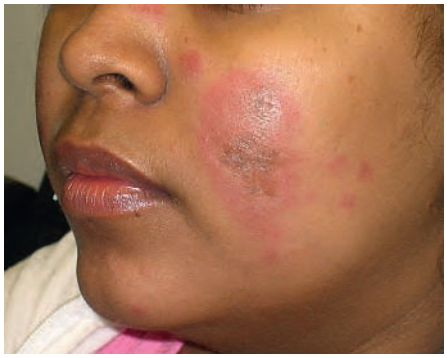


FIG. 52.5 Discoid Lesion in Systemic Lupus Erythematosus.

demonstrated presence of a keratinocyte IFN signature and Langerhans cell depletion, suggesting an underlying propensity for inflammation.[55]

Acute Cutaneous Systemic Lupus Erythematosus

ACLE includes malar rash, bullous lupus, toxic epidermal necrolysis (TEN) variant, maculopapular lupus rash, and photosensitivity. Malar rash is the most common ACLE lesion, typically occurring across the cheeks and nose but sometimes including the forehead and chin, sparing the nasolabial folds (unlike seborrheic dermatitis) (Fig. 52.4). It usually begins as small discrete erythematous macules or papules that coalesce with or without facial swelling, is frequently associated with sun exposure, and heals without scarring. The differential diagnosis includes acne rosacea, seborrheic dermatitis, erysipelas, dermatomyositis, and contact dermatitis. Microscopic analysis reveals a sparse inflammatory lymphocytic dermatitis with occasional histiocytes engulfing nuclear debris resembling LE cells found close to the dermoepidermal junction. Immunofluorescent staining for complement components and immunoglobulin at the dermoepidermal junction is positive in 70% to 80% of patients.

Subacute Cutaneous Systemic Lupus Erythematosus

This distinctive rash is mostly diagnosed in White populations. It consists of erythematous papules and plaques, with or without adherent pityriasiform scaling, that erupt on the extremities and trunk, usually sparing the head and neck. These nonscarring lesions may assume an annular polycyclic form with central pallor and tiny vesicles at the active margins and can be mistaken for erythema multiforme. The differential diagnosis includes psoriasis, polymorphic light eruption, and tinea corporis. SCLE is exacerbated by UV light and a growing list of medications, including thiazides and calcium channel blockers. Biopsy reveals a lymphocytic dermatitis confined to the superficial and mid dermis, frequently with associated dermal edema, mucinosis, and degenerating keratinocytes. From 60% to 90% of SCLE patients have circulating anti-Ro antibodies; however, they are also deposited in non-lesional skin.

Chronic Cutaneous Systemic Lupus Erythematosus

CCLE includes discoid lupus erythematosus (DLE), verrucous lupus, panniculitis, tumidus, and chilblain lupus. DLE is the most common form of CCLE. Lesions and are usually localized to the head and neck in photo-exposed areas (Fig. 52.5) and are typically circular, vary in size, and scar with significant disfigurement. Early lesions appear as erythematous plaques with

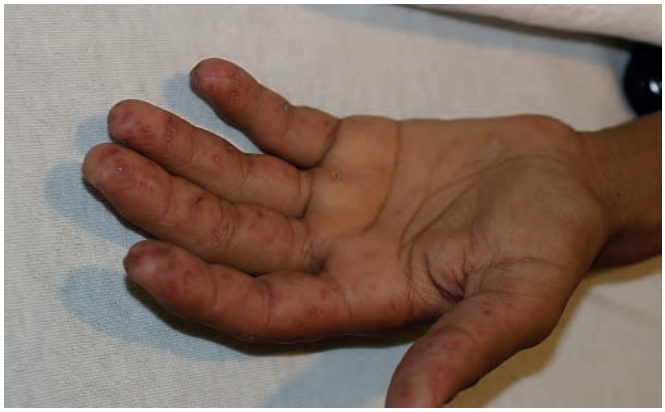


FIG. 52.6 Cutaneous Vasculitis Affecting the Hands in a Patient With Active Systemic Lupus Erythematosus.

or without follicular hyperkeratosis, plugging, and scale and progress to scarring annular lesions with an erythematous, indurated border, adherent scale, and a central area with atrophy and telangiectasias. There are no autoantibody associations with DLE, and only 5% of patients with DLE develop systemic lupus. High-titer ANA, Raynaud phenomenon, and the presence of arthralgias may identify patients at risk for systemic evolution. Histopathology characteristically reveals a lymphocytic interface dermatitis with CD4 lymphocytes and pDCs involving follicles and epidermis. There is vacuolar degeneration of the basal layer of epidermal keratinocytes and prominent keratotic follicular plugging. Dermal mucin deposition is also present, and there is usually dense granular deposition of immunoglobulin (predominantly IgG) and C3 at the dermal-epidermal junction.

Lupus profundus typically presents as firm, tender, deep subcutaneous nodules that may atrophy over time. Overlaying epidermal changes include DLE, ulcerations, and dystrophic calcification. Biopsy reveals a lobular panniculitis with patchy lymphoplasmacytic infiltrate in subcutaneous fat lobules. Panniculitis (10% to 20% of patients) must be differentiated from a subcutaneous T-cell lymphoma, erythema nodosum, pancreatic panniculitis, and morphea.

Nonspecific skin lesions reported in SLE are typically seen during disease flares and are associated with greater disease severity. These lesions include, but are not limited to, cutaneous vasculitis (Fig. 52.6), urticaria, Raynaud phenomenon, livedo reticularis, alopecia, sclerodactyly, calcinosis cutis, atrophie blanche, bullous lesions, erythema multiforme, and leg ulcers.

Hair and Nail

Different patterns of hair loss occur in SLE.[60] Scarring alopecia with permanent hair loss is associated with DLE, and biopsy shows typical DLE infiltrates with interface dermatitis and immunofluorescence that differentiates it from other common forms of scarring alopecia. Patchy or diffuse non-scarring alopecia is frequently associated with disease activity and is SLE-specific according to the 2012 SLICC criteria for SLE. Although clinically it may appear similar to alopecia universalis, anagen, or telogen effluvium, SLE-related non-scarring alopecia is distinguished histologically by interface dermatitis, obscuring of the dermoepidermal junction with inflammatory cells. Non-scarring alopecia resolves with complete hair regrowth with control of disease activity.

A wide spectrum of nail abnormalities, including pitting, ridging, onycholysis, and dyschromia with blue or black hyperpigmentation, is reported in up to 30% of SLE patients, but none are lupus specific. Nail fold erythema with ragged cuticles and splinter hemorrhages resembling the changes of dermatomyositis are common.

Oral Lesions

The spectrum of oral lesions reported in SLE includes cheilitis, ulcerations, erythematous patches, lichen planus–type plaques on the buccal mucosa and palate, and DLE.[61] Most oral lesions are asymptomatic. Positive immunofluorescent staining on biopsy may be useful to differentiate DLE from lichen planus–like lesions and leukoplakia. Lupus mucosal ulcerations demonstrate an interface mucositis and not leukocytoclastic vasculitis.

Gastrointestinal Manifestations

GI symptoms occur commonly in SLE, with reported incidences of 15% to 75%; attribution is critical because at least half are attributable to side effects of medications and to infectious complications.[62–64]

Esophagus

The prevalence of esophageal involvement varies. Dysphagia and heartburn are reported in approximately 50% of patients, although many reviews citing high incidences of dysphagia and odynophagia predate the advent of proton pump inhibitors and H_2 blockers, and the relationship of medication use to symptoms is not clear. Esophageal dysmotility is observed in up to 70% and is attributed to an inflammatory process involving esophageal muscle or vasculitic damage to the Auerbach plexus. Ulceration is rarely seen outside the context of infections such as invasive candidiasis, herpes simplex, or cytomegalovirus. SLE patients with secondary Sjögren syndrome may have salivary gland dysfunction, resulting in decreased saliva contributing to dysphagia.

Abdominal Pain/Vasculitis

Acute abdominal pain is common in SLE with reported incidences as high as 40%. The differential diagnosis includes intestinal vasculitis (lupus enteritis, mesenteric vasculitis) (45.5%), pancreatitis (10.8%), hepatobiliary disease (18.8%), and intestinal pseudo-obstruction (IPO) (3.3%).[65] The most catastrophic and potentially fatal GI disturbances are related to ischemia of the small and large intestines resulting from medium- and small-vessel vasculitis or thrombotic complications of APL antibodies. Approximately half of SLE patients with acute abdominal pain will have intestinal ischemia; associated mortality is high, so early consideration and intervention are critical. Lupus enteritis and mesenteric vasculitis are frequently associated with active disease elsewhere, whereas SLE patients with inactive disease and acute abdominal pain will likely have intra-abdominal pathology unrelated to SLE. The clinical presentation may be acute, severe abdominal pain or an insidious, stuttering course with nausea, vomiting, bloating, diarrhea, postprandial fullness, anorexia, and weight loss. Mesenteric vasculitis preferentially affects the superior mesenteric artery, involving the small intestine more commonly than the large bowel. Vasculitis can also occur in the esophagus, stomach, peritoneum, rectum, gallbladder, pancreas, and liver. Computed tomography (CT) and/or magnetic resonance with or without angiography are the preferred imaging tests for evaluation of abdominal pathology; the radiographic signs of intestinal ischemia do not differ based on pathogenesis.

In cases with an insidious clinical course, endoscopy and colonoscopy may provide evidence of ischemia demonstrating ulcerating or heaped-up lesions with overt vasculitis on biopsy. The lesions are segmental and focal. Histologically, there is a small-vessel arteritis and venulitis with neutrophilic, lymphocytic, and macrophage infiltrates and fibrinoid necrosis of the vessel walls, associated thrombosis, and mononuclear infiltrate in the lamina propria. There may be immunoglobulin, C3, and fibrin deposition in the adventitia and media.

Intestinal Pseudo-obstruction

Although rare, SLE-associated IPO (SLE-IPO) may be the initial manifestation of SLE. Clinical symptoms (abdominal pain and distension with diminished or absent peristalsis) and radiographic findings of SLE-IPO mimic those of mechanical obstruction. Distinguishing features of SLE-IPO include concomitant active SLE in other organ systems and associations with hematologic cytopenias, hypocomplementemia, and serositis. Findings of coexisting ureterohydronephrosis and hepatobiliary dilatation in the absence of obstructing lesions suggest underlying smooth muscle dysmotility associated with vasculitis or autonomic nervous system dysfunction. Poor prognosis is associated with older age at SLE diagnosis, GI symptoms as the initial SLE manifestation, longer disease duration, and delayed diagnosis of IPO.[63]

Peritonitis

Symptomatic lupus peritonitis is seen in only 10% of patients, despite evidence of peritoneal inflammation in greater than 60% of autopsy studies. Acute peritonitis may be attributed to peritoneal vasculitis or ischemia and presents with abdominal pain (see earlier). The finding of ascitic fluid by CT scan or ultrasound mandates an evaluation of the fluid to exclude infection and malignancy. Rarely, ascites may be attributable to hepatic or portal vein thrombosis. Chronic peritonitis characterized by large amounts of painless ascites attributable to SLE and not to heart failure, constrictive pericarditis, or severe hypoalbuminemia due to nephrotic syndrome, liver disease, or a protein-losing enteropathy (PLE) is rare. In lupus peritonitis, the ascitic fluid is generally exudative with a predominance of lymphocytes; LE cells, autoantibodies, and low complement levels are frequent. On biopsy, the peritoneum is usually edematous; it is sometimes hemorrhagic with lymphocytic perivascular infiltrates.

Pancreatitis

Pancreatitis attributable to SLE is rare, occurring in 8% to 11% of patients with abdominal pain and having an annual reported

incidence of 1/1000 or less. Although both corticosteroids and azathioprine can trigger pancreatitis, 34% of reported cases are not on these medications at the onset of pancreatitis, and most patients respond to steroid therapy. The clinical presentation and diagnosis of pancreatitis are similar in patients with and without SLE. Specific findings in SLE include leukopenia, thrombocytopenia, and anemia. Inflammation and necrosis are common on biopsy; pathophysiology has been attributed to vasculitis, APL-related thrombosis of pancreatic vessels, and intimal thickening of pancreatic arterial walls with immune complex deposition.[62] Mortality rates are reported from 18% to 27%; poor outcome is associated with increased systemic SLE activity (particularly, low complement, and thrombocytopenia).

Liver

The term "lupoid hepatitis" was coined in the 1950s to describe cases of young women with chronic active hepatitis characterized by a lymphoplasmacytic infiltrate on biopsy, hypergammaglobulinemia, and a positive LE cell test in the blood. With the development of ANA testing, it became apparent that "lupoid" and "chronic active hepatitis" were indistinguishable in terms of clinical presentation, pathology, and response to treatment, and the term "autoimmune hepatitis" (AIH) was endorsed in 1993.[66] Importantly, only 10% of AIH patients meet criteria for SLE, and 2.4% to 4.7% of SLE patients have noninfectious "lupus hepatitis" (LH).[64] Mild/moderate transaminitis is reported in up to 55% of patients; however, medication-related effects, infection, venous congestion related to cardiopulmonary disease, hemolysis, myositis, and veno-occlusive disease must all be considered prior to SLE attribution. The distinction between subclinical LH and AIH is important because therapy and prognosis are different. LH is associated with mild enzyme abnormalities, whereas AIH is a progressive disease frequently leading to hepatic failure. Both conditions share a predilection for young women, and both demonstrate features of autoimmunity, including hypergammaglobulinemia, arthralgias, and serum autoantibodies. Histologically, LH biopsies reveal lobular and periportal lymphocytic infiltrates, in contrast to the periportal and piecemeal necrosis with dense lymphocytic infiltrates seen in AIH with progression to panlobular or multilobular necrosis and cirrhosis. Serologically, antibodies to ribosomal P (51%), dsDNA (70%), and Ro (60%) have been associated with LH, whereas AIH is associated with antibodies to liver and kidney microsomes. Anti–smooth muscle antibodies are observed in 60% to 80% of patients with AIH, compared with 30% of patients with LH. Although both have a favorable response to steroids, AIH usually requires additional immunosuppressive agents.[66]

Distinguishing LH from hepatitis C virus (HCV) can be difficult. Up to 30% of patients chronically infected with HCV have low titers of ANA and other autoantibodies (anti-DNA, anticardiolipin antibodies, and rheumatoid factor). They can also have cryoglobulins and associated cryoglobulinemic vasculitis. It is necessary to confirm a positive enzyme-linked immunosorbent assay (ELISA) for HCV with polymerase chain reaction (PCR) in patients presenting with arthritis, cutaneous vasculitis, and a positive ANA, because SLE patients may have false-positive serologic tests for HCV.

Protein-Losing Enteropathy

Profound hypoalbuminemia in the absence of severe nephrotic syndrome, liver disease, or constrictive pericarditis should trigger concern for PLE. The most common clinical presentations are related to hypoalbuminemia and include peripheral edema (94%), ascites (58%), pleural effusion (54%), and pericardial effusion (24%).[63] Clinically significant PLE is uncommon in SLE, with reported prevalence rates of 1% to 8%; it can present individually or in the context of severe disease activity with other organ involvement. The diagnosis is confirmed with an increased α_1-antitrypsin level in a 24-hour stool collection or with the presence of intravenously administered labeled human albumin in the stool. Imaging findings are nonspecific and include ascites and bowel wall thickening. Protein leakage occurs most commonly in the small intestine but may be multifocal, and intestinal histology reveals lymphangiectasis, edematous villi, inflammatory infiltrate, vasculitis, and mucosal atrophy. Biopsy results suggest a role of TNF, IFN-γ, and IL-6 in the increased vascular and enterocyte permeability in PLE.

Pulmonary Involvement

Lupus affects the lungs in diverse ways involving the pleura, lung parenchyma, and blood vessels. It is more common in late-onset lupus, in patients diagnosed at 50 years of age or older.[67] The most frequent and important complicating feature is infection.

Pleuritis

Pleuritis is the most common pulmonary manifestation of SLE, reported in 40% to 56% of patients.[68] Pleural involvement in up to 93% of lupus patients at autopsy suggests that much pleuritis may be asymptomatic. Clinically, patients note typical pleuritic pain. On physical examination, the most frequent abnormality is tachypnea; a pleural friction rub is present in some cases, and pleural effusions occur in more severe cases. Pleural fluid is usually exudative with normal glucose and pH, and elevated protein and lactate dehydrogenase levels. The leukocyte count can range from several hundred to 20 000 cells/μL; both a lymphocytic and neutrophilic predominance are reported. When performed, immunologic testing on pleural fluid may show reduced complement levels and the presence of ANA, anti-DNA antibodies, and LE cells. Although these tests are commonly obtained, these results are neither sensitive nor specific enough to diagnose lupus pleuritis.

Lupus Pneumonitis

Lupus pneumonitis occurs in up to 10% of patients. Patients present with dyspnea, cough, mild pleuritic chest pain, and fever. Pulmonary infiltrates are present on plain radiograph or CT. This presentation must be distinguished from an infectious etiology. Histologic examination of affected lung tissue shows alveolar edema and hemorrhage with hyaline membrane formation; immunofluorescent staining reveals immune complex deposition.

Pulmonary Hemorrhage

Pulmonary hemorrhage is a rare but potentially fatal complication of SLE which may be associated with APL antibodies. Symptoms include shortness of breath with or without hemoptysis, which may be stuttering in onset, accompanied by a fall in hemoglobin, usually occurring in the context of multiorgan involvement from SLE. Imaging may show patchy infiltrates. Pulmonary function testing is marked by an increased diffusion capacity (DL_{CO}) secondary to the presence of alveolar blood, whereas arterial O_2 saturation is decreased. Histopathology shows bland intra-alveolar hemorrhage and hemosiderin-laden macrophages. Microangiitis with an inflammatory infiltrate and necrosis of the alveolar septa can occur. As hemorrhage into the

lung may be secondary to thrombotic thrombocytopenia, infections, or pulmonary hypertension, demonstration of an inflammatory process in the pulmonary vessels or tissue is helpful to establish a diagnosis of primary pulmonary hemorrhage.

Chronic Diffuse Interstitial Lung Disease

Chronic diffuse interstitial lung disease is a relatively uncommon manifestation of SLE and is occasionally associated with anti-Ro antibodies. It usually has a progressive course with a chronic nonproductive cough, dyspnea, and pleuritic chest pains. Physical examination is frequently remarkable for basilar rales with diminished diaphragmatic movement. Pulmonary function tests demonstrate a restrictive pattern with decreased diffusion capacity; oxygen saturation is decreased. Imaging often shows interstitial fibrosis that is more prominent at the lung bases. High-resolution CT (HRCT) may also help to determine the extent of treatable disease (*i.e.*, fibrosis [honeycombing] versus inflammation [ground glass]). However, the most reliable method to assess the extent of pulmonary inflammation in comparison with fibrotic damage is histologic examination. Biopsies showing nonspecific interstitial pneumonia (NSIP) or lymphocytic interstitial pneumonia (LIP) are more frequent than usual interstitial pneumonia (UIP). Evaluation of bronchial alveolar lavage fluid helps to exclude infection.

Pulmonary Hypertension

Pulmonary hypertension unrelated to chronic pulmonary emboli or interstitial lung disease occurs in SLE and is associated with increased mortality. Severe cases are rare; the recent recognition of milder cases may be partially attributed to increased awareness and findings of elevated pulmonary artery pressures on echocardiograms obtained for evaluation of shortness of breath. Patients typically present with progressive dyspnea occurring in the absence of infiltrates on chest radiographs or significant hypoxemia. Chest pain and a chronic nonproductive cough are also frequently present. Autoantibodies to endothelin receptor type A may be present; brain natriuretic peptide is usually abnormal.[69] Pulmonary function testing reveals a reduced DL_{CO} with normal lung volumes. Elevated pulmonary artery pressure is confirmed with cardiac angiogram. Biopsy or autopsy specimens of the lung reveal "plexiform" lesions that resemble those seen in primary pulmonary hypertension.

Shrinking-Lung Syndrome

The shrinking-lung syndrome refers to the rare findings of shortness of breath occurring in the absence of pleuritis or interstitial lung disease plus a chest x-ray showing elevated hemidiaphragms. Pulmonary function testing shows a restrictive pattern with loss of lung volume. It was generally accepted that this syndrome results from diaphragmatic weakness (from a myopathic process) or chest wall restriction; however, more recent studies suggest that pleuritis and impaired ability to take deep inspirations, leads to parenchymal reorganization and reduced lung compliance may also play a causative role. There is no definitive therapy, although immunosuppressive therapy with cytotoxic agents usually results in an improvement of lung function and respiratory symptoms.

Cardiac Involvement

There are several ways in which lupus affects the cardiovascular system; targets include the myocardium, valves, pericardium, and blood vessels.

Myocardium

Myocardial dysfunction in SLE is likely to be secondary to factors other than lupus, such as hypertension, medications, or coronary artery disease (CAD). However, a cardiomyopathy resulting from immune-mediated myocardial inflammation does occur, either in isolation or concomitant with myositis or other manifestations of systemic disease. Inflammatory myocarditis is often associated with anti-RNP antibodies. Histopathology typically shows a mononuclear, inflammatory cell infiltrate. Perivascular or myocardial wall deposits of immune complexes and complement also occur. Cardiac MRI can facilitate a diagnosis of myocarditis; however, myocardial biopsy is the gold standard and provides additional information on the extent of active inflammatory disease and fibrosis. Symptoms and signs of myocarditis include unexplained tachycardia, an abnormal electrocardiogram (with ST- and T-wave abnormalities), cardiomegaly, and heart failure. Echocardiography may show systolic and diastolic ventricular dysfunction. Myocardial involvement without overt clinical signs occurs commonly and may be documented using Doppler echocardiography. A noninflammatory cardiomyopathy may be seen in association with high-dose cyclophosphamide and, although rarely, with hydroxychloroquine.

Valvular Heart Disease

Valvular abnormalities, with thickening, regurgitation, or verrucous vegetations, occur commonly in SLE (50% to 60%) and are best documented by transesophageal echocardiography.[70] They are observed more frequently in patients with high-titer APL antibodies. The characteristic Libman-Sacks lesion, nonbacterial verrucous vegetations, is noted at autopsy in 15% to 60% of patients. Mitral, aortic, and tricuspid valves are most frequently involved. Clinically, these lesions are usually asymptomatic, and hemodynamic compromise, rupture of the chordae tendineae, or infection are rare events. They are thought to result from valvulitis and subsequent healing with fibrosis and valvular thickening. On histologic examination, mononuclear cells, hematoxylin bodies, fibrin and platelet thrombi, and immune complexes are present.

Pericarditis

Pericardial inflammation in SLE is frequent. Although asymptomatic pericarditis occurs in more than 50% of patients, clinically apparent pericarditis occurs in only 25%, and cardiac tamponade and constrictive pericarditis are infrequent. Pericardial fluid and thickening are easily detected by echocardiography; cardiac silhouette enlargement on plain films is seen in the presence of large effusions. There are no unique signs and symptoms of pericarditis in lupus patients. The histologic findings of acute pericarditis in lupus are inflammation with a mononuclear cell infiltrate accompanied by immunoglobulin and complement deposition. Results of pericardial fluid analysis are neither sensitive nor specific; the fluid is usually an exudate with elevated protein concentrations, normal or low glucose levels, and an elevated WBC count that is primarily neutrophilic. Complement levels in the fluid are low, and autoantibodies (ANA, dsDNA) and LE cells have been reported.

Coronary Artery Disease

Accelerated atherosclerosis is well documented in SLE patients. Myocardial infarction, angina, and sudden death resulting from

CAD are described. Estimates of the prevalence of CAD vary widely depending on the methodology used for ascertainment. There is a 50-fold greater risk of myocardial infarction in young women with lupus compared with normal age-matched controls.[70] Cardiac events occur both late in the course of SLE and early and may even predate the SLE diagnosis. Coronary artery vasculitis is a potential cause of CAD but is exceedingly rare, and bland atherosclerotic plaque is typically described in surgical and postmortem specimens. Traditional risk factors such as hypertension, diabetes, and hyperlipidemia, are enriched in lupus patients, and an increased prevalence of the metabolic syndrome likely contributes additional risk. Lupus-related risk factors include duration of SLE, duration of corticosteroid use, renal disease, and absence of use of hydroxychloroquine. The potential contributing influence of antibodies to phospholipids or disease activity continue to be explored. Atherosclerosis is an inflammatory condition with the innate and adaptive immune system contributing to all stages of disease. Inflammation is now recognized to accelerate the formation and the rupture of atherosclerotic plaque. If endothelial dysfunction and vascular injury are the events triggering atherosclerosis, there are multiple potentially responsible processes in lupus; these include, but are not limited to, autoantibodies directed to endothelial cells, oxidized low-density lipoprotein (LDL) and/or APL antibodies, low-density granulocytes (LDGs), LDG NETs,[71] and immune complexes may contribute to accelerated atherosclerosis in SLE.

Renal Involvement

Lupus patients with kidney disease have increased morbidity and mortality compared with patients without this feature of disease. Lupus nephritis is a common with significant impact on morbidity and mortality. The prevalence of nephritis ranges from 50% to 75% overall, with increased prevalence of proliferative nephritis and more aggressive disease in African Americans, Asians, and Hispanics compared with Caucasians. Low socioeconomic status, independent of ethnicity, is predictive of poor prognosis, and pediatric lupus and male lupus are both associated with a greater incidence of, and more aggressive, nephritis. Onset of nephritis frequently occurs within 2 years after diagnosis but may occur at any time, so monitoring for potential renal activity is an ongoing obligation. Clinically, patients are asymptomatic unless they are nephrotic or have developed end-stage renal disease. Detection typically relies on examination of the urine, although a rising creatinine or hypertension may herald renal involvement. The presence of proteinuria on urinalysis, hematuria (>5 red blood cells [RBCs]/high-power field), or pyuria (>5 WBCs/high-power field) in the absence of other etiologies should prompt an evaluation for nephritis. A 24-hour urine collection remains the most accurate measurement of urinary protein loss; however, the protein/creatinine ratio in a spot urine is accepted and more commonly used for monitoring patients. Monitoring serum creatinine as a surrogate for the glomerular filtration rate is standard; however, creatinine is an insensitive marker of lupus renal disease and should be used in conjunction with other assays. Ultimately, a kidney biopsy should be performed to ascertain the type of kidney disease as well as the amount of fibrosis and degree of reversibility. Proliferative renal activity (see later) is usually preceded or accompanied by serological activity. Antibodies to dsDNA are almost always elevated or rising, whereas measurements of serum complement (C3, C4, or CH_{50}) are usually low or dropping.

Histologic findings of the kidney in lupus nephritis can be defined using the classification proposed by the International Society of Nephrology/Renal Pathology Society (ISN/RPS).[72] In general, membranous nephritis (class V) presents with a bland urinary sediment (*i.e.*, no RBCs, WBCs, or casts), nephrotic-range proteinuria, a normal to mildly elevated creatinine, a normal blood pressure, and normal serologies. Patients with mesangial disease (class II) present with a bland or minimally active sediment, low-grade proteinuria (less than 500 mg/24 hours), and a normal serologic profile. Class III (focal) and class IV (diffuse) proliferative nephritis are characterized by active urinary sediment, proteinuria (>500 mg/24 hours), active serologies, and, frequently, hypertension and elevated serum creatinine. Class III is defined as 50% or less glomerular involvement, and class IV is defined as greater than 50%. The extent of proteinuria, urinary sediment activity, serologic abnormalities, and creatinine elevation is often less in class III than in class IV renal disease. Most cases of class II nephritis do not require initiation of cytotoxic therapy, and progression to end-stage kidney disease is rare. The prognosis of class III disease is dependent on the degree of activity; patients with 40% to 50% of glomerular involvement have a prognosis similar to that of patients with class IV disease. In addition to the number of involved glomeruli, renal biopsy assessment includes measures of activity (proliferative response) and chronicity (sclerotic response). Therapeutic intervention in class III or IV disease requires cytotoxic therapy in addition to high-dose corticosteroids, provided the chronicity index is not too high, indicating irreversible damage. Even with potent immunosuppressant therapy such as cyclophosphamide or mycophenolate mofetil (MMF), a complete response is induced in only approximately 20%, with a partial response occurring in approximately 80% of patients. Moreover, repeat histopathologic studies of kidney biopsies in patients achieving a complete clinical response show ongoing renal inflammation in almost half. Relapses and flares of renal disease are not infrequent, particularly when tapering corticosteroids or discontinuing immunosuppressive treatment. Long-term prognosis is favorable in patients in whom proteinuria declines to 800 mg/day or less. Potential therapies that maintain podocyte integrity or prevent activation of renal endothelial cells as well as agents directed against inflammatory cytokines, B cells, or T cells may offer therapeutic advantage and improved renal outcome.[73]

Hematologic

Hemocytopenias occur frequently in SLE and are included in both the ACR and SLICC classification criteria; prevalence varies among lupus cohorts (Table 52.2). Most of the cytopenias are associated with increased antibody-mediated peripheral destruction; however, the bone marrow is also recognized as an immune target. Evaluation for medication effects or nutritional deficiencies is essential before attributing a cytopenia to an immune-mediated mechanism.

Anemia

Antibody-mediated peripheral destruction of RBCs, autoimmune hemolytic anemia (AHA), occurs in 5% to 14% of SLE patients. In the multiethnic Latin American Study Group for the Study of Lupus (GLADEL-(Grupo Latino Americano De Estudio del Lupus)) GLADEL cohort, AHA was an independent predictor of damage accrual and decreased survival and was associated with disease activity.[74] RBC antibodies are

usually opsonizing warm-type IgG antibodies that trigger removal in the spleen. Antigen specificities of the RBC antibodies in SLE remain elusive. The presence of AHA is readily diagnosed by an elevated reticulocyte count, positive Coombs test, elevated lactate dehydrogenase and total bilirubin, low serum haptoglobin, and the presence of spherocytes on the peripheral smear. AHA in SLE has also been associated with APL antibodies, which may reflect cross-reactivity with erythrocyte membrane antigens.[75]

Anemia of chronic disease (ACD) is the most common cause of anemia in SLE. It is characteristically normochromic and normocytic with normal or high serum ferritin and pathogenesis is related to abnormalities in iron homeostasis and/or erythropoietin response. The hepatic hormone hepcidin regulates serum iron by preventing gut absorption and iron release from hepatocytes and macrophages. Under inflammatory conditions, hepcidin production is increased, resulting in impaired erythropoiesis.[75] Low erythropoietin levels in SLE may be attributable to renal disease and antierythropoietin antibodies. Although hemophagocytosis of hematopoietic cells is frequently noted on bone marrow biopsies, the hemophagocytic syndrome characterized by spiking fevers, tender hepatosplenomegaly, anemia, leukopenia, and markedly elevated serum ferritin is rare.

Lupus is also associated with a thrombotic microangiopathic hemolytic anemia (TMHA) with schistocytes, helmet cells, and triangular fragments of RBCs on peripheral smear with a negative Coombs test. Clinical symptoms may include fever, renal insufficiency, neurologic symptoms, and thrombocytopenia, correlating with local or diffuse microvascular thrombosis. TMHA is associated with antiphospholipid syndrome (APLS), which in many cases precedes the diagnosis, prompting the observation that TMHA may be a manifestation of APLS.[76] The differential diagnosis includes catastrophic APLS, malignant hypertension, and thrombotic thrombocytopenic purpura (TTP). Coexistent TTP and SLE is a rare and frequently fatal phenomenon that has been associated with antibodies to ADAMTS 13 in 16% of SLE-associated TTP (ADAMTS 13 is *A D*isintegrin *A*nd *M*etalloprotease with *T*hrombo*S*pondin type 1 repeats; *13*th member).

Leukopenia

Leukopenia, either neutropenia or lymphopenia, occurs in 20% to 40% of patients, and attribution to SLE rather than medication effects is essential.[77] Lymphopenia has a reported prevalence of 15% to 80%, whereas severe lymphopenia (<500/mm^3) occurs in 4% to 10% of SLE patients. The prevalence of neutropenia, variably defined in the literature as less than 1800 to less than 2500/mm^3, is 20% to 40%, whereas severe neutropenia (<1000/mm^3) is rare (1% to 4% prevalence). Both neutropenia and lymphopenia can reflect disease activity; however, associations with infection remain controversial. Carli et al. found that associations between leukopenia and major infection lost significance in half of published studies after controlling for confounding factors.[77] Nonetheless, careful monitoring for infection is warranted, and prophylaxis against opportunistic infections should be considered in severe cases.

Autoantibodies are the pathogenic mechanism generally implicated in the leukopenias. Antineutrophil antibodies directed against membrane components of mature and progenitor cells, resulting in decreased phagocytosis and accelerated apoptosis, are implicated, as are antibodies against granulocyte–colony-stimulating factor (G-CSF) and hyposensitivity of myeloid cells to G-CSF. Binding of TNF-related apoptosis-inducing ligand (TRAIL) to TRAIL receptors on neutrophils also accelerates neutrophil apoptosis.[78] Lymphocytotoxic antibodies, particularly to CD4 T cells, increased apoptosis related to Fas and Fas ligand upregulation, and increased serum IL-10 levels have all been implicated in the pathogenesis of lymphopenia. B cells expressing the 9G4 idiotype found on V_H 4.34 heavy chains may be responsible for production of antilymphocyte antibodies.

Exclusion of drug effects, malignancy, or myelofibrosis is required before attribution to SLE. In the absence of recurrent infection, leukopenia in SLE rarely warrants treatment because increased steroids can also contribute to the risk of infection. Severe leukopenia has been treated with G-CSF; however, G-CSF has also been associated with disease flare in 30% of cases.

Thrombocytopenia

Low platelet counts (<100 × 10^9/mm^3) are seen in approximately 25% of patients, although severe thrombocytopenia is reported in fewer than 10%. Immune-mediated consumption is the most frequent cause, but in rare instances thrombocytopenia occurs as a manifestation of TMHA, TTP, or disseminated intravascular coagulation (DIC) or as part of the hemophagocytic syndrome, all of which are associated with high mortality and morbidity. A pathogenic role for antibodies against platelet membrane glycoproteins IIb/IIIa is well established, and glycoproteins Ia/IIa and IbIX have also been identified as antigens.[79] Other possible mechanisms include antibodies to the thrombopoietin receptor (TPOR), and APL and anti-CD40-ligand antibodies that bind to platelets, resulting in platelet sequestration.

Bone marrow involvement may present as pure red cell aplasia (PRCA), aplastic anemia, marrow fibrosis, agranulocytosis, and myelodysplastic syndrome.[75,80] PRCA attributed to SLE is associated with anti-erythropoietin antibodies and is characterized by a normocytic, normochromic anemia with low reticulocytes and normal leukocytes and platelets. Aplastic anemia, characterized by pancytopenia, is associated with antibodies to marrow progenitor cells.

Central and Peripheral Nervous System

Neurologic and psychiatric manifestations of SLE (NPSLE) are diverse. They can involve the peripheral nervous system (PNS) and/or the central nervous system (CNS) and frequently occur irrespective of systemic disease activity. All clinical manifestations of NPSLE can occur in non-autoimmune individuals; thus attribution to SLE is critically important as treatment decisions reflect attribution. Limited access to brain tissue and difficulties with assessment and attribution of individual symptoms have hindered progress in understanding SLE-related mechanisms for NPSLE and identification of reliable biomarkers for individual syndromes.

NPSLE Nomenclature and Epidemiology

In 1999, a consensus committee established by the ACR developed reporting standards, case definitions, and recommendations for laboratory and imaging studies for 19 neurologic, psychiatric, and cognitive syndromes involving the CNS and PNS. CNS syndromes are further categorized as focal or diffuse (Table 52.4). The clinical syndromes are well characterized with case definitions and attribution to SLE requires the exclusion

TABLE 52.4 Neuropsychiatric Lupus: Central and Peripheral Nervous System Syndromes

Central Nervous System		Peripheral Nervous System
Diffuse manifestations	Focal manifestations	
Cognitive dysfunction	Cerebrovascular disease	Cranial neuropathy
Mood disorder	Seizures	Autonomic neuropathy
Anxiety disorder	Aseptic meningitis	Mononeuropathy
Psychosis	Movement disorder	Polyneuropathy
Acute confusional state	Myelopathy	Myasthenia gravis
Headache	Demyelinating syndrome	Acute inflammatory demyelinating polyradiculopathy (Guillan Barré)
		Plexopathy

of comorbid illness, medication effects, or other conditions that may produce similar symptoms. Using this nomenclature, NPSLE is common overall, with a prevalence of 57% to 95%.[81] Reports of prevalence for NPSLE and individual syndromes vary widely and reflect differences in study methodology, racial and ethnic differences among cohorts, and criteria used to assess syndromes despite the ACR guidance. Nonetheless, cognitive impairment (7% to 95%) and mood (7% to 65%) and anxiety (6% to 40%) disorders occur most frequently, followed by headache (12% to 28%), seizures (7% to 20%), cerebrovascular disease (8% to 15%), and psychosis (1% to 11%). Others occur rarely (<7%), but all have significant impact on quality of life. Several attribution models have been developed that include the timing of NPSLE events relative to SLE diagnosis, presence of concurrent systemic disease activity, and selective criteria for inclusion. Autopsy studies have reported a range of neuropathologic findings including vasculopathy, micro and large ischemic infarcts, micro and large hemorrhage, microthrombi, vasculitis, myelopathy, small white matter lesions, and cortical atrophy.[82,83] Although pathologic lesions are more frequent in NPSLE, they are also found in SLE patients without diagnosed neuropsychiatric disease. Similarly, brain imaging studies using advanced imaging techniques have identified multiple gray and white matter abnormalities in SLE patients without NPSLE.[84] In aggregate, these findings support a continuum of subclinical brain involvement in SLE that is occurring even prior to SLE diagnosis. Therefore accurate assessment of true NPSLE prevalence awaits establishment of reliable and objective biomarkers for SLE-mediated mechanisms resulting in CNS or PNS pathology.

NPSLE Pathogenesis

Many of the focal CNS NPSLE syndromes occur as a result of vascular pathology: acute thrombotic events associated with APL antibodies (anticardiolipin, anti-β2 glycoprotein 1 antibodies, and lupus anticoagulant), and embolic events related to Libman-Sacks endocarditis or atherosclerotic disease, which is prevalent in SLE. More rarely, inflammation related to small-vessel vasculitis leads to intraluminal thrombosis. Proposed pathologic mechanisms for the diffuse CNS NPSLE syndromes include vasculitis, vasculopathy, immune complexes, brain-reactive autoantibodies, microglial cell activation, cytokine-induced, and/or cell-mediated inflammation and thrombosis,[81] leading to perfusion abnormalities, neuronal dysfunction, axonal damage, and microstructural damage. Given the fundamental importance of autoantibodies in SLE pathogenesis, it is not surprising that a growing list of autoantibodies have been associated with diffuse CNS disease, including anti–*N*-methyl D-aspartate receptor (NMDAR), anti–ribosomal-P (also termed anti–neuronal surface P antigen [NSPA]), anti–α tubulin, anti-Sm, and anti-RNP antibodies. These autoantibodies have been shown to either directly alter neuronal function or activate microglia (MG) cells to produce a proinflammatory state. APL antibodies have also been shown to impact neurologic function and structure through effects on vascular endothelium, separate from their prothrombotic effects. Evidence for these antibodies comes from clinical associations and murine models. Cytokines also contribute to CNS pathology. Type I IFN in particular has been associated with 50% to 90% of cases.[85] MG, resident macrophage-like cells in the brain that perform routine surveillance and debris clean-up under normal circumstances, can also promote damage. Activated MG are important pathologic mediators in other neurodegenerative diseases, and evidence from lupus mouse models and autopsy findings suggest they may play a role in NPSLE.

Presence of autoantibodies in CSF and brain tissue suggests an important role for a permeabilized blood-brain barrier (BBB) in NPSLE because antibodies are not made in the CNS and an intact BBB does not allow access to the brain. Indirect evidence for BBB disruption in SLE is supported by reports of an elevated albumin concentration gradient between cerebrospinal fluid (CSF) and plasma, an elevated IgG index, and elevated serum levels of proteins whose origins are exclusive to the brain parenchyma, such as S100B. Advanced neuroimaging studies of the BBB also suggest abnormal BBB permeability in SLE patients.[86,87] Proposed mechanisms for BBB disruption in SLE include brain endothelial cell disruption mediated by inflammatory cytokines, chemokines, complement C5a, antiendothelial cell antibodies, anti-NMDAR antibodies, and TWEAK (TNF-like weak inducer of apoptosis).[88–90] Anti–glucose-regulated protein-78 (anti-GRP78) antibodies have also been shown to alter BBB permeability through effects on brain endothelial cells.

This distinction between focal and diffuse CNS NPSLE and potential mechanisms is essential for the development of treatment strategies. Thrombotic events associated with APL antibodies are treated with anticoagulation, whereas high-dose corticosteroids and immunosuppression in combination with neuroleptic and antiseizure medications are generally required for treatment of diffuse CNS events. Because there is no intervening BBB, peripheral nerves are more accessible to circulating complement, autoantibodies, and inflammatory molecules, and vasculitis of epineural arteries is a common finding in PNS NPSLE.

NPSLE Assessment and Attribution

Currently, the diagnosis of NPSLE relies on clinical evaluation and an exhaustive search for other potential causes. CSF examination is useful for excluding infection or malignant cells. Although CSF in NPSLE can be characterized by a lymphocytosis and increased immunoglobulin with elevated total protein, IgG index, and oligoclonal bands, these abnormalities are not consistently present. Although numerous autoantibodies

and cytokines have been identified in the CSF of patients with SLE, none are specific for individual CNS NPSLE syndromes, and routine testing is not recommended. Although extremely sensitive for detection of structural and ischemic lesions associated with focal CNS NPSLE or non-SLE related pathology, MRI is frequently not helpful in diagnosing active diffuse CNS NPSLE because MRI studies are frequently normal in patients with psychiatric syndromes and global CNS dysfunction.[84]

In the CNS NPSLE syndromes, cognitive impairment tends to develop insidiously, irrespective of peripheral disease activity. CNS NPSLE syndromes may occur in isolation or in the context of globally increased disease activity. Although serologic evidence of disease activity (elevated anti-DNA antibodies and low complement) may help to diagnose CNS NPSLE, particularly if combined with other clinical signs of active disease, no serologic tests are specific for CNS NPSLE. Anti-NMDAR and anti-P antibodies have been identified in the serum, CSF, and brain of SLE patients and have been associated with cognitive and depressive syndromes and acute confusional state in some SLE patients. Advanced neuroimaging techniques, including structural and functional imaging studies, are currently evaluating associations between CNS microstructural integrity, cerebral perfusion, regional metabolism, BBB permeability, and clinical, serologic, and biologic parameters.[84]

PNS NPSLE events are recognized on the basis of clinical presentation with diagnostic studies such as electromyography and peripheral nerve biopsy.

APL antibody syndrome is described in detail in Chapter 61.

Drug-Induced Lupus

Until recently, drug-induced lupus referred to a clinical syndrome characterized by constitutional, musculoskeletal symptoms and serositis that resembles mild lupus and occurs after exposure to multiple drugs (most notably TNF inhibitors, IFN-α, procainamide, hydralazine, chlorpromazine, minocycline, and methyldopa). Drug-induced lupus has been associated with ANA and antihistone antibodies, whereas generation of more specific autoantibodies such as anti-dsDNA antibodies and hypocomplementemia are rare. Symptoms generally begin weeks to months after initiation of the inciting therapeutic agent and resolve within weeks after the drug is discontinued; autoantibodies can persist for up to 12 to 24 months. Host factors affecting drug metabolization (*e.g.*, slow acetylation of procainamide and hydralazine) and genetic predisposition contribute to the risk of developing drug-induced lupus. Multiple potential mechanisms resulting in the loss of self-tolerance have been suggested for this classic model. One of the most extensively explored is inhibition of DNA methylation with TLR activation, resulting in overexpression of costimulatory molecules such as leukocyte function–associated antigen-1 (LFA-1) on T cells and enhanced T-cell help.[91]

Since the introduction of anti-TNF agents, a different variant of drug-induced lupus has been recognized. Up to 30% of patients receiving TNF blockade develop autoantibodies, including a positive ANA and antibodies to dsDNA. However, clinical disease is uncommon, characteristically affecting the skin and joints while nephritis, and vasculitis are rare. The pathogenesis of this immunologic deregulation is not known.

Interestingly, although autoimmune inflammatory disease is well described following receipt of checkpoint inhibitors for treatment of malignancy, lupus-like syndromes occur less frequently.[92]

Treatment

THERAPEUTIC PRINCIPLES

Goals of Therapy in SLE

I. Suppression of inflammation
- Induction of remission/low lupus disease activity state (LLDAS)
- Maintenance of remission/LLDAS
- Preservation of organ function

II. Suppression of immune activation
- Modulation of the immune response

III. Prevention and management of drug-related toxicities

The goals of lupus treatment are to stop and reverse ongoing organ inflammation, to prevent or limit irreversible organ damage, and to suppress the immune response driving the inflammation and prevent flares. Therapeutic agents, and combinations thereof, are used for induction of remission, maintenance of remission, or prevention of flare. The efficacy of therapeutic agents must be balanced against their potential toxicity. Thus treatment must be tailored to the individual patient based on disease manifestations. In general, milder disease requires treatment with less potent or lower doses of antiinflammatory and immunosuppressive medications than more active, severe disease affecting major organs such as the kidney or brain. Individual patient responses to a given medication will vary, and patients must be monitored closely for response as well as toxicity. "Treat-to-target" recommendations in SLE from an international SLE task force include the following more specific guidelines: the treatment target should be remission of organ manifestations; factors such as pain that negatively influences health-related quality of life should be addressed; maintenance treatment should aim for the lowest glucocorticoid dose needed to control disease; and relevant therapies adjuvant to any immunomodulation should be considered to control comorbidity in SLE.[93] Low lupus disease activity state (LLDAS) is a less stringent target. Achieving LLDAS is associated with decreased accrual of damage, flare reduction, and improved quality of life.[94]

Some genetic factors predicting risk of toxicity or therapeutic benefit for individual agents have been identified. Polymorphisms of a key enzyme in the metabolism of azathioprine, thiopurine methyltransferase (TPMT), are common; 0.3% and 11% of Caucasians are homozygous and heterozygous, respectively, for mutations associated with altered expression of TPMT.[95] TPMT-deficient patients are especially susceptible to leukopenia and pancytopenia associated with azathioprine. Cyclophosphamide is metabolized to its active form by cytochrome P450. Individuals heterozygous or homozygous for a specific cytochrome P450 polymorphism *(CYP2C19*2)* have a lower probability of developing premature ovarian failure, but they also show a poorer response to therapy.[96]

Although corticosteroids are the foundation of treatment, exposure must be minimized to the greatest extent given their multiple and frequent side effects, including hypertension, diabetes mellitus, increased susceptibility to infection, bone loss, and weight gain (Chapter 83). Additional disease-modifying agents that are also steroid-sparing include antimalarials (hydroxychloroquine), antimetabolites such as azathioprine (1 to 2.5 mg/kg per day), methotrexate (7.5 to 25 mg/week), leflunomide (10 to 20 mg/day), MMF (2 to 3 g/day), and alkylating agents such as cyclophosphamide (monthly pulse

0.5 to 1.0 g/m^2) (Chapter 84). When more conventional therapies have failed, anecdotal reports, case series, and open-label studies suggest the use of intravenous immunoglobulin, thalidomide (50 to 100 mg/day), tacrolimus or cyclosporine. Medications such as dapsone, danazol, colchicine, and chlorambucil may be efficacious in cutaneous disease, hematologic disease, serositis, or arthritis and in severe refractory disease, respectively, but in general these medications are not commonly used due to their toxicity and the introduction and availability of better-tolerated, efficacious agents. Belimumab, a monoclonal antibody directed against BAFF (Chapter 84), was approved by the US Food and Drug Administration (FDA) for treatment of SLE in 2011 and represents a therapeutic option for patients with non-CNS active disease, including nephritis.[97]

Several newer agents are on the horizon for treatment of SLE. Some are currently available but approved for other indications, whereas others are still in the developmental stages. These therapies generally have more specific immunologic targets than standard treatments. Anifrolumab, a monoclonal antibody directed against the IFN-α receptor, was approved in 2021 for moderate to severe SLE; a study in lupus nephritis is ongoing. Voclosporin, a second-generation calcineurin inhibitor, was also approved for treatment of lupus nephritis in 2021.[97] Several new agents have been recently evaluated in phase III clinical trials and have been unable to demonstrate superiority over treatment with placebo or standard of care (SOC). Although imperfect trial design and imperfect outcome measures may account for some of these failures, the possibility remains that these agents are not effective treatments for SLE and do not confer a substantial improvement over current SOC therapies. Rituximab (anti-CD20 antibody), a B-cell–directed therapy approved for use in rheumatoid arthritis, targeting all B cells except plasma cells, failed to demonstrate improved performance over SOC in trials of patients with active SLE, or in patients with active lupus nephritis added to a background therapy of MMF and corticosteroids. Elevated BAFF levels are present after B-cell depletion that can facilitate maturation of autoreactive cells and exacerbate the impaired tolerance of SLE. This has led to ongoing evaluation of sequential therapy with belimumab following B-cell depletion.[98] Although conceptually attractive, the list of agents that have failed to show efficacy over standard of care or have been associated with increased toxicity are many and include abatacept (CTLA4-Ig), an approved (rheumatoid arthritis) monoclonal antibody which blocks T-cell activation; epratuzumab (anti-CD22 antibody), which targets all B cells by blocking both BAFF and APRIL; atacicept, a monoclonal antibody directed against BAFF and April (however, efficacy was demonstrated in a subset of patients with higher baseline disease activity); tabalumab and blisibumab, monoclonal antibodies directed against BAFF; ocrelizumab, an approved (multiple sclerosis) anti-CD20 monoclonal antibody; tocilizumab, an approved antibody to the IL-6 receptor, a cytokine involved in B-cell survival and activation as well as in differentiation of Th17 and Tfh cells; monoclonal antibodies to IFN-α; and ustikinumab, an approved monoclonal antibody targeting IL-12 and IL-23. Janus tyrosine kinase (JAK) inhibitors are approved for rheumatoid arthritis and are under study for discoid lupus (topical administration) and for systemic disease (given orally). Other agents in clinical development include CD40 and CD40 ligand antibodies, tyrosine kinase 2 (Tyk 2) inhibitors, Bruton tyrosine kinase inhibitors, SYK inhibitors, second-generation B-cell depleters (obinutuzumab), and immunoproteasome inhibitors.

Potential therapies aimed at DCs include vitamin D, which blocks DC maturation and T-cell activation, and inhibitory oligodeoxynucleotides, which block TLR9 signaling and DC maturation. However, a study of vitamin D supplementation designed to assess vitamin D inhibition of DC activation failed to demonstrate efficacy. Hydroxychloroquine, a standard agent for lupus, interferes with TLR7 and TLR9 signaling by preventing acidification of the endosomal compartment. A study of a novel RNase is under way and a phase 1 trial of anti-BDCA2 receptors on pDCs, inhibiting IFN-α production, demonstrated efficacy in CLE patients.[99]

Approved therapies or those in development for other autoimmune diseases may also prove useful in SLE. Eculizumab, currently approved for paroxysmal nocturnal hemoglobinuria, is an antibody directed against C5 that blocks cleavage of C5 and the subsequent triggering of the complement cascade. Alicaforsen, an antisense oligodeoxynucleotide that inhibits ICAM-1 expression, decreases inflammation in both rheumatoid arthritis and Crohn disease. Efalizumab, a monoclonal antibody against CD11a (an LFA-1 subunit that interacts with ICAM-1), benefits patients with psoriasis. Although no chemokine-targeted therapies are in clinical trial in lupus, a CCR1 antagonist slowed disease progression in a mouse model of lupus and has been tested in patients with rheumatoid arthritis. FTY720 (fingolimod), an agonist for the sphingosine 1 phosphate receptor that prevents egress of lymphocytes from secondary lymphoid organs and inflamed tissues, has beneficial effects in the MRL/*lpr* mouse model of lupus; it has been given to transplant recipients and to patients with multiple sclerosis and may provide benefit in CNS disease. The use of anti-TNF agents for SLE is complicated by its lupus-inducing effects in some patients. Paradoxically, TNF is involved in renal inflammation in SLE, and short-term treatment with monoclonal antibody to TNF appears to improve lupus nephritis, showing that agents can abort inflammation and yet exacerbate autoimmunity.

TRANSLATIONAL RESEARCH

ON THE HORIZON

- Increased understanding of the functional consequences of single nucleotide polymorphisms identified as susceptibility loci in systemic lupus erythematosus may provide insight into:
- Dysregulation of tolerance mechanisms and pathogenesis of disease in different ethnic populations
- Predictors of response to therapy
- New therapeutic targets
- Development of organ-specific biomarkers will allow for improved use of immunomodulating and immunosuppressive therapies and possibly for use of preventive therapies.

Laboratory research continues to inform us about dysregulated immunologic pathways in SLE. The current challenge of translational medicine is to shed light on the clinical relevance of these altered pathways. Increased understanding of the qualitative and quantitative differences of these molecular perturbations not only should result in improved diagnostic capabilities and biomarkers for disease but also will lead to the ability to subset patients for prognosis and response to therapy. Improved organ-specific biomarkers and methods for subsetting patients

on a molecular level are likely to result in improved use of immunologic therapies. In particular the brain, which has been relatively inaccessible to mechanistic and diagnostic probing, is likely to be an organ to be more fully explored. Ultimately, advances in translational studies should result in more effective and less toxic therapeutic interventions.

CONCLUSIONS

In the past several years, much has been learned about genetic susceptibility to SLE, and it is encouraging that many of the genes identified are associated with pathways that also have been implicated in disease pathogenesis. The role of B cells in disease pathogenesis has been confirmed, but recent studies have also highlighted the role of T cells, DCs, and neutrophils. Enhanced understanding of each of these contributing factors and the crosstalk between them has allowed identification of numerous potential therapeutic targets. The disappointing results from clinical trials of immunobiologic agents for lupus have highlighted the critical importance of study design and the complexity of disease in humans. The overall goal in therapy must be to eliminate autoreactivity while maintaining immunocompetence. Among the challenges now faced are the careful phenotyping of patients to identify etiopathologically distinct subpopulations and new clinical trial designs to allow the use of combinations of agents, each of which alone may have a minor effect on disease course.

REFERENCES

1. Petri M, Orbai AM, Alarcon GS, Gordon C, Merrill JT, Fortin PR, et al. Derivation and validation of the Systemic Lupus International Collaborating Clinics classification criteria for systemic lupus erythematosus. *Arthritis Rheum*. 2012;64(8):2677–2686.
2. Aringer M, Costenbader K, Daikh D, Brinks R, Mosca M, Ramsey-Goldman R, et al. 2019 European League Against Rheumatism/American College of Rheumatology Classification Criteria for Systemic Lupus Erythematosus. *Arthritis Rheumatol*. 2019;71(9):1400–1412.
3. Stojan G, Petri M. Epidemiology of systemic lupus erythematosus: an update. *Curr Opin Rheumatol*. 2018;30(2):144–150.
4. Gonzalez LA, Toloza SM, Alarcon GS. Impact of race and ethnicity in the course and outcome of systemic lupus erythematosus. *Rheumatic Diseases Clinics of North America*. 2014;40(3):433–454. vii–viii.
5. Ocampo-Piraquive V, Nieto-Aristizabal I, Canas CA, Tobon GJ. Mortality in systemic lupus erythematosus: causes, predictors and interventions. *Expert Rev Clin Immunol*. 2018;14(12):1043–1053.
6. Gatto M, Iaccarino L, Ghirardello A, Punzi L, Doria A. Clinical and pathologic considerations of the qualitative and quantitative aspects of lupus nephritogenic autoantibodies: A comprehensive review. *J Autoimmun*. 2016;69:1–11.
7. Gregersen PK, Olsson LM. Recent advances in the genetics of autoimmune disease. *Annu Rev Immunol*. 2009;27:363–391.
8. Ueda H, Howson JM, Esposito L, Heward J, Snook H, Chamberlain G, et al. Association of the T-cell regulatory gene CTLA4 with susceptibility to autoimmune disease. *Nature*. 2003;423(6939):506–511.
9. Lim J, Kim K. Genetic variants differentially associated with rheumatoid arthritis and systemic lupus erythematosus reveal the disease-specific biology. *Sci Rep*. 2019;9(1):2739.
10. Orozco G, Sanchez E, Gonzalez-Gay MA, Lopez-Nevot MA, Torres B, Caliz R, et al. Association of a functional single-nucleotide polymorphism of PTPN22, encoding lymphoid protein phosphatase, with rheumatoid arthritis and systemic lupus erythematosus. *Arthritis and Rheumatism*. 2005;52(1):219–224.
11. Manderson AP, Botto M, Walport MJ. The role of complement in the development of systemic lupus erythematosus. *Annu Rev Immunol*. 2004;22:431–456.
12. Baumann I, Kolowos W, Voll RE, Manger B, Gaipl U, Neuhuber WL, et al. Impaired uptake of apoptotic cells into tingible body macrophages in germinal centers of patients with systemic lupus erythematosus. *Arthritis and Rheumatism*. 2002;46(1):191–201.
13. Anolik JH, Barnard J, Cappione A, Pugh-Bernard AE, Felgar RE, Looney RJ, et al. Rituximab improves peripheral B cell abnormalities in human systemic lupus erythematosus. *Arthritis and Rheumatism*. 2004;50(11):3580–3590.
14. Mackay M, Stanevsky A, Wang T, Aranow C, Li M, Koenig S, et al. Selective dysregulation of the Fc{gamma}IIB receptor on memory B cells in SLE. *J Exp Med*. 2006
15. Wong M, Tsao BP. Current topics in human SLE genetics. *Springer Semin Immunopathol*. 2006
16. Tucci M, Barnes EV, Sobel ES, Croker BP, Segal MS, Reeves WH, et al. Strong association of a functional polymorphism in the monocyte chemoattractant protein 1 promoter gene with lupus nephritis. *Arthritis and Rheumatism*. 2004;50(6):1842–1849.
17. Crow MK. Interferon pathway activation in systemic lupus erythematosus. *Curr Rheumatol Rep*. 2005;7(6):463–468.
18. Nath SK, Harley JB, Lee YH. Polymorphisms of complement receptor 1 and interleukin-10 genes and systemic lupus erythematosus: a meta-analysis. *Hum Genet*. 2005;118(2):225–234.
19. Schotte H, Willeke P, Tidow N, Domschke W, Assmann G, Gaubitz M, et al. Extended haplotype analysis reveals an association of TNF polymorphisms with susceptibility to systemic lupus erythematosus beyond HLA-DR3. *Scand J Rheumatol*. 2005;34(2):114–121.
20. Wu J, Metz C, Xu X, Abe R, Gibson AW, Edberg JC, et al. A novel polymorphic CAAT/enhancer-binding protein beta element in the FasL gene promoter alters Fas ligand expression: a candidate background gene in African American systemic lupus erythematosus patients. *J Immunol*. 2003;170(1):132–138.
21. Mehrian R, Quismorio Jr. FP, Strassmann G, Stimmler MM, Horwitz DA, Kitridou RC, et al. Synergistic effect between IL-10 and bcl-2 genotypes in determining susceptibility to systemic lupus erythematosus. *Arthritis and Rheumatism*. 1998;41(4):596–602.
22. Ballestar E, Esteller M, Richardson BC. The epigenetic face of systemic lupus erythematosus. *J Immunol*. 2006;176(12):7143–7147.
23. Chung SA, Nititham J, Elboudwarej E, Quach HL, Taylor KE, Barcellos LF, et al. Genome-Wide Assessment of Differential DNA Methylation Associated with Autoantibody Production in Systemic Lupus Erythematosus. *PloS One*. 2015;10(7):e0129813.
24. Javierre BM, Fernandez AF, Richter J, Al-Shahrour F, Martin-Subero JI, Rodriguez-Ubreva J, et al. Changes in the pattern of DNA methylation associate with twin discordance in systemic lupus erythematosus. *Genome Res*. 2010;20(2):170–179.
25. Jeffries MA, Sawalha AH. Epigenetics in systemic lupus erythematosus: leading the way for specific therapeutic agents. *Int J Clin Rheumtol*. 2011;6(4):423–439.
26. Watad A, Neumann SG, Soriano A, Amital H, Shoenfeld Y. Vitamin D and Systemic Lupus Erythematosus: Myth or Reality? *Isr Med Assoc J*. 2016;18(3-4):177–182.
27. Milner EC, Anolik J, Cappione A, Sanz I. Human innate B cells: a link between host defense and autoimmunity? *Springer Semin Immunopathol*. 2005;26(4):433–452.
28. Suurmond J, Atisha-Fregoso Y, Marasco E, Barlev AN, Ahmed N, Calderon SA, et al. Loss of an IgG plasma cell checkpoint in patients with lupus. The. *Journal of Allergy and Clinical Immunology*. 2019;143(4):1586–1597.
29. Liu Z, Davidson A. BAFF and selection of autoreactive B cells. *Trends Immunol*. 2011;32(8):388–394.
30. Doreau A, Belot A, Bastid J, Riche B, Trescol-Biemont MC, Ranchin B, et al. Interleukin 17 acts in synergy with B cell-activating factor to influence B cell biology and the pathophysiology of systemic lupus erythematosus. *Nat Immunol*. 2009;10(7):778–785.
31. Pugh-Bernard AE, Cambier JC. B cell receptor signaling in human systemic lupus erythematosus. *Curr Opin Rheumatol*. 2006;18(5):451–455.
32. Floto RA, Clatworthy MR, Heilbronn KR, Rosner DR, MacAry PA, Rankin A, et al. Loss of function of a lupus-associated FcgammaRIIb polymorphism through exclusion from lipid rafts. *Nat Med*. 2005;11(10):1056–1058.

33. Karnell JL, Kumar V, Wang J, Wang S, Voynova E, Ettinger R. Role of CD11c(+) T-bet(+) B cells in human health and disease. *Cellular Immunology*. 2017;321:40–45.
34. Lande R, Ganguly D, Facchinetti V, Frasca L, Conrad C, Gregorio J, et al. Neutrophils activate plasmacytoid dendritic cells by releasing self-DNA-peptide complexes in systemic lupus erythematosus. *Sci Transl Med*. 2011;3(73):73ra19.
35. Garcia-Romo GS, Caielli S, Vega B, Connolly J, Allantaz F, Xu Z, et al. Netting neutrophils are major inducers of type I IFN production in pediatric systemic lupus erythematosus. *Sci Transl Med*. 2011;3(73):73ra20.
36. Bennett L, Palucka AK, Arce E, Cantrell V, Borvak J, Banchereau J, et al. Interferon and granulopoiesis signatures in systemic lupus erythematosus blood. *J Exp Med*. 2003;197(6):711–723.
37. Kis-Toth K, Tsokos GC. Dendritic cell function in lupus: Independent contributors or victims of aberrant immune regulation. *Autoimmunity*. 2010;43(2):121–130.
38. Crispin JC, Kyttaris VC, Terhorst C, Tsokos GC. T cells as therapeutic targets in SLE. *Nat Rev Rheumatol*. 2010;6(6):317–325.
39. Scheinecker C, Bonelli M, Smolen JS. Pathogenetic aspects of systemic lupus erythematosus with an emphasis on regulatory T cells. *J Autoimmun*. 2010;35(3):269–275.
40. Ohl K, Tenbrock K. Regulatory T cells in systemic lupus erythematosus. *Eur J Immunol*. 2015;45(2):344–355.
41. Brandt D, Hedrich CM. TCRalphabeta(+)CD3(+)CD4(-)CD8(-) (double negative) T cells in autoimmunity. *Autoimmun Rev*. 2018;17(4):422–430.
42. McMurray RW. Bromocriptine in rheumatic and autoimmune diseases. *Semin Arthritis Rheum*. 2001;31(1):21–32.
43. Grimaldi CM, Hill L, Xu X, Peeva E, Diamond B. Hormonal modulation of B cell development and repertoire selection. *Mol Immunol*. 2005;42(7):811–820.
44. Zhang W, Wu K, He W, Gao Y, Huang W, Lin X, et al. Transforming growth factor beta 1 plays an important role in inducing CD4(+)CD25(+) forhead box P3(+) regulatory T cells by mast cells. *Clin Exp Immunol*. 2010;161(3):490–496.
45. Cunningham M, Gilkeson G. Estrogen receptors in immunity and autoimmunity. *Clin Rev Allergy Immunol*. 2011;40(1):66–73.
46. Mu Q, Zhang H, Luo XM. SLE: Another Autoimmune Disorder Influenced by Microbes and Diet? *Front Immunol*. 2015;6:608.
47. Zoma A. Musculoskeletal involvement in systemic lupus erythematosus. *Lupus*. 2004;13(11):851–853.
48. Urowitz MB, Gladman DD, Ibanez D, Sanchez-Guerrero J, Romero-Diaz J, Gordon C, et al. American College of Rheumatology criteria at inception, and accrual over 5 years in the SLICC inception cohort. *J Rheumatol*. 2014;41(5):875–880.
49. Grossman JM. Lupus arthritis. *Best Pract Res Clin Rheumatol*. 2009;23(4):495–506.
50. Aranow C, Zelicof S, Leslie D, Solomon S, Barland P, Norman A, et al. Clinically occult avascular necrosis of the hip in systemic lupus erythematosus. *J Rheumatol*. 1997;24(12):2318–2322.
51. Fialho SC, Bonfa E, Vitule LF, D'Amico E, Caparbo V, Gualandro S, et al. Disease activity as a major risk factor for osteonecrosis in early systemic lupus erythematosus. *Lupus*. 2007;16(4):239–244.
52. Hong J, Aspey L, Bao G, Haynes T, Lim SS, Drenkard C. Chronic Cutaneous Lupus Erythematosus: Depression Burden and Associated Factors. *Am J Clin Dermatol*. 2019;20(3):465–475.
53. Werth VP, Merrill JT. A double-blind, randomized, placebo-controlled, phase II trial of baricitinib for systemic lupus erythematosus: how to optimize lupus trials to examine effects on cutaneous lupus erythematosus. *Br J Dermatol*. 2019;180(5):964–965.
54. Jarrett P, Werth VP. A review of cutaneous lupus erythematosus: improving outcomes with a multidisciplinary approach. *J Multidiscip Healthc*. 2019;12:419–428.
55. Kahlenberg JM. Rethinking the Pathogenesis of Cutaneous Lupus. *The Journal of Investigative Dermatology*. 2021;141(1):32–35.
56. Kunz M, Konig IR, Schillert A, Kruppa J, Ziegler A, Grallert H, et al. Genome-wide association study identifies new susceptibility loci for cutaneous lupus erythematosus. *Experimental Dermatology*. 2015;24(7):510–515.
57. Vermi W, Lonardi S, Morassi M, Rossini C, Tardanico R, Venturini M, et al. Cutaneous distribution of plasmacytoid dendritic cells in lupus erythematosus. Selective tropism at the site of epithelial apoptotic damage. *Immunobiology*. 2009;214(9-10):877–886.
58. Berthier CC, Tsoi LC, Reed TJ, Stannard JN, Myers EM, Namas R, et al. Molecular Profiling of Cutaneous Lupus Lesions Identifies Subgroups Distinct from Clinical Phenotypes. *J Clin Med*. 2019;8(8).
59. Stannard JN, Kahlenberg JM. Cutaneous lupus erythematosus: updates on pathogenesis and associations with systemic lupus. *Curr Opin Rheumatol*. 2016;28(5):453–459.
60. Concha JSS, Werth VP. Alopecias in lupus erythematosus. *Lupus Sci Med*. 2018;5(1):e000291.
61. Orteu CH, Buchanan JA, Hutchison I, Leigh IM, Bull RH. Systemic lupus erythematosus presenting with oral mucosal lesions: easily missed? *Br J Dermatol*. 2001;144(6):1219–1223.
62. Brewer BN, Kamen DL. Gastrointestinal and Hepatic Disease in Systemic Lupus Erythematosus. *Rheum Dis Clin North Am*. 2018;44(1):165–175.
63. Li Z, Xu D, Wang Z, Wang Y, Zhang S, Li M, et al. Gastrointestinal system involvement in systemic lupus erythematosus. *Lupus*. 2017;26(11):1127–1138.
64. Ebert EC, Hagspiel KD. Gastrointestinal and hepatic manifestations of systemic lupus erythematosus. *J Clin Gastroenterol*. 2011;45(5):436–441.
65. Janssens P, Arnaud L, Galicier L, Mathian A, Hie M, Sene D, et al. Lupus enteritis: from clinical findings to therapeutic management. *Orphanet J Rare Dis*. 2013;8:67.
66. Mackay IR. Historical reflections on autoimmune hepatitis. *World J Gastroenterol*. 2008;14(21):3292–3300.
67. Medlin JL, Hansen KE, McCoy SS, Bartels CM. Pulmonary manifestations in late versus early systemic lupus erythematosus: A systematic review and meta-analysis. *Semin Arthritis Rheum*. 2018;48(2):198–204.
68. Mittoo S, Fell CD. Pulmonary manifestations of systemic lupus erythematosus. *Semin Respir Crit Care Med*. 2014;35(2):249–254.
69. Guo L, Li M, Chen Y, Wang Q, Tian Z, Pan S, et al. Anti-Endothelin Receptor Type A Autoantibodies in Systemic Lupus Erythematosus-Associated Pulmonary Arterial Hypertension. *Arthritis Rheumatol*. 2015;67(9):2394–2402.
70. Miner JJ, Kim AH. Cardiac manifestations of systemic lupus erythematosus. *Rheumatic diseases clinics of North America*. 2014;40(1):51–60.
71. Carmona-Rivera C, Kaplan MJ. Low-density granulocytes: a distinct class of neutrophils in systemic autoimmunity. *Semin Immunopathol*. 2013;35(4):455–463.
72. Weening JJ, D'Agati VD, Schwartz MM, Seshan SV, Alpers CE, Appel GB, et al. The classification of glomerulonephritis in systemic lupus erythematosus revisited. *J Am Soc Nephrol*. 2004;15(2):241–250.
73. Davidson A, Aranow C. Lupus nephritis: lessons from murine models. *Nat Rev Rheumatol*. 2010;6(1):13–20.
74. Gonzalez-Naranjo LA, Betancur OM, Alarcon GS, Ugarte-Gil MF, Jaramillo-Arroyave D, Wojdyla D, et al. Features associated with hematologic abnormalities and their impact in patients with systemic lupus erythematosus: Data from a multiethnic Latin American cohort. *Semin Arthritis Rheum*. 2016;45(6):675–683.
75. Velo-Garcia A, Castro SG, Isenberg DA. The diagnosis and management of the haematologic manifestations of lupus. *J Autoimmun*. 2016;74:139–160.
76. Espinosa G, Bucciarelli S, Cervera R, Lozano M, Reverter JC, de la Red G, et al. Thrombotic microangiopathic haemolytic anaemia and antiphospholipid antibodies. *Annals of the Rheumatic Diseases*. 2004;63(6):730–736.
77. Carli L, Tani C, Vagnani S, Signorini V. Mosca M. Leukopenia, lymphopenia, and neutropenia in systemic lupus erythematosus: Prevalence and clinical impact--A systematic literature review. *Semin Arthritis Rheum*. 2015;45(2):190–194.
78. Matsuyama W, Yamamoto M, Higashimoto I, Oonakahara K, Watanabe M, Machida K, et al. TNF-related apoptosis-inducing ligand is involved in neutropenia of systemic lupus erythematosus. *Blood*. 2004;104(1):184–191.
79. Michel M, Lee K, Piette JC, Fromont P, Schaeffer A, Bierling P, et al. Platelet autoantibodies and lupus-associated thrombocytopenia. *Br J Haematol*. 2002;119(2):354–358.

80. Chalayer E, Costedoat-Chalumeau N, Beyne-Rauzy O, Ninet J, Durupt S, Tebib J, et al. Bone marrow involvement in systemic lupus erythematosus. *QJM*. 2017;110(11):701–711.
81. Schwartz N, Stock AD, Putterman C. Neuropsychiatric lupus: new mechanistic insights and future treatment directions. *Nat Rev Rheumatol*. 2019;15(3):137–152.
82. Cohen D, Rijnink EC, Nabuurs RJ, Steup-Beekman GM, Versluis MJ, Emmer BJ, et al. Brain histopathology in patients with systemic lupus erythematosus: identification of lesions associated with clinical neuropsychiatric lupus syndromes and the role of complement. *Rheumatology (Oxford)*. 2017;56(1):77–86.
83. Sibbitt Jr. WL, Brooks WM, Kornfeld M, Hart BL, Bankhurst AD, Roldan CA. Magnetic resonance imaging and brain histopathology in neuropsychiatric systemic lupus erythematosus. *Semin Arthritis Rheum*. 2010;40(1):32–52.
84. Mackay M, Tang CC, Vo A. Advanced neuroimaging in neuropsychiatric systemic lupus erythematosus. *Curr Opin Neurol*. 2020;33(3):353–361.
85. Ronnblom L, Leonard D. Interferon pathway in SLE: one key to unlocking the mystery of the disease. *Lupus Sci Med*. 2019;6(1):e000270.
86. Chi JM, Mackay M, Hoang A, Cheng K, Aranow C, Ivanidze J, et al. Alterations in Blood-Brain Barrier Permeability in Patients with Systemic Lupus Erythematosus. *Ajnr*. 2019;40(3):470–477.
87. Kamintsky L, Beyea SD, Fisk JD, Hashmi JA, Omisade A, Calkin C, et al. Blood-brain barrier leakage in systemic lupus erythematosus is associated with gray matter loss and cognitive impairment. *Annals of the Rheumatic Diseases*. 2020;79(12):1580–1587.
88. Abbott NJ, Mendonca LL, Dolman DE. The blood-brain barrier in systemic lupus erythematosus. *Lupus*. 2003;12(12):908–915.
89. O'Carroll SJ, Kho DT, Wiltshire R, Nelson V, Rotimi O, Johnson R, et al. Pro-inflammatory TNFalpha and IL-1beta differentially regulate the inflammatory phenotype of brain microvascular endothelial cells. *J Neuroinflammation*. 2015;12:131.
90. Stock AD, Wen J, Putterman C. Neuropsychiatric Lupus, the Blood Brain Barrier, and the TWEAK/Fn14 Pathway. *Front Immunol*. 2013;4:484.
91. Kaplan MJ, Deng C, Yang J, Richardson BC. DNA methylation in the regulation of T cell LFA-1 expression. *Immunological Investigations*. 2000;29(4):411–425.
92. Raschi E, Antonazzo IC, Poluzzi E, De Ponti F. Drug-induced systemic lupus erythematosus: should immune checkpoint inhibitors be added to the evolving list? *Ann Rheum Dis*. 2019
93. van Vollenhoven RF, Mosca M, Bertsias G, Isenberg D, Kuhn A, Lerstrom K, et al. Treat-to-target in systemic lupus erythematosus: recommendations from an international task force. *Annals of the Rheumatic Diseases*. 2014;73(6):958–967.
94. Golder V, Kandane-Rathnayake R, Hoi AY, Huq M, Louthrenoo W, An Y, et al. Association of the lupus low disease activity state (LLDAS) with health-related quality of life in a multinational prospective study. *Arthritis Research & Therapy*. 2017;19(1):62.
95. Yates CR, Krynetski EY, Loennechen T, Fessing MY, Tai HL, Pui CH, et al. Molecular diagnosis of thiopurine S-methyltransferase deficiency: genetic basis for azathioprine and mercaptopurine intolerance. *Ann Intern Med*. 1997;126(8):608–614.
96. Takada K, Arefayene M, Desta Z, Yarboro CH, Boumpas DT, Balow JE, et al. Cytochrome P450 pharmacogenetics as a predictor of toxicity and clinical response to pulse cyclophosphamide in lupus nephritis. *Arthritis and Rheumatism*. 2004;50(7):2202–2210.
97. Parodis I, Houssiau FA. From sequential to combination and personalised therapy in lupus nephritis: moving towards a paradigm shift? *Annals of the Rheumatic Diseases*. 2021
98. Dall'Era M, Aranow C, Byron M, Ding L, Smilek D, Diamond B, Wofsy D. Phase 2 Trial of Induction Therapy with Anti-CD20 (Rituximab) Followed By Maintenance Therapy with Anti-BAFF (Belimumab) in Patients with Active Lupus Nephritis. *Arthritis Rheumatol*. 2018;70(suppl 10).
99. Furie R, Werth VP, Merola JF, Stevenson L, Reynolds TL, Naik H, et al. Monoclonal antibody targeting BDCA2 ameliorates skin lesions in systemic lupus erythematosus. *J Clin Invest*. 2019;129(3):1359–1371.

Rheumatoid Arthritis

Andrew P. Cope

Rheumatoid arthritis (RA) is one of the most common chronic inflammatory diseases and in modern times has become a prototype disease entity for defining the molecular and pathological basis of chronic inflammatory syndromes. The term *rheumatoid arthritis* was coined by Garrod in 1859. However, this was probably an inappropriate use of the term because it encompassed polyarticular osteoarthritis as well as inflammatory polyarthritis. In spite of references to inflammatory afflictions of joints by the likes of Galen, Sydenham, and Heberden, the first convincing case reports of the disease, described in terms that would be recognizable today, were published in 1800 by Landré-Beauvais, who labeled the disease "la goutte asthénique primitive." This description was distinct because all patients were female, an observation that was significant when the most important differential diagnosis at that time was polyarticular gout, a disease predominantly of males.

Today we recognize RA as a chronic inflammatory disorder of joints of unknown etiology in which the major target tissue is the synovial lining of joints, bursae, and tendon sheaths. Although traditionally considered an autoimmune disease, RA differs from organ-specific autoimmune disease entities in several respects.[1] From the outset of clinically apparent disease, the systemic immuno-inflammatory process, driven by cytokines and other inflammatory mediators, promotes the activation and proliferation of stromal joint tissues, in particular the fibroblastic synovial lining layer. This appears to contrast with organ-specific autoimmune diseases, such as type 1 diabetes or autoimmune thyroiditis, characterized by an antigen-driven immune-inflammatory response *in situ* leading to targeted cellular destruction of autoantigen-expressing pancreatic β-islet or thyroid tissue cells. In RA, once the inflammatory process is established, the inflammatory synovium, or pannus, may invade and erode underlying cartilage and bone. Unlike autoimmune diseases that target single organs or tissues, RA is a systemic inflammatory disease that likely encompasses a heterogeneous syndrome with marked variation in clinical expression that most clinicians today would acknowledge is more than one disease entity. Indeed, it is now apparent that the disease is heterogeneous not only clinically but also pathologically, serologically, and genetically.

EPIDEMIOLOGY

The incidence of RA (the rate of new cases arising in a given period) is approximately 0.4 per 1000 when the 2010 American College of Rheumatology/European League Against Rheumatism (ACR/EULAR) classification criteria are applied. Large cross-sectional population samples indicate that disease prevalence, which ideally should include all past and inactive cases, ranges from 0.5% to 2% for Caucasian European and North American populations over the age of 15, with a female to male excess of 2 to 4 times. Rates plateau between the ages of 45 and 75 years in some series, but can increase steadily with age until the seventh decade, declining thereafter. Despite similar prevalence estimates for these geographically diverse populations, greater diversity has been documented for rural African populations, where the prevalence has been reported to be as low as 0.1%, and for Native Americans (including the Pima, Yakima, and Chippewa tribes), where the prevalence may be as high as 5%. Such variance across geographical borders likely reflects distinct environmental factors and sociodemographic determinants, as well as a spectrum of genetic admixture. Lifetime risk has been estimated at 2% for males and 4% for females.

Complex polygenic autoimmune syndromes like RA are diseases of low penetrance, where thresholds of disease expression may be higher in males. Recent insights into familial clustering indicate that family history remains a strong independent risk factor for RA, but that this risk does not differ by gender.[2] This implies that nongenetic factors influence gender bias. Twin studies also provide compelling evidence for genetic effects, given the excess concordance rates for monozygotic (12% to 15%) compared with dizygotic twins (<5%, and probably nearer to 3.5%).[2] These concordance rates appear somewhat low, but compared with a background prevalence of 1% in outbred populations, genetic epidemiology studies report heritability estimates of 68% for those patients who carry antibodies to citrullinated protein antigens ($ACPAs^+$) and 66% for those who do not ($ACPAs^-$), indicating substantial genetic influence.[3] Human leukocyte antigen (HLA) is thought to contribute 35% to 40% of this value. This "missing heritability" leaves a substantial contribution to disease susceptibility from environmental factors, influenced by occupation, socioeconomic status, exposure to infectious pathogens, and lifestyle factors—a conglomeration of factors termed the "exposome."

Two of the more intriguing factors contributing to disease occurrence are age and gender. Age-associated changes in susceptibility to infection, neoplastic disease, and autoimmunity suggest that a common mechanism could be responsible. Immune senescence is one possibility, where age-related decline in host immunity is characterized at the cellular and molecular level by expansions of lymphocyte clones, corresponding contractions of the naïve T- and B-cell repertoires linked to depletion of lymphocyte precursors in thymus and bone marrow, and telomere erosion of leukocytes, features indicative of an extensive proliferative history. When combined with dysregulation of costimulatory receptors such as CD28, oxidative stress,

KEY CONCEPTS

Important Risk Factors for Developing Rheumatoid Arthritis

- Female gender; impact of X chromosome, micro-chimerism, lifestyle
- Age; associated with accelerated immune aging
- Inheritance of genetic variants, *e.g., HLADRB1* and *PTPN22*
- Autoantibodies to modified protein antigens (AMPAs), rheumatoid factor
- Family history; first-degree relatives have higher prevalence of genetic and serological risk factors
- Hormonal factors; nulliparity, the first 3 months postpartum, low androgen or high estrogen status (in males); longer-duration breastfeeding
- Smoking status; >25 cigarettes/day for >20 years confers a 15-fold risk in subjects who carry disease associated human leukocyte antigen (HLA)-*DRB1* alleles
- Low alcohol intake
- Environmental antigens (the "exposome"); dietary factors; exposure to infectious (and noninfectious/microbiota) pathogens at mucosal surfaces such as the lung, periodontium, and gut; non-inherited maternal antigens (NIMA)
- Urban dwelling, relating to airborne pollutants

and a range of biochemical derangements of pathways integral to antigen responsiveness and immune regulation, these factors may combine to (1) increase susceptibility to foreign pathogens, (2) augment reactivity to self-tissue antigens (which may be modified post-translationally by the aging process), and (3) generate a repertoire of lymphocytes defective in terms of tumor surveillance. Thus, physiological immune senescence could be considered a risk factor for RA in the elderly, while premature senescence may contribute to early onset RA.

The female sex preponderance implies that hormonal and reproductive factors strongly influence risk. On the one hand, nulliparity is a risk factor for RA. Women entering the first 3 months of the postpartum period are also at increased risk. By contrast, oral contraceptive use, pregnancy, and hormonal replacement therapy have all been associated with reduced risk or less severe disease, whereas extended periods of breastfeeding appear to increase risk. An influence of hormonal factors is further suggested in studies of men where disease is associated with lower androgenic testosterone and dehydroepiandrosterone (DHEA) levels, and increased estradiol, compared with healthy control male subjects.

ETIOLOGY AND PATHOGENESIS

Environmental and Nongenetic Factors

Most, but not all studies have reported an association between RA and smoking. One of the largest studies, comprising over 370,000 women from Women's Health Cohort Study, reported a relative risk of 1.4 for women who smoked more than 25 cigarettes per day for more than 20 years, compared with nonsmokers.[4] The association appears to be more closely related to duration than to the amount of tobacco exposure, with smoking status being a risk factor for older age of RA onset; smoking also may influence severity because smokers are more likely to have seropositive, erosive disease with extraarticular manifestations. Further evidence of gene-environment interactions with respect to smoking has been documented in a population-based case–control study of Swedish RA patients.[5] In this study the relative risk of developing rheumatoid factor (RF) positive (RF^+) RA was calculated according to smoking status and HLA-DRB1 genotype. The relative risk of developing RA increased from 2.5 in nonsmokers with disease-associated HLA-DRB1 genes to 7.5 and 15.7 in smokers who carried one or two copies of the susceptibility alleles, respectively. Follow-up gene-environment interaction studies demonstrate robust associations between heavy smoking, HLA-DRB1 alleles encoding specific amino acids at positions 11 and 13, and the presence of antibodies to citrullinated protein antigens (ACPA). Strongest associations were with antibodies to citrullinated α and β chains of fibrinogen (epitopes Fibα580–600 and Fibβ36–52) and α-enolase (CEP-1); IgA ACPAs were especially prevalent in smokers.

Being female, a smoker, and carrying specific disease-associated genetic variants may be necessary but not sufficient to initiate chronic inflammatory arthritis. Other environmental triggers may be involved. Not least among these is exposure to foreign pathogens. This association has gained credibility because of the presumed link not only between infection and autoimmunity but also between immunodeficiency and autoimmune disease. Nonetheless, no single pathogen or group of pathogens has been defined. This could imply that aberrant host responses (either exaggerated innate inflammatory responses or failure to terminate such responses) may arise after a wide range of infectious insults. Indeed, bacterial products including superantigens, mycoplasma species, viruses (including herpes family, parvovirus, and retroviruses), and fungi have all been implicated, but data are insufficient to prove causation. Epstein-Barr virus (EBV) infection is common, and antibodies to EBV nuclear antigens have been reported in patients with RA. EBV is a polyclonal activator of B lymphocytes and EBV-specific T cells reactive to EBV gp110 have been identified in RA synovial joints, in keeping with the detection of EBV RNA in synovium. There exists a tantalizing link between infection with *Porphyromonas gingivalis*, which expresses its own enzymatic machinery for generating bacterial or host-derived citrullinated proteins, severe periodontitis (which shares risk factors for RA), and RA.[6]

A comparison of the microbial genomes from the small intestine and colon of mice housed in conventional versus germ-free facilities suggests that a single gut-residing species, in this case segmented filamentous bacteria, can profoundly influence the clinical expression of inflammatory arthritis in mouse models of autoimmune arthritis. In the K/B × N serum transfer model of arthritis, segmented filamentous bacteria enhanced generation of interleukin (IL)-17-expressing T cells in lamina propria. Together with rapid advances in sequencing technologies, these data have prompted a systematic analysis of the symbiont microbial communities in patients with RA in comparison to those derived from healthy control populations. Microbiota are attractive environmental risk factors because they are acquired at around the time of birth and are modified by diet as well as the host genome. Studies from several groups have demonstrated irrefutable evidence of dysbiosis, with distinct patterns of microbiota depending on the stage of disease and the population studied. For example, the first US study reported enrichment of *Prevotella* spp. and *Bacteroides* spp., gram-negative anaerobes in the gut of patients with early but not established RA. Metagenomic sequencing of oral and fecal microbiota from a cohort of Chinese RA patients revealed enrichment of *Lactobacillus*, particularly in severe disease, and depletion of *Haemophilus* spp.[7] Recent studies indicate that enrichment of *Prevotella* spp. is associated with an RA polygenic risk score in unaffected twins, as well as subjects at risk of RA,[8] suggesting a link between host genetic factors and dysbiosis prior to disease onset.

Immunogenetics

RA is a clinically heterogeneous disease, and so comprehensive identification of disease susceptibility genes has been challenging, in spite of heritability estimates in excess of 60%. With the exception of the MHC, where extensive gene polymorphism contributes about one-third of genetic susceptibility, and *PTPN22* (odds ratio 1.8), individual genetic variants confer low to moderate risk and have low penetrance (odds ratios of 1.1 to 1.5). Numerous genome-wide linkage scans of multiplex families with RA have established and confirmed an important contribution of the MHC (Chapter 5). This lends support to a wealth of epidemiological and genetic data describing associations between RA and specific HLA-*DRB1* alleles, in particular HLA-DR4 subtypes. Although this association was first described by Stastny in the 1970s, it was shown more than a decade later that susceptibility to RA across different ethnic populations correlated closely with the expression of a specific consensus amino acid sequence (referred to as the "shared epitope," hereafter SE) within the HLA-DR β chain α-helix (Fig. 53.1).[9] This sequence was subsequently shown by several groups of investigators to be encoded by HLA-*DRB1* alleles, including *HLA-DR4 (*04:01, *04:04, *04:05*, and **04:08)*, but also *HLA-DR1 (*01:01), DR6 (*14:02)*, and *DR10 (*10:01)* alleles, among others. *HLA-DR9 (*09)* is also associated, but not in Caucasians. According to fine-mapping data, RA risk is linked not only to amino acids encoded by the RA SE sequence (71 and 74 positions of the alpha helix of the DRβ chain) but also to amino acid positions 11 and 13, located in the peptide binding groove of the HLA-DR heterodimer.[10] Single amino acid variants have also been defined in HLA-B (position 9) and HLA-DPβ (position 9). Differences in odds ratios associated with these five amino acid variants between ACPA-positive and ACPA-negative disease provide further evidence that seropositive and seronegative disease are genetically distinct.

Specific genotypes co-segregate with distinct clinical features. For example, in population-based studies, different HLA-*DRB1* alleles influence the severity of disease, including radiographic progression (confirmed in meta-analyses), with *DRB1*0401* being found in patients with severe, seropositive, erosive RA (often with extraarticular features such as vasculitis and Felty syndrome in *04:01 homozygous or *04:01/*04:04 compound homozygote individuals). Valine at position 11 of HLA-*DRB1* confers the strongest association with radiological damage and highest all-cause mortality. Statistical modeling points to HLA genetic polymorphism being associated with young-onset RA. *DRB1**01:01 and *10:01 are observed at a higher frequency in patients with less severe, seronegative, nonerosive disease. Inheriting two copies of alleles expressing the consensus sequence increases disease penetrance, time of onset, and severity. Thus, distinct genotypes can manifest as distinct clinical entities.

On the basis of early observations, two principal models were proposed to account for the association between RA and the consensus DR β-chain sequence. Both were based on the assumption

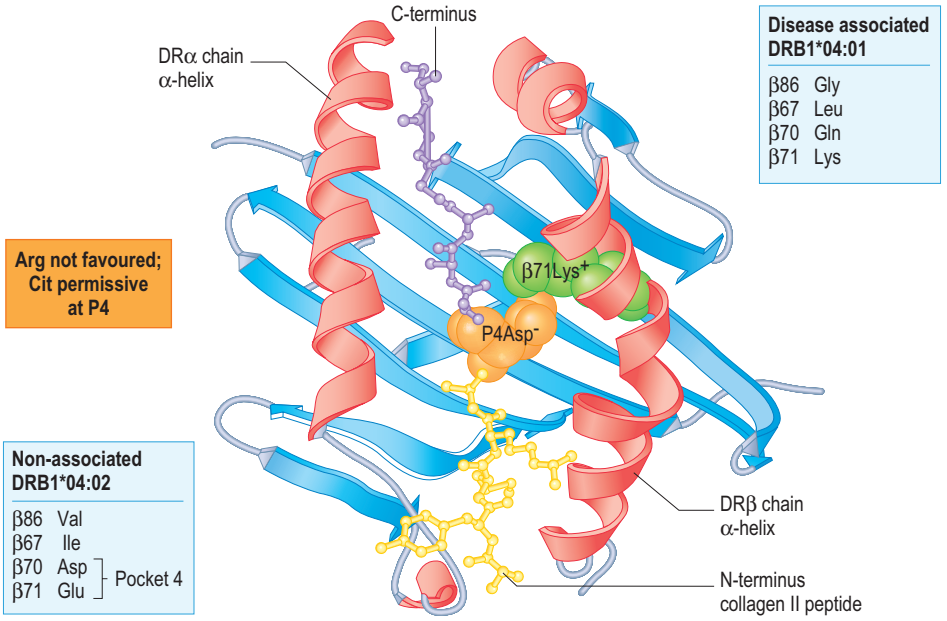


FIG. 53.1 Crystal Structure of a Collagen II Peptide/Human Leukocyte Antigen (HLA)-DR4 Complex. Ribbon model of an immunodominant collagen II peptide (1168 to 1180) complexed to HLA-DR4 (*DRA**01:01/*DRB1**04:01); a view of the major histocompatibility complex/peptide complex as seen from the T-cell surface. DRα and DRβ chain helices are shown in red, whereas the β-pleated sheet comprising the floor of the peptide-binding groove is shown in blue. Here, amino acids 11, 13, 71, and 74 form part of the fourth anchoring peptide-binding pocket, the same positions that confer highest risk of disease. Residues 67 to 74 of the DRβ chain, components of the third hypervariable region, derive the "shared epitope." The ball and stick model of the CII peptide is shown. Interacting residues of the peptide position 4 (Asp, orange) and DRβ chain residue (β71Lys, green), which make up part of pocket 4, are depicted as van der Waal's spheres. Differences in amino acid sequence between the closely related disease-associated DRB1*04:01 and non-associated *DRB1**04:02 gene products are illustrated. Note that although Arg would not be favored at position 4 in the peptide, modification of Arg → Cit by deimination would be permissive. These findings point to SE associating not with RA per se, but with an immunological phenotype, in this case autoantibodies to modified protein antigens (AMPA). Figure generated by R. Visse and A. Cope, based on crystal data derived by Wiley and colleagues (Courtesy Dessen A, Lawrence CM, Cupo S, Zaller DM, Wiley DC. X-ray crystal structure of HLA-DR4 (DRA*0101, DRB1*0401) complexed with a peptide from human collagen II. *Immunity*. 1997;7(4):473–481).

that the SE is the critical genetic element linked directly to disease. The first model proposed that the SE determines specific peptide binding and that "pathogenic" peptides bind preferentially to disease-associated HLA class II molecules (see Fig. 53.1). This model predicted that a gradient of affinities of disease-inducing peptide for MHC class II molecules might account for the differences in susceptibility and/or severity conferred by different HLA-DR molecules. Along the same lines, disease-associated alleles may preclude the binding of peptides required for the generation of naturally occurring Tregs specific for self-peptide antigens. The second model proposed that the SE influences T-cell receptor (TCR) recognition by binding and selecting autoreactive T cells during thymic maturation and expanding these populations in the peripheral compartment; again perturbations of a repertoire of Tregs could arise through opposing influences of the SE sequence. Based on crystal structures, both models hold up, with the shared epitope sequence conferring dual functionality by determining the repertoire of specific peptides for presentation and providing a determinant for TCR recognition.

Another important line of evidence pointing to specific functions of SE^+ alleles has arisen through analysis of autoantibodies in RA patients typed at the HLA-*DRB1* locus. These studies, replicated in European and US cohorts, demonstrated associations between SE frequencies and antibodies to cyclic citrullinated peptides (anti-CCPs), as distinct from rheumatoid factor (RF). When compared with healthy controls, the odds ratios for the association between one or two copies of the SE and anti-CCPs positivity was 4.4 and 11.8, respectively. The mechanism underlying this association is illustrated in Fig. 53.1, and discussed further below. The importance of citrulline as a molecular feature of human T-cell–specific autoantigenic determinants is further suggested through identification of autoreactive $CD4^+$ T lymphocytes by flow cytometry using HLA-citrullinated-peptide tetramer complexes.[11] These tools permit determination not only of the relative proportion of antigen-reactive effector T cells, but also regulatory T-cell subsets.

Meta-analyses and fine-mapping studies, including custom-designed single nucleotide polymorphism (SNP) arrays have identified more than 100 susceptibility loci with genome-wide significance, many of which have been validated in RA populations of diverse ancestry, and numerous small effect causal variants.[12] Among the strongest associations outside HLA is *PTPN22*, initially identified in candidate gene association studies. The *PTPN22* gene variant confers risk with odds ratios ranging from 1.5 to 1.9 in Caucasian and European populations, and it is unusual among most genetic variants in that it is a coding frame mutation (R620W). *PTPN22* encodes a hematopoietic cytoplasmic protein tyrosine phosphatase whose substrates include Src and Syk family kinases, CD3ε, TCRζ, and signaling intermediates such as Vav. These signaling intermediates operate downstream of antigen receptors, integrins, and pattern recognition receptors, and so the genetic variant that alters phosphatase function has become a highly plausible susceptibility allele from an immunobiological perspective. Indeed, studies in *Ptpn22* mutant mice have demonstrated that loss of PTPN22 function perturbs activation of dendritic cells and uptake of, and presentation to, T-cells of immune-complex-derived antigens. This, together with enhanced LFA-1 dependent cell adhesion at the immunological synapse,[13] and augmented antigen receptor signaling, places dysregulation of the initiation of adaptive immune responses firmly at the center of autoimmune susceptibility linked to *HLA* and *PTPN22* polymorphisms.

Scrutiny of other, non-MHC susceptibility loci point to genes whose products are involved in proximal signaling pathways that regulate T-cell activation, differentiation, and persistence. Besides *HLA* and *PADI4*, which influence the molecular determinants of T-cell "input signals," these include *CD28, CTLA-4*, and *CD2-CD58* (regulation of T-cell costimulation); *CD247, PTPN22, PRKCQ, TAGAP*, and *REL* (transducer modules of TCR signaling); *STAT4* and *TNFRSF14* (inducers of lineage-specific cytokine gene expression and persistence of memory T cells); and *REL, IL2-IL-21, IL2RA*, and *IL2RB* (regulators of IL-2 gene expression and IL-2R signaling). Notable overlap with these and other allelic variants has been reported in other autoimmune diseases, indicating that susceptibility to RA is linked to fundamental perturbations of immune tolerance. Variants mapping to cell surface receptors (*IL6R, CCR6, CD40, CD5, FCGR2A*) and intracellular signaling intermediates (*TNFAIP3, TYK2, TRAF1, TRAF6, RASGRP1, BLK*) are well represented, as are transcription factors linked to cell differentiation and effector function (*GATA3, IRF5, IRF8, IKZF3, RBPJ, RUNX1*). Perturbations in expression or function of these genes will influence the function of a broad range of immune cell subsets.

The identification of disease-associated genetic variants raises questions about precisely how they alter cellular function, and in which cell types. This is a challenge, not least because 90% of genome-wide association studies (GWAS) associations reside in non-coding sequences, and means that gene-targeting approaches to reveal functions in mammalian cells is not trivial. Nonetheless, identification of expressed quantitative trait loci (eQTL) has supported the view that a proportion of variants alter gene expression. For example, RA genetic studies have shown that SNPs are overrepresented in memory $CD4^+$ T-cell regulatory elements. Understanding cell-type-specific gene expression enrichment in GWAS risk loci remains a topic of considerable interest.

Synovial Pathology

RA targets diarthrodial joints, structures characterized by hyaline cartilage lining opposing articulating surfaces and a cavity of viscous synovial fluid lined by synovial membrane lacking a basement membrane but encased by a fibrous joint capsule. Normal synovial tissue comprises a lining layer, no more than a few cells in depth, of stromal fibroblast-like synoviocytes (FLSs—also known as type B synoviocytes) and sublining macrophages (type A synoviocytes). The normal synovium serves to line non-cartilaginous surfaces, and although blood vessels are sparse, it functions to provide essential nutrients to avascular structures including cartilage, tendons, and bursae.

Increased Vascularity and Cell Migration

The range of pathology observed in patients with RA perhaps most convincingly underscores the heterogeneity of the disease.[14] The earliest changes observed relate to increases in vascularity characterized by vascular congestion and thrombosis with obliteration of small vessels in association with perivascular inflammatory infiltrates. Hyperplasia of the synovial lining layer is another typical early finding. These changes are rather nonspecific and certainly not diagnostic.

A key checkpoint defining the switch from acute to chronic persistent inflammation is the sustained activation of microvascular endothelium, phenotypic changes in the high endothelial venules (reminiscent of tissue injury), and the concomitant upregulation of adhesion molecules such as intercellular

adhesion molecule (ICAM)-1 and vascular cell adhesion molecule (VCAM)-1 (Chapter 16). According to current thinking, the expression of chemoattractants derived from synovial stromal cells heralds the rolling, adhesion, and transmigration of mononuclear cells through endothelial barriers into the synovial membrane. It also contributes to the progressive synovial hypertrophy and hyperplasia, sometimes with villous-like projections more typical of chronic, established inflammation. Intravital imaging of synovial joints of mice injected with arthritogenic antibodies derived from the serum of *K/B* × *N* mice indicates that enhanced vasopermeability at sites destined to become arthritic is a crucial early event, at least in antibody-induced disease. This process is dependent on mast cells and neutrophils and on the release of the vasoactive amines histamine and serotonin, which contribute to heightened vascular permeability. Neovascularization promotes further influx of inflammatory cells. To what extent this is driven by the hypoxic environment is not entirely clear, but expression of angiogenic growth factors such as vascular endothelial growth factor (VEGF), angiopoietin, Tie-2, and hypoxia inducible factor (HIF), as well as lymphangiogenic factors VEGF-C and VEGF-R3, are increased. The abundance of lymphatic vessels in inflamed synovium, suggested by expression of podoplanin and CD31 *in situ*, suggests that active lymphangiogenesis exists that may promote efflux of cells and fluid from the synovium.[14]

Organization of Lymphoid Tertiary Microstructures

Tissue microstructure both dictates and facilitates immune responses in secondary lymphoid organs and mucosa-associated lymphoreticular tissues (MALTs). These structures have evolved to coordinate responses to pathogens and to direct lymphocyte recirculation, and although their role in immune homeostasis is established, quite how they contribute to pathological states is less well understood. Thus, the inflamed, noncapsulated synovium appears to be uniquely suited to supporting distinct patterns of cellular infiltrates, including inducible lymphoid structures that promote pathways of cell activation, differentiation, and survival.[14] These include diffuse, rather disorganized lymphocytic infiltrates that comprise the most common form of synovitis, occurring in ~30% in prospective cohort studies; up to 70% has been described in late-stage disease (at arthroscopy, joint replacement surgery). In 40% to 50% of patients more organized follicular structures may exist (Fig. 53.2). Based on immunohistochemical analysis, approximately 25% of these follicular structures include organized germinal centers in which there are zones of proliferating B cells with affinity maturation, in addition to a distinct T-cell zone. In aggregates lacking germinal centers, follicular dendritic cells (DCs) are absent.

A fourth histological pattern, characterized by granulomatous reactions, has been described in a much smaller subset of patients. The cellular and molecular determinants of these structures include the homeostatic lymphoid chemokines CXCL13 and the CCR7 ligands CCL21 and CCL19, VCAM-1$^+$ICAM-1$^+$LTβR$^+$ mesenchymal-organizer cells, and hematopoietic-derived CD3$^-$CD4$^+$IL-7R$^+$RANK$^+$ lymphoid-inducer cells. Lymphoid-tissue-inducer cells produce LT-β, which is required for high endothelial venule differentiation, amplification of chemokine expression, and development of the stromal architecture. Documentation of class-switch recombination and somatic mutation of immunoglobulin genes *in situ* is further evidence of active adaptive immunity within synovial lymphoid follicles and suggests that supporting *in situ* production of pathogenic antibodies is of considerable pathobiological importance.

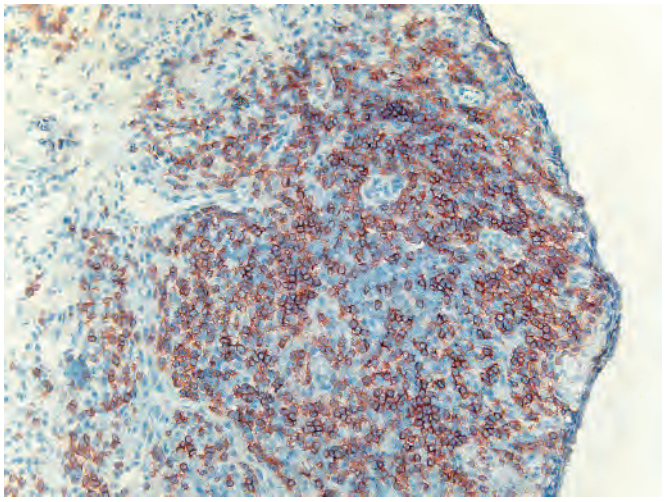


FIG. 53.2 Lymphoid Follicular Structures in Inflamed Rheumatoid Arthritis Synovial Tissue. A characteristic hematoxylin and eosin–stained tissue section from a patient with active RA showing a large follicular-like structure (original magnification ×100). This section is also stained with monoclonal antibodies to CD3ε, followed by a three-step immunoperoxidase staining protocol (CD3$^+$ T cells stained dark red). (Courtesy Tak PP, Taylor PC, Breedveld FC, Smeets TJ, Daha MR, Kluin PM, Meinders AE, Maini RN. Decrease in cellularity and expression of adhesion molecules by anti-tumor necrosis factor alpha monoclonal antibody treatment in patients with rheumatoid arthritis. *Arthritis Rheum.* 1996;39(7):1077–1081, with permission from J. Wiley and Sons, Inc.)

Functional correlates of these dynamic structures continue to emerge. Using conventional histology, immunohistochemistry, and RNA-seq on synovial tissue from early RA patients it has been possible to identify three broad groups (designated pathotypes) based on cell-specific gene modules. These bear similarities to those outlined above, and are described as fibroblastic pauci-immune pathotype, macrophage-rich diffuse-myeloid pathotype, and a lympho-myeloid pathotype characterized by infiltration of lymphocytes and myeloid cells.[15] Importantly, these phenotypes map to disease severity and clinical outcomes since pro-myeloid inflammatory synovial gene signatures correlate with clinical response to initial drug therapy, whereas plasma cell genes identified a poor prognosis subgroup associated with ultrasonographic synovial thickening and increased power Doppler scores, and progressive structural damage. In addition, a more inflammatory synovial phenotype at baseline correlated with a greater fall in DAS28-CRP after 6 months of disease-modifying anti-rheumatic drugs (DMARDs) treatment, while a robust type I interferon (IFN) response in peripheral blood was associated with synovial B-cell infiltration.

Immunobiology of Rheumatoid Arthritis

Initiation of the Immune Response

Synovial fibroblasts are exquisitely sensitive to inflammatory cytokines such as IL-1, TNF, and IL-6. Fibroblast-like synoviocytes (FLSs) also express a range of toll-like receptors (TLRs) that can respond to exogenous, pathogen-associated molecular

patterns (PAMPs) and a growing range of self-tissue proteins, some of which could be considered damage-associated molecular patterns (DAMPS). Endogenous ligands especially relevant to inflammatory arthritis include heat shock proteins, fibrinogen fragments, antibody-DNA complexes, high-mobility group box (HMGB)-1, and hyaluronan oligosaccharides. Stimulation of FLS through these pathways induces cytokines such as IL-6, matrix metalloproteinases, adhesion molecules, and an array of chemokines including granulocyte chemotactic protein (GCP)-2, RANTES, monocyte chemoattractant protein (MCP)-2, IL-8, growth-related oncogene-2, and, to a lesser extent, macrophage-inflammatory protein 1α, MCP-1, EXODUS, and CXCL-16. This creates an inflammatory niche for recruiting and sustaining leucocytes in the synovial joint.

Cell-surface phenotyping of synovial FLS by mass and flow cytometry has uncovered seven subsets based on expression of podoplanin, cadherin 11, and Thy-1, segregated on the basis of the hematopoietic marker CD34.[16] In-depth transcriptomic profiling identified three broader subsets, discrete from OA synovial FLS. While $CD34^+$ fibroblasts are observed in both superficial lining and deeper sublining areas, $CD34^-Thy1^+$ fibroblasts in RA form a discrete perivascular zone surrounding capillary structures in the deep sublining layer of the synovium, especially near accumulations of lymphocytes. $CD34^-Thy1^-$ fibroblasts were mostly observed in the lining area. Synovial tissue from active, clinically swollen joints had fewer $CD34^-Thy1^-$, more $CD34^-Thy1^+$, and more $CD34^+$ fibroblasts, and the proportion of $CD34^-Thy1^+$ fibroblasts positively correlated with the proportion of infiltrated leukocytes determined by flow cytometry. These expanded, putative pathogenic populations are more proliferative and express genes associated with a migratory response, and expression of inflammatory cytokine genes such as IL6, CXCL12, and CCL2 that support matrix invasion, immune cell recruitment, and osteoclastogenesis. Adoptive transfer experiments in mice have demonstrated that two anatomically discrete and functionally distinct populations of fibroblast-activation protein alpha (FAPα) expressing FLS segregate based on Thy1 expression; $Thy1^+$ FLS are effector cells largely confined to the sub-lining layer, while the lining layer $Thy1^-$ subset express *Ccl9* and *Tnfsf11*, both potent inducers of osteoclast activity, as well as *Mmp3*, *Mmp9*, and *Mmp13*, and promote cartilage destruction.[17]

Dendritic cells (DCs) are thought to be the most important antigen-presenting cells in RA. Indeed, the proinflammatory environment favors DC maturation in regional lymph nodes as well as inflamed tissue. Thus in peripheral blood, DC precursors express either an immature $CD33^{dim}CD14^{dim}CD16^-$ phenotype or a more mature MHC class II^{bright} $CD11c^+CD33^{bright}CD14^{dim}$ surface phenotype typical of conventional myeloid DC (mDC); neither population expresses costimulatory molecules. In contrast, synovial fluid and tissue DC subsets resemble mature peripheral blood cells; in addition, a subset expresses high levels of CD86 that can support allogeneic mixed leukocyte reactions. More recent data indicate that they may differentiate further *in situ* as suggested by nuclear translocation of RelB in DC localized within perivascular infiltrates, consistent with prior cytokine receptor or Toll-like receptor (TLR) engagement *in vivo*. Perivascular RA synovium also contains populations of MHC class $II^+CD11c^-CD123^+$ plasmacytoid DC (pDC); in contrast to the conventional myeloid DC subset, these are $RelB^-$ and comprise ~30% of all synovial DC. A subset of pDC express BDCA2, capable of producing IFN-α *in situ*. Unlike their peripheral blood counterparts, synovial pDC efficiently activate allogeneic T cells to proliferate as well as to produce IFN-γ, TNF, and IL-10.

Although the common myeloid precursor cell is the precursor for all myeloid cells, including DC and tissue macrophages, the precise role of monocytes—namely $CD14^+CD16^-$, $CD14^+CD16^+$, and the more recently described $CD14^{dim}CD16^+$ subset—in synovial inflammation is uncertain. They are good candidates as persistence factors through their capacity to activate and polarize T-cell subsets, to respond to the environment through TLR expression, and to produce a wide range of inflammatory mediators, including IL-1, TNF, IL-6, IL-8, CCL2, NO, and type I IFN.

It turns out that not all myeloid cells are inflammatory. A comprehensive spatiotemporal analysis of the composition, origin, and differentiation of subsets of macrophages within healthy and inflamed joints of mice using fate-mapping approaches, fluorescence microscopy, and single-cell RNA sequencing has identified a population of CX_3CR1^+ tissue-resident, lining synovial macrophages with barrier functions typical of epithelial cells.[18] This population was derived from a proliferative subset of CX_3CR1^- cells in the deeper layers of the synovium, which acquire transcription factors for terminal differentiation into the CX_3CR1^+ lining subset, and genes with immune regulatory functions for clearance of apoptotic debris, and genes encoding tight-junction proteins. Similar populations have been defined in human synovium expression genes associated with gate keeping functions. In the serum transfer model of arthritis, inflammation-associated barrier breakdown occurred after the deposition of autoantibody-containing immune complexes.

The initial wave of inflammation has two major consequences. First, inflammatory cytokines will promote the activation of vascular endothelium, changes that occur very early in disease (see above and Fig. 53.3).[19] Under the influence of locally generated cytokines and chemokines, synovial postcapillary venules undergo morphological changes to an extent that they resemble high endothelial venules similar to those observed in secondary lymphoid organs. The second major consequence is the migration of inflammatory leukocytes, including polymorphonuclear leukocytes and immature or undifferentiated monocytes, orchestrated by chemokines produced by resident stromal as well as infiltrating cells (see Fig. 53.3). CXC, CC, C, and CX_3C chemokines all play a role, exerting chemotactic activity toward neutrophils, lymphocytes, and monocytes but also influencing the topology of inflammatory infiltrates. Besides the homeostatic chemokines described above, the key players include IL-8/CXCL8, RANTES/CCL5, MIP-1α/CCL3, SDF-1/CXCL12, IP-10/CXCL10, and MCP-1/CCL2. Upregulation on endothelium of cell-surface adhesion molecules, including ICAM-1, VCAM-1, and E-selectin, permits the rolling and adhesion of leukocytes as they migrate. In synovial joints, resident stromal cells and infiltrating macrophages are a dominant source of such factors. Crucially, the expression of cognate chemokine receptors such as CCR4, CCR5, CCR6, CXCR3, and CX_3CR1 on inflammatory cell subsets contributes selectivity of cellular recruitment.

Autoantigens in Rheumatoid Arthritis

Although current models of adaptive immune responses would suggest that DC carry antigens derived from damaged or dying synovial tissue, the molecular nature of disease-associated antigens has, until recently, remained an enigma. Many RA-associated

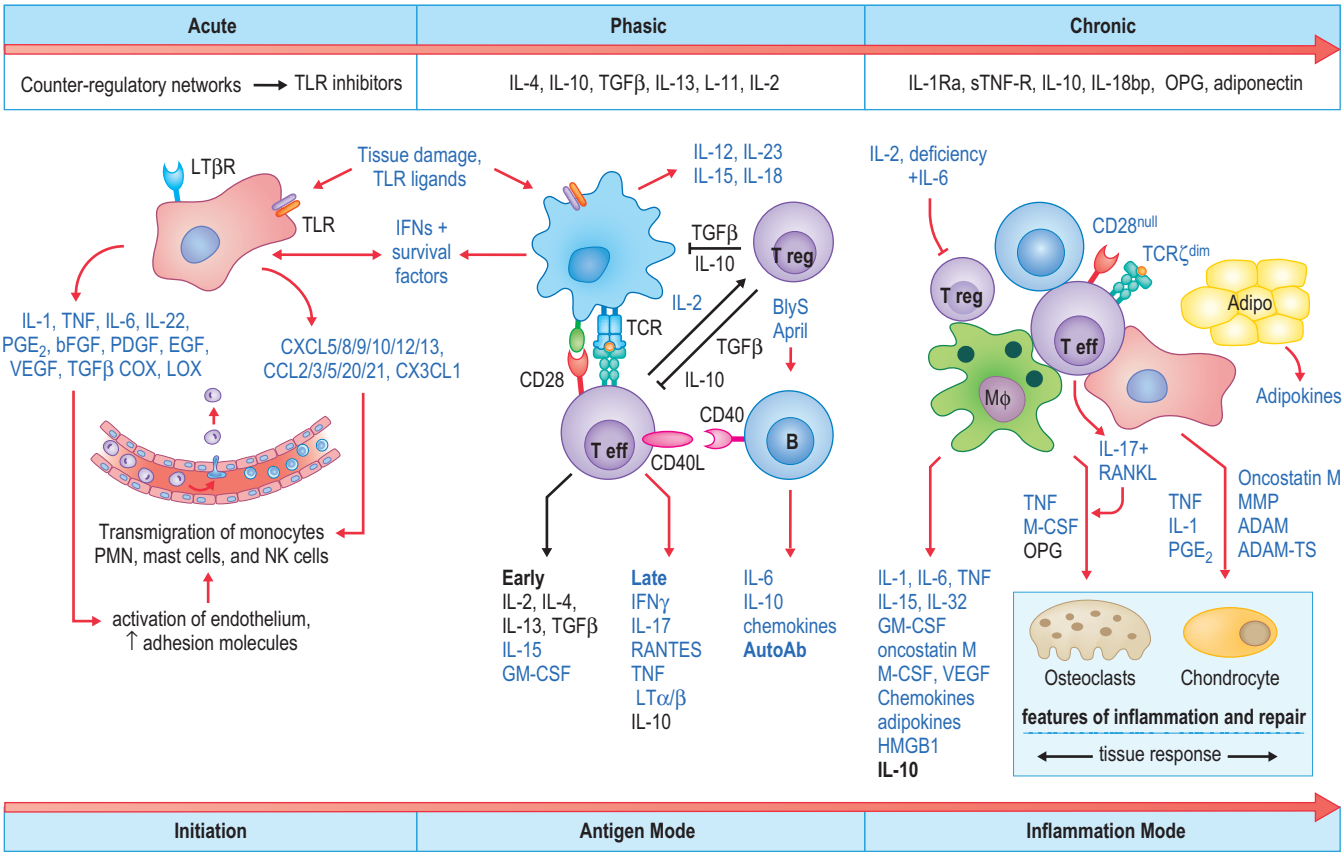


FIG. 53.3 Cytokine Networks in Rheumatoid Arthritis. The pathogenesis of rheumatoid arthritis can be thought of as a series of complex and closely related pathways temporally and spatially regulated. These include (1) an acute insult that may trigger the disease, characterized by stimulation of FLSs by inflammatory stimuli and the generation of cytokines and chemokines that promotes the migration and infiltration by cells of the innate immune system; (2) repeated episodes of antigen-specific adaptive immune responses (in lymph node, bone marrow, and *in situ*). Failure to resolve adaptive immunity is a key checkpoint that may lead to (3) a cytokine-driven chronic inflammatory phase when multiple cellular and molecular components sustain the response. Through multiple pathways acting on many cell types, this process leads to tissue injury. Proinflammatory pathways are shown in blue (text) and red (arrows), whereas anti-inflammatory, counterregulatory pathways are shown in black (text and arrows). *Adipo*, Adipocyte; *AutoAb*, autoantibodies; *B*, B cell; *DC*, dendritic cell; *FLSs*, fibroblast-like synoviocytes; *MΦ*, macrophage; *TCR*, T-cell antigen receptor; T_{eff}, effector T-helper cell; *Treg*, regulatory T cell.

autoantigens have been described (Table 53.1 for examples), and for some there exist clear correlates linked to *in vivo* arthritis models. The best described are collagen II, proteoglycans, HCgp-39, glucose-6-phosphate isomerase, α-enolase, vimentin, and citrullinated fibrinogen. However, when used as recombinant native antigen, few have been found to elicit reproducible and/or robust PB or SF T- or B-cell responses in a significant proportion of patients, when compared to healthy donors. There are several plausible explanations for this. Perhaps the most obvious is that the autoantigens used to test lymphocyte reactivity in vitro do not carry the posttranslational modifications (i.e., the neoepitopes) recognized by autoantibody or antigen receptor.

The Discovery of Citrulline as a Key Target for Autoimmunity in Rheumatoid Arthritis

In 1998, van Venrooij and colleagues first reported that patients with RA carried serum autoantibodies that recognized deiminated peptides of fillagrin.[20] Using new-generation anti-cyclic citrullinated peptide (anti-CCP) based assays, the presence of these antibodies, now collectively termed ACPAs, is now confirmed to be both sensitive (up to 80%) and highly specific (>95%) for the diagnosis of RA. Indeed, serum anti-CCPs levels are stable in established disease, can be detected many years before disease onset, and have been shown to be predictors of radiographic progression. Changes in isotype usage and spreading of antigenic specificities suggest that ACPA responses mature before disease onset. While citrullination is not specific for RA, it may be inflammation specific, having been documented in inflamed synovium from patients with reactive arthritis and psoriatic arthritis as well as RA, but not osteoarthritis (OA). What appears specific for RA is the immune response to citrulline (Fig. 53.4). Initially, linkage analysis across chromosome 6 documented a peak with logarithm of odds (LOD) scores in excess of 10 for $ACPA^+$ patients, but not for those who do not carry these antibodies. This relationship was independent of RF status because the SE allele frequencies in $ACPAs^+$ patients were twice those of $ACPAs^-$ patients, even for those patients who are RF^+.

An emerging model proposes that SE^+ DRB1 alleles are not involved in the initial breach of tolerance, but play a role in boosting the immune response to citrullinated proteins, hence

TABLE 53.1 Autoantigens in Rheumatoid Arthritis

Established	T or B Cell[a]	Molecular Specificity	Assay[b]
Immunoglobulin G	B	Human Fc IgG	Rheumatoid factor
Cyclic peptides	T and B	Citrullinated peptides	Anti-CCPs
Various peptides	B	Carbamylated peptides	Research[b]
Various peptides	B	Acetylated peptides	Research[b]
Fibrinogen peptides	T and B	Citrullinated α- and β-chain epitopes	Research[b]
Enolase peptides	T and B	Citrullinated CEP-1 peptide	Research[b]
Vimentin peptides	T and B	Citrullinated vimentin peptides	MCV assay
Collagen II	T and B	Multiple epitopes, including glycosylated epitopes	Research[b]
HnRNPA2	B	Multiple epitopes	Research[b]
Aggrecan	T and B	Citrullinated epitopes	Research[b]
HCgp-39	T	Multiple epitopes	Research[b]
Glucose-6-phosphate isomerase	B	Multiple epitopes	Research[b]

[a]Denotes autoantigens recognized by T or B cells, or both.
[b]Assay is either not commercially available or not in routine clinical use. Details of assays may be found in primary research communications.
CCPs, Cyclic citrullinated peptides; *MCV*, modified citrullinated vimentin.

the strong association between SE$^+$ DRB1 alleles and ACPA$^+$ arthritis. This concept is further reinforced by crystallographic studies suggesting that the conversion of positively charged arginine at key residues in antigenic peptides from candidate autoantigens such as fibrinogen to neutral citrulline is permissive for peptide binding and recognition by autoreactive T cells *in vivo*.[11]

Citrullination is widespread in multiple tissues in response to appropriate provocations. Although the molecular basis for these triggers is poorly understood, recent data point to a link between smoking—an environmental exposure known to be linked with RA—citrullination, and individuals carrying SE.[21] Thus, cells derived from bronchoalveolar lavage from smokers, but not from nonsmokers, express citrulline. The association between the development of RA and smoking has now been linked to ACPAs$^+$ patients, whose relative risk increases 20-fold if they smoke and carry two copies of the SE$^+$ DRB1 alleles; in comparison, the relationship in ACPAs$^-$ patients appears to be much weaker or nonexistent. A second example relates to the fact that *Porphyromonas gingivalis*, a pathogen associated with severe periodontitis, expresses its own deiminating enzyme and has the capacity to modify host proteins such as fibrinogen and α-enolase in inflamed gingival tissue. Molecular mimicry is suggested by the finding that anti-α-enolase autoantibodies from patients with RA cross-react with *P. gingivalis*–derived α-enolase. The links between autoantibodies to the immunodominant α-enolase epitope CEP1, smoking, SE$^+$DRB1, and disease-associated *PTPN22* alleles are some of the strongest defined to date, and show direct links between environmental exposure and disease-specific immune responses governed by immune-response genes.

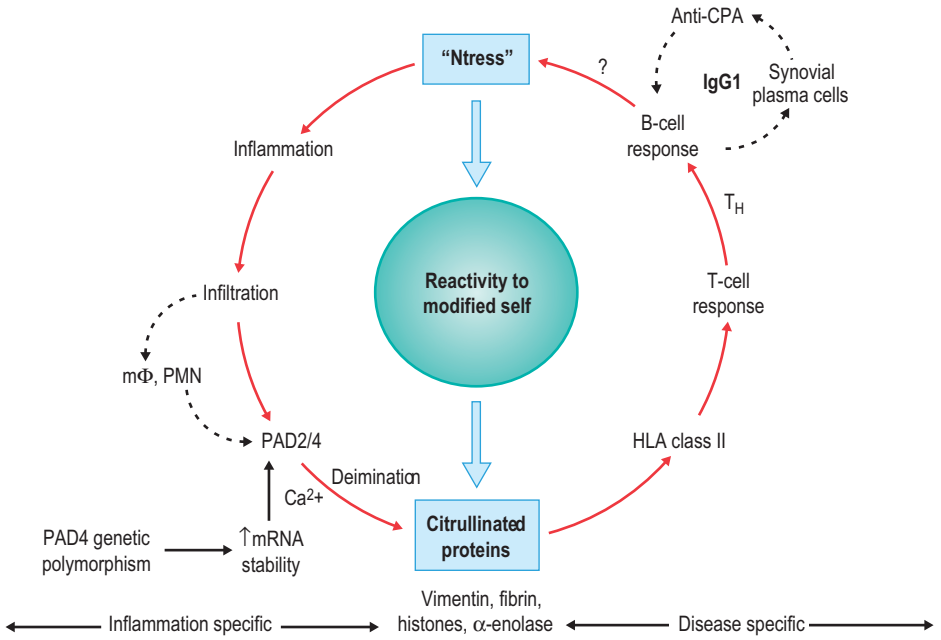


FIG. 53.4 The Generation of Autoantibodies to Citrullinated Protein Antigens. The stressed and inflamed synovium is characterized by an influx of inflammatory cells, including macrophages and neutrophils that express peptidyl-arginine deiminases (PAD). In the presence of sufficient Ca^{2+}, PADs deiminate target proteins including, among others, vimentin, fibrinogen, aggrecan, type II collagen, histones, and α-enolase. This reaction is inflammation, but not disease specific. The combination of environmental stimuli (including inflammation and exposure to tobacco smoke) and the inheritance of specific HLA-DRB1 alleles favor T- and B-cell immune responses to the host's derivatized neoepitope peptide antigens. Neoepitopes may also be generated by changes in peptide cleavage during antigen processing as a consequence of the Arg → Cit modification. *Anti-CPA*, Autoantibodies to citrullinated protein antigens; T_H, T-helper effector cell.

Our growing appreciation of the importance of post-translational modifications (PTMs) in general, and citrullination in particular, has driven a search for alternative PTMs that might contribute to the triggering or perpetuation of autoimmune responses to altered self.[22] For example, autoantibodies recognizing peptides carrying amino acids modified by carbamylation (to homocitrulline) or acetylation (to acetyllysine) are readily detectable, collectively termed anti-modified protein antibodies (AMPA). Mapping antibody reactivities against multiple peptide specificities indicates that citrulline-specific B cells are highly cross-reactive—a feature that likely underpins the basis of breech of immune tolerance prior to the onset of RA, and supported by findings from mouse immunization experiments revealing that different AMPA responses emerge from exposure to a single type of modified protein.

Lymphocyte Biology

Flow cytometric analysis of dissociated synovial mononuclear cell cultures indicates that infiltrating T lymphocytes make up approximately 10% to 35% of cells in inflamed tissue. Synovial T cells express phenotypic markers of antigen experienced, terminally differentiated T cells with enhanced migratory capacity. Thus, synovial T cells typically carry cell surface markers such as HLA-DR$^+$, LFA-1$^+$, VLA-1$^+$, CXCR3, CD69$^+$, CD45RO$^+$, CD45RA$^-$, CD45RB^{dim}, CD29^{bright}, CD27$^-$, and CD25$^-$, and low expression of TCRζ. While synovial fluid T cells are also FasL$^+$, Bcl2$^-$, Bax^{bright}, favoring a proapoptotic state, it is thought that environmental cues transduced through common γ chain receptor signaling cytokines such as IL-2, IL-7, and IL-15, as well as type I interferons, prevent apoptosis of T cells *in situ*. Synovial tissue–derived lymphocytes may be different. Consistent with their state of terminal differentiation, a subset of synovial T cells are CD28$^-$, while at the same time expressing a range of natural killer (NK) cell-surface receptors that are thought to contribute to effector function independently of cognate antigen.

Recent analysis of citrulline-reactive B cells from RA patients indicate that they display markers of class-switched memory B cells and plasmablasts (CD20$^+$CD27$^+$IgD$^-$). Analysis of the variable (V) regions of the immunoglobulin heavy and light chains confirm that antigen-specific activation and differentiation of B cells into plasma cells takes place in draining lymph nodes as well as in the chronically inflamed synovial tissue of patients with RA. Likewise, analysis of the T-cell repertoire indicates that the synovium provides a niche for supporting expansion of specific T cells, whose clonality is shared between different joints from the same patient, but not substantially with clones from paired blood samples.

For over a decade it has been known that one of the dominant cytokines expressed in synovial T cells from patients with established disease is IFN-γ. Somewhat surprisingly, a significant proportion of IFN-γ$^+$ cells also express IL-10. Recent data support a model in which there exists, during differentiation, a transition from IFN-γ$^+$ Th1 T cells through an IFN-γ$^+$IL-10$^+$ double-positive stage to a single-positive IL-10$^+$ stage. This last phase could represent part of the normal life cycle of an effector T cell, where IL-10 expression promotes the resolution of the adaptive immune response, attenuating the function of DCs. Interestingly data indicate that this Th1 life cycle involving a switch from IFN-γ to IL-10 production may be defective in individuals at risk of developing RA.

Recent, systematic approaches for studying synovial T cells at a single cell level has uncovered a markedly expanded population of PD-1^{hi}CXCR5$^-$CD4$^+$ T cells in synovium of patients with RA.[23] These are unusual in light of high expression of PD-1 and lack of CXCR5, and while they express activation markers such as MHC class II, they are not exhausted. Defined as peripheral helper T cells (T_{PH}), and distinct from T_{FH} cells, synovial PD-1^{hi}MHCII^{hi}CXCR5$^-$ cells support B-cell help via expression of IL-21, CXCL13, ICOS, MAF, and BLIMP1, and have been shown to induce plasma cell differentiation at least *in vitro*.

Finally, PTM of immunoglobulins, most notably glycosylation of the Fc domain, has been a recognized feature of the evolving immune response for decades, with distinct moieties being linked to specific IgG effector functions. Besides glycosylation at position 297 in the IgG Fc tail of APCAs, where animal studies suggest that disease-associated glycosylation patterns are regulated by IL-23 and Th17 cells, a recent addition to the repertoire of glycosylation of ACPAs has been the identification of highly sialylated N-linked glycans in the antigen variable domain of up to 90% IgG ACPA, arising as a consequence of T-cell dependent, variable region somatic hypermutation.[22] This is rather uncommon in other IgG molecules. Analysis of IgG ACPA in first-degree relatives of RA patients suggests that extensive glycosylation precedes the onset of disease, and thus represents a signature of high risk. While Fc tail agalactosyl "pro-inflammatory modifications" likely alters immunity through effects on FcR binding and effector responses, current thinking suggests that ACPA IgG variable domain N-glycans could influence antigen binding avidity, B-cell receptor signaling and tolerance, or binding to glycan receptors such as lectins.[22] Further work is needed to evaluate each of these possibilities.

Immune Regulation

Investigation of regulatory cell subsets and their anti-inflammatory properties has perhaps more firmly established the concept that failure of the host's intrinsic mechanisms of immune regulation can underpin autoimmune diseases. Experiments in gene-deficient mice (*e.g.*, Foxp3, IL-2, IL-2R, IL-2R signaling, STAT5, IL-10, TGF-β) lend support to this concept. In RA the data remain less clear. For example, there is *in vitro* evidence for a relative deficiency of constitutive IL-10 expression in synovial cell cultures, and yet clinical trials of IL-10 have been disappointing. These results may reflect the complex role of these cytokines in disease pathogenesis. The identification of defective numbers and/or function of CD4$^+$CD25^{bright} Tregs has been suggested by several investigators, but reports are conflicting.[24] Some studies have shown clear reductions in numbers of peripheral blood Tregs in patients with RA, whereas others have shown no difference. In synovial joints, the data are more consistent, with many reports showing substantial increases in Treg numbers in synovial tissue and fluid compared with paired PB. However, some studies have reported normal function at a cellular level, whereas others have shown depressed regulatory function. One possible mechanism is that synovial effector T cells are refractory to regulatory pathways.

The advent of immune checkpoint inhibition (CPI) for the treatment of a wide range of cancers has, serendipitously, provided incontrovertible *in vivo* evidence that immune checkpoints CTLA-4 and PD-1 contribute to immune homeostasis in individuals at risk of inflammatory arthritis.[25] Immune-related adverse events (irAEs) are now a well-recognized complication of CPI with anti-PD-1/PDL-1 and anti-CTLA4, targeting any organ or tissue in the body. These inflammatory syndromes, which phenocopy idiopathic autoimmune diseases, can arise within weeks or months of treatment initiation, and are thought to represent an abrupt breach of immune

tolerance. Both seronegative and seropositive subsets of inflammatory polyarthritis, indistinguishable from RA, have been reported by many groups; some cases are transient and self-remitting, while others persist, requiring disease-modifying therapy. What remains unclear is whether this RA-like syndrome arises only in susceptible individuals or whether the inflammatory process arises *de novo*, regardless of genetic or environmental factors.

Impact of the Inflammatory Response on Cartilage and Bone

For many years, it was considered that the terminal effector phase of chronic inflammation that led to cartilage destruction and bone resorption was driven almost exclusively by inflammatory cytokines and proteinases. IL-1, MMPs (MMP1, 3, 8, 13), and aggrecanases (ADAMTS 4 and 5) were, and remain, major drivers. Attempts to establish more directly a link between adaptive immunity and destruction of target tissue failed, not least because of the lack of a direct physical link between lymphocytes, chondrocytes, and bone. A breakthrough came in the late 1990s with the identification of the TNF/TNFR family member receptor for activation of nuclear factor (NF)-κB ligand (RANKL)/TRANCE/ODF and its counter-receptor RANK, and the dissection of the molecular and cellular components required for osteoclast differentiation from monocyte precursors.[26] According to contemporary paradigms, RANKL is necessary and sufficient for osteoclast differentiation. TNF, M-CSF, IL-1, and IL-17 contribute, and RANKL-independent pathways may also play a role. RANKL is expressed on synovial fibroblasts and osteoblasts but also on activated T cells, its counter-receptor being expressed on myeloid lineage cells including monocytes, osteoclast precursors, and DCs. Its expression is regulated by inflammatory mediators including TNF and PGE_2. RANKL is shed, probably through the action of several membrane-associated proteases including MT1-MMP (MMP14). Gene targeting of RANKL or RANK in mice leads to inhibition of osteoclastogenesis and a profound osteoporotic bone phenotype. Deletion of osteoprotegerin (OPG), the naturally occurring decoy soluble receptor for RANK, leads to unbridled osteoclast differentiation and bone resorption and substantially reduced bone mass. In RA, several studies have demonstrated perturbations of serum RANKL/OPG ratios, and recent clinical experience with denosumab, a fully humanized monoclonal antibody that binds to RANKL, demonstrates increased bone mineral density and reduced bone turnover in patients with RA.

Distinct from enhanced bone resorption, bone formation is impaired in RA, although until recently the pathways involved were rather obscure. The inflammatory process itself, along with its effects on osteoblast maturation and function, has been directly implicated because bone formation at surfaces adjacent to bone marrow, as opposed to inflamed synovium, are relatively well preserved. Recent data suggest that ACPAs can also directly promote osteoclastogenesis, providing a mechanistic link between autoantibodies and joint destruction that also depends on IgG glycosylation status. Evidence now points to the canonical Wingless (Wnt) signaling pathway as being an essential control point in osteoblast function, based on the observation that in animal models of RA antibodies to Dickkopf homologue 1 (DKK1), a secreted antagonist of Wnt blocking, signals at the level of its cognate receptor Frizzled, promote bone formation and inhibit bone resorption indirectly by increasing production of OPG. This indicates that Wnt signals negatively regulate osteoclastogenesis.

The interplay between bone formation and resorption is fundamental to bone homeostasis, especially in the context of chronic inflammatory disease. Recent evidence suggests that the molecular basis for this so-called coupling lies with the osteoprotective factor semaphorin 3 A. This factor regulates bone mass in both osteoblasts (through effects on the canonical Wnt/β-catenin signaling pathway) and osteoclasts (by inhibiting RANKL-dependent osteoclast differentiation). Together these data support the view that bone and joint integrity are regulated by a delicate balance of catabolic and anabolic immune and inflammatory mediators influencing the maturation and function of osteoblasts (Wnt:DKK1) and osteoclasts (RANKL:OPG).

CLINICAL FEATURES

Disease Onset

RA is a heterogeneous disease that does not conform to a single clinical entity. Whereas 10% of patients may have an acute severe onset and 20% a more subacute onset, the onset of signs and symptoms may be insidious in up to 70% of patients. A more episodic or palindromic onset has also been described. A common presentation, more likely during the winter months, will be that of a female in her fifth to sixth decade of life who complains of diffuse symmetrical joint pain, swelling, and stiffness of peripheral joints. Patients may complain that they can no longer make a fist, especially in the early morning. The targeting of afflicted synovial joints may be symmetrical in most, but not all, cases, typically affecting small joints of the hands and feet as well as wrists. Less frequent are those presenting with slow-onset monoarticular disease. Patients not fulfilling the diagnostic classification criteria for RA may be ascribed the more appropriate diagnostic label of undifferentiated arthritis, because in a proportion of cases signs and symptoms may resolve spontaneously. Systemic disease is much less common in current clinical practice, in part through application of earlier, more intensive treatment regimens (see below). Nonetheless, the systemic nature of the disease can be manifested through a wide array of systemic, extraarticular clinical features that may occur in patients with disease at the more severe end of the spectrum. A spectrum of disease severity may also be evident from hand radiographs; examples are shown in Fig. 53.5.

Diagnosis

Classification Criteria

The American College of Rheumatology criteria for the classification of RA are a set of clinical and laboratory parameters, established largely for epidemiological purposes, that serve as a guide for the diagnosis of RA. They are relatively straightforward and easy to apply, especially to patients with established disease. However, failure of a patient with early signs and symptoms of an inflammatory arthropathy to fulfill them does not mean that the individual does not have RA. The 1987 criteria were simplified further by removing the "probable," "definite," and "classic" subclassifications. These criteria returned a sensitivity for RA of 91% to 94% and a specificity of 89% in the clinical setting. New criteria, established by a steering group comprising members of the ACR and EULAR and published in 2010, are based on a weighted score around four domains, including distribution and type of joint involvement (tender as well as swollen joints are included, scoring 0 to 5 points), serology (0 to 3) that includes RF or ACPAs and is weighted according to antibody

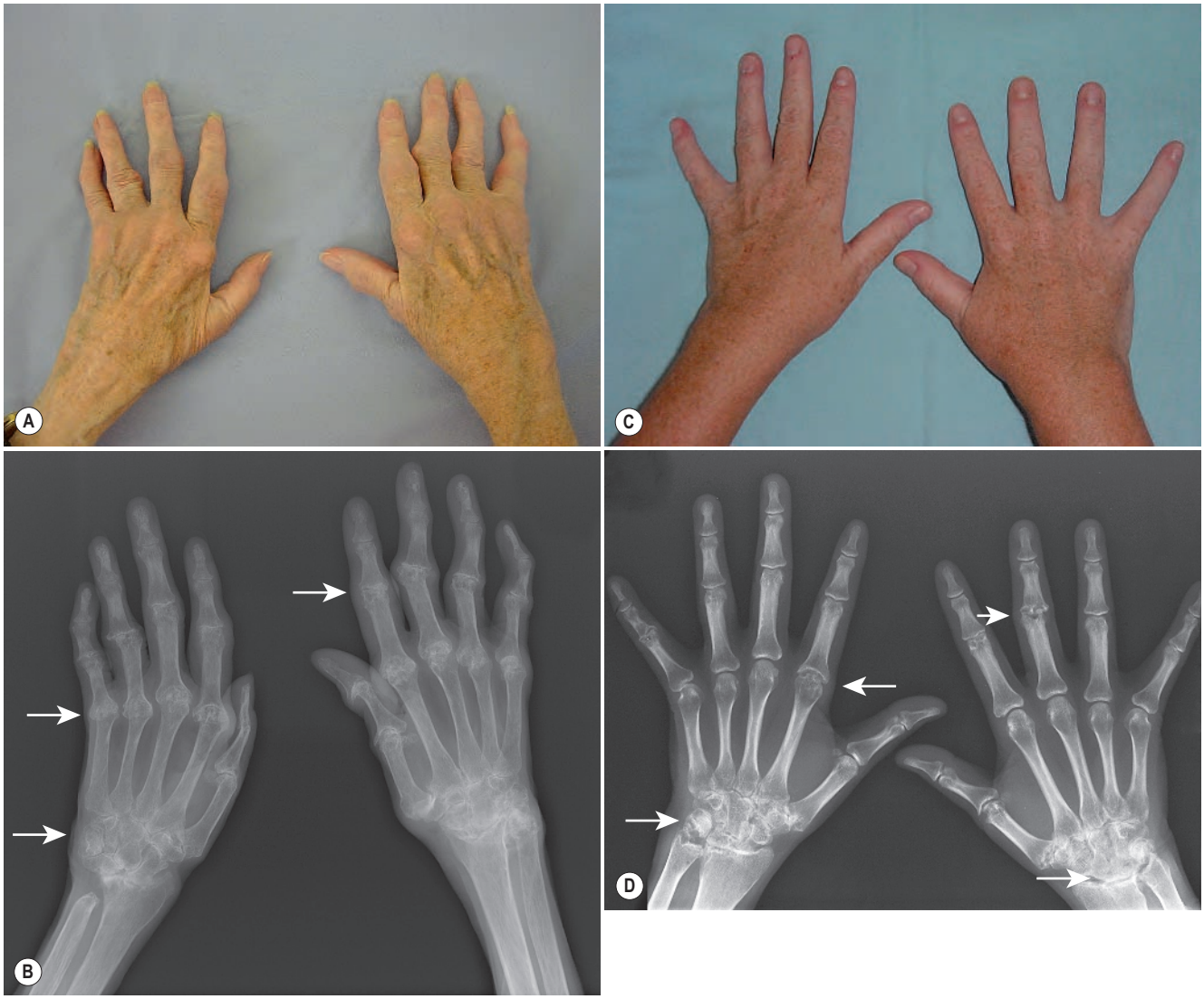


FIG. 53.5 Photomicrographs and Radiographs of the Rheumatoid Hand. Chronic, severe, and erosive RA, refractory to treatment and showing joint swelling and classical deformities (A and B), is compared with erosive disease whose progression has been attenuated by combination and biological therapy (C and D). White arrows indicate major areas of bone and cartilage destruction. (Reproduced with kind permission from the patients.)

levels, duration of synovitis (0 to 1), and acute-phase reactants erythrocyte sedimentation rate or C-reactive protein (0 to 1). A score of 6 or above is indicative of an early disease state requiring initiation of DMARDs such as methotrexate (Table 53.2), and so these criteria are thought to better reflect an intention to treat on the part of the supervising physician. Although the presence of synovitis in at least one joint is required (in the absence of an alternative diagnosis that better explains the synovitis), there remains debate as to whether subclinical synovitis of specific joints, defined by magnetic resonance imaging (MRI) or high-resolution ultrasonography (HRUS), should be included in the joint score.

Laboratory Findings

Until the late 1990s, IgM RFs, autoantibodies that recognize the Fc subunit of IgG, remained one of the few parameters of value in the clinical setting, forming the basis of the seropositive versus seronegative RA stratification and identifying those patients more likely to progress to erosive disease with or without extraarticular features. Nevertheless, RF can be detected in up to 5% of the healthy population and in 10% to 20% of the elderly population (>65 years of age); RF is found in a range of rheumatic conditions including Sjögren syndrome, SLE, and cryoglobulinemia, as well as in acute infectious and neoplastic disease entities, influencing its diagnostic utility. In general RF is not of value for monitoring responses to therapy.

The discovery of ACPAs, which can be detected very early in the disease process, has had a major impact on diagnostic practice as the assays have become more widely available.[20] They also have prognostic value in terms of radiographic progression, and titers may alter with therapy. The new-generation anti-CCPs kits have demonstrated diagnostic sensitivity of 80% and specificity of 98%. As the range of RA-associated autoantigens has expanded and the repertoire of citrullinated target autoantigens has become better defined, multiplex assays of serum autoantibodies are likely to play an increasingly important role in the diagnosis and prognosis of subsets of autoantibody-positive inflammatory arthritides.

TABLE 53.2 2010 ACR/EULAR Classification Criteria for Rheumatoid Arthritis

Domain	Weighted Score
Joint Involvement (0–5)	
1 medium to large joint	0
2–10 medium to large joints	1
1–3 small joints	2
4–10 small joints	3
>10 joints (with at least one small joint)	5
Serology (0–3)	
Neither RF- nor ACPAs-positive (≤ULN)	0
At least one test, low positive titer (>1 ≤3 × ULN)	2
At least one test, high positive titer (>3 × ULN)	3
Duration of Synovitis (0–1)	
<6 weeks	0
≥6 weeks	1
Acute-Phase Reactants (0–1)	
Neither ESR nor CRP abnormal	0
Abnormal ESR or CRP	1
TOTAL (≥6 indicates *definite* RA)	_____

ACPAs, Antibodies to citrullinated protein antigens; *ACR*, American College of Rheumatology; *CRP*, C-reactive protein; *ESR*, erythrocyte sedimentation rate; *EULAR*, European League Against Rheumatism; *RF*, rheumatoid factor; *ULN*, upper limit of normal.

CLINICAL PEARLS

Predictors of Poor Outcome in Rheumatoid Arthritis

- Chronic, unremitting disease onset, especially at advanced age
- High disease activity scores at baseline
- Female gender
- Poor functional status as determined by validated functional disability indices such as the Stanford Health Assessment Questionnaire (HAQ) and the Arthritis Impact Measurement Scale (AIMS)
- Low socioeconomic status
- Systemic and extraarticular features
- Depression and anxiety
- Comorbidity, for example, infection, cardiovascular disease, renal impairment
- Early erosive disease (in first 6–12 months; may be associated with ACPAs autoantibodies)
- Persistent acute-phase response (e.g., time-integrated CRP levels)
- Autoantibodies (RF and ACPAs) and HLA-DRB1 status (SE+)
- Significant delay in early use of DMARDs and corticosteroids

ACPAs, Antibodies to citrullinated protein antigens; *CRP*, C-reactive protein; *DMARDs*, disease-modifying antirheumatic drugs; *RF*, rheumatoid factor; SE, shared epitope.

The Preclinical Phase of Rheumatoid Arthritis

The identification of a preclinical phase of RA, that can be detected by virtue of serum autoantibodies up to 14 years before disease onset, has provided opportunities to characterize the RA prodrome in more detail. Work has centered around evaluation of acute-phase proteins and serum cytokines and chemokines, which appear to rise 1 to 2 years before onset of clinical disease; ACPAs isotype usage, N-glycan modifications, and epitope spreading; whole-blood gene expression analysis, which shows similarities to patients with established disease (including type I interferon-inducible gene signatures in a subset); imaging by MRI and ultrasonographic modalities; and, most recently, analysis of lymph node immune phenotypes, showing evidence of lymphocyte activation. The strongest predictors of progression in those at-risk subjects are joint pains without clinically detectable synovitis (arthralgia), combined with the presence of high-titer ACPAs and RF autoantibodies. Risk scores, which take into account other demographic factors, have been developed and are currently being validated. In 2016, a EULAR definition of the clinical characteristics of patients at risk of developing RA was published.[27]

Progression rates for ACPA+ arthralgia patients with subclinical synovitis detected by ultrasound examination are of the order of 20% over a median of 28 months,[28] while arthritis-free survival curves suggest that 40% to 50% of those at highest risk will develop clinically apparent synovitis in 2 to 3 joints within 24 months. Somewhat surprisingly, genetic markers, such as HLA-DRB1 SE, have not contributed dramatically to risk estimates, while emerging patterns of tenosynovitis by ultrasonography or MRI, or bone marrow oedema (or osteitis) by MRI define some of the earliest articular features of this at-risk state. Rates of progression are remarkably consistent across at-risk cohorts and are considered sufficient to justify secondary prevention intervention studies. Thus, in 2019 the effects of a single IV dose of rituximab were studied in ACPA+ arthralgia subjects. Compared to a control group (who also received IV hydrocortisone at baseline), rituximab delayed disease onset by about 12 months. As of spring 2020, there are three other clinical trials actively examining the impact of a fixed period of therapy with hydroxychloroquine, methotrexate, or abatacept versus placebo and followed up for a 2- to 3-year period. If delay or prevention is observed, even in a subset of individuals, this is likely to transform the way we approach the treatment of RA.

ON THE HORIZON

The Preclinical Phase of Rheumatoid Arthritis

- High-risk individuals include those with inflammatory joint pain (arthralgia) and serum RA–associated autoantibodies. From 40% to 50% of those with high-titer ACPAs (with or without rheumatoid factor) may develop clinical synovitis within 24 months.
- Targeting these high-risk individuals with preventative strategies provides the best chance of achieving cure. Trials of rituximab, abatacept, or hydroxychloroquine are ongoing.
- In-depth molecular and cellular studies for characterizing the preclinical phase of disease will be critical to the success of this endeavor.
- Models for progression to RA have identified the following phases:
  - I. Genetic risk
  - II. Genetic risk with autoimmunity (e.g., ACPAs, rheumatoid factor)
  - III. Genetic risk with autoimmunity and arthralgia (but absence of clinically apparent synovitis)
  - IV. Undifferentiated arthritis (clinically apparent synovitis, not fulfilling RA classification criteria)
  - V. Early RA (fulfilling disease criteria; need for DMARDs)
- Stratification of risk, including genetic, serological, and demographic factors, will permit the identification of subjects most suitable for intervention studies.
- Studying the impact of lifestyle modifications, for example, diet, stopping smoking, would be of great interest.
- Reestablishing immune homeostasis and/or induction of immune tolerance may provide the best chance of achieving cure in subjects during the preclinical phase of disease. Early-phase studies of peptide immunotherapy indicate that this approach is well tolerated.
- Establishing robust assays for signatures of immune tolerance (e.g., immune phenotyping by flow cytometry, extended autoantibody serotyping, multiplex assays for detection of inflammatory mediators in serum, whole-blood transcriptomic profiles) that reflect a healthy immune system will greatly facilitate these endeavors.

ACPAs, Antibodies to citrullinated protein antigens; *DMARDs*, disease-modifying antirheumatic drugs; *RA*, rheumatoid arthritis.

TREATMENT

Disease-Modifying Antirheumatic Drugs

THERAPEUTIC PRINCIPLES

Treatment Paradigms in Rheumatoid Arthritis

- Education and counseling through early involvement of a multidisciplinary team, including specialist nurse and other health-care professionals; appropriate balance of rest and exercise during disease flares
- Adequate nutrition (especially important with severe, active disease)
- Comprehensive assessment of disease activity, especially during early phase of disease to achieve rapid disease control
- Complete suppression of inflammation early in the disease, with tight control through regular and frequent reassessments focused around disease activity scores
- Yearly imaging assessments to monitor radiographic progression, for example, X-rays or high-resolution ultrasonography of hands and feet
- Early use of DMARD/SAARDs
- Early use of corticosteroids, including use of intraarticular joint injections to suppress inflammation, and use of step-up combination therapy in severe disease
- Appropriate use of biologicals, for example, early use of anticytokine, T- or B-cell–targeted therapies in severe disease
- Early relief of pain with judicious use of NSAID or COX2 inhibitors according to safety/risk profile
- Monitoring for drug toxicity
- Effective contraception, where appropriate
- Bone protection
- Monitoring for risk factors of key comorbidities, including cardiovascular disease
- Prevention of infection through vaccination (preferably before instituting immunosuppressive agents), for example, against influenza, pneumococcus, and herpes zoster

COX, Cyclooxygenase; *DMARD*, disease-modifying antirheumatic drugs; *NSAID*, nonsteroidal anti-inflammatory drugs; *SAARD*, slow-acting antirheumatic drugs.

Over the past two decades there has been a dramatic paradigm shift in the therapy of RA from control of symptoms to the control of the disease process and aggressive suppression of inflammation. This shift has come about through a growing appreciation of the relationship between joint inflammation and joint destruction, as well as the development of imaging technology for detecting the very earliest changes in joint structures. The impact of this paradigm shift in therapeutic terms is striking. Traditional "go-low, go-slow" regimens of the 1970s and 1980s included the initiation of nonsteroidal anti-inflammatory drugs (NSAIDs), followed by implementation of DMARDs only after destructive disease became evident. Depending on the clinical response, sequential monotherapy was the norm. Although this strategy may still be appropriate for patients with mild disease, current practice now dictates aggressive combination therapy (two or more conventional DMARDs) from the outset for patients with poor prognostic factors, with preference for the faster-acting DMARDs such as methotrexate, leflunomide, and sulfasalazine (onset 3 to 6 weeks) over slower-acting agents such as hydroxychloroquine, gold, and D-penicillamine (onset 3 to 6 months), but with the addition of oral or parenteral prednisolone. More recent data suggest that the specific choice of therapy may be less important than the strategy. For example, the pioneering TICORA and BeST studies both indicate that intensive treatment when combined with intensive control most convincingly influences outcome measures, including clinical response, retention, functional status, and radiographic progression. The benefits of tight control have been substantiated in many subsequent clinical trials, as well as in "real-life" clinical practice.[29]

Anti-Cytokine Therapy

The introduction of targeted therapy to the clinic using biological agents (e.g., chimeric or fully humanized antibodies to ligands or receptors, soluble receptor fusion proteins, or recombinant receptor antagonists) has transformed the treatment of RA. The prototype biologic, developed in the 1990s, was TNF blockade.[30] The rationale for inhibiting TNF bioactivity is based upon its pleiotropic effects on cell activation, cellular adhesion and migration, induction of cytokine and inflammatory gene mRNA and protein, neoangiogenesis, and the regulation of cartilage catabolic factors such as IL-1 and matrix metalloproteinases (see Fig. 53.3). TNF and other inflammatory cytokines such as IL-1, IL-6, IL-15, IL-17, and GM-CSF are expressed constitutively in inflamed synovial tissue at mRNA and protein level. In many cases the expression of their high-affinity cognate receptors is also upregulated and the functional activity of the corresponding naturally occurring inhibitors (e.g., soluble TNF-R or IL-1Ra) is reduced further, promoting persistence of the inflammatory response (although levels of protein may be increased, reflecting an attempt at restoring homeostasis).

Chimeric anti-TNF monoclonal antibodies (infliximab) were first used to treat RA in open-label clinical trials in 1992.[30] Humanized antibodies (adalimumab) and the soluble p75 TNF-R IgG fusion protein (etanercept) were tested soon after with comparable therapeutic effects; golimumab and a construct comprising a PEGylated anti-TNF antibody Fab fragment (certolizumab) are also licensed for use in patients with RA, as are a growing number of "biosimilar" TNF inhibitors. TNF-α blockade leads to dramatic and rapid reductions in symptoms (pain, stiffness, and fatigue) and signs (joint pain and swelling) of arthritis in a dose-dependent fashion, and in a significant proportion of patients (~60% to 70%) who have failed conventional DMARDs.[30] As a class, and when used in combination with methotrexate, the majority are superior to either drug alone.

The clinical benefit of TNF blockade has prompted extensive mechanism of action studies.[30] Anti-TNF reduces the acute-phase response, including IL-6 serum levels. Leukocyte trafficking is inhibited, as demonstrated through an early (within hours) and dramatic rise in lymphocyte counts through demargination, a more prolonged and sustained exclusion of leukocytes based on reductions in cellularity of synovial tissue biopsies after treatment and suppression of markers of angiogenesis, including VEGF; TNF blockade promotes lymphangiogenesis and may facilitate cell egress from inflamed synovium. TNF blockade has also been shown to downregulate markers of cartilage and bone destruction, including the collagenases MMP1 and 3, and to reduce the ratio of RANKL and OPG in serum, effects that might explain in part the joint-preserving effects of anti-TNF *in vivo*.

The IL-1 receptor antagonist (IL-1Ra) is the only IL-1 inhibitor currently licensed for use in RA. It has disease-suppressing effects in animal models of arthritis, with potent joint protection, but has proven less effective than anti-TNF in patients with RA. Nevertheless, it has been used effectively to treat patients who have failed TNF blockade, slowing radiographic progression, and may be efficacious in a subset of individuals with more systemic autoinflammatory

syndromes. Anti-IL-6R blockade (tocilizumab) is licensed for use in patients with established RA. Evidence suggests that in early disease, tocilizumab as monotherapy is as effective in suppressing signs and symptoms of RA as when it is combined with methotrexate. These effects are mediated through blocking IL-6 actions on the immune response, the acute-phase response, osteoclastogenesis, B-cell activation and immunoglobulin production, angiogenesis, and cell adhesion (reviewed in Kishimoto).[31] Clinical responses are comparable to those observed with TNF blockade, and a range of IL-6 inhibiting drugs are now licensed for use. Unlike in psoriasis, the effect of inhibiting IL-17 in RA (with humanized monoclonal antibodies ixekizumab or secukinumab that block IL-17A) has been more modest, although there may exist a subset of RA patients in whom IL-17-driven inflammatory responses dominate. A role for blocking GM-CSF is emerging.

Finally, the impact of using small molecule inhibitors of cytokine signals, transduced through receptors that utilize members of the Janus kinase (JAK) family of tyrosine kinases,[32] has been evaluated in some detail. These are interesting, immunologically important kinases, because gain-of-function mutants are associated with leukemia and lymphoma, whereas loss-of-function is associated with primary immunodeficiency. Jakinibs, including tofacitinib, baricitinib, and upadacitinib, have been studied in phase III clinical trials of RA patients. Tofacitinib has high affinity for JAK3, but is also able to partially inhibit JAK1 and, to a lesser extent, JAK2; baricitinib blocks JAK1 and JAK2 to a similar extent, with some Tyk2 targeting activity, while upadacitinib has more JAK1 selectivity. These agents are well tolerated, with acceptable safety profiles and now provide physicians with an orally available agent with efficacy comparable to biologic agents. This is likely due to the wide range of inflammatory pathways that are targeted, including common gamma chain cytokines IL-2, IL-7, IL-15, and IL-21, type I and type II interferons, and IL-6 family cytokines.

Anti-T-Cell Therapy

The contribution of costimulatory signals ("signal 2") transduced through CD28 to priming and activation of naïve T cells and amplification of cytokine gene expression and proliferative responses has provided a rationale for testing costimulatory blockade in patients. This has been achieved using a nondepleting humanized IgG1-CTLA-4 fusion protein that prevents CD80 and CD86 (expressed on antigen-presenting cells) from engaging CD28 (but also CTLA-4) expressed on T cells. Initial studies confirmed that the agent was safe and well tolerated.[33] As well as suppressing disease activity, CTLA-4-Ig (licensed as abatacept as intravenous and subcutaneous formulations) inhibits radiographic progression and structural damage and it is also effective in treating those patients who have had inadequate responses to TNF blockade as well as to methotrexate. Head-to-head studies indicate that the kinetics of response are remarkably similar to those of anti-TNF when compared as monotherapy, indicating that a significant number of costimulation-dependent T cells actively participate in the ongoing inflammatory response. Clinical improvement may continue beyond 12 months. Recent data suggest that inhibition of CD28 signals reduce the number of T_{FH}. Abatacept targets one of the earliest stages of adaptive immunity, and so offers a biologically plausible pathway to target in subjects at high risk of progressing to RA.

Anti-B-Cell Therapy

Rituximab is a humanized monoclonal antibody that recognizes human CD20, a 33- to 37-kDa membrane-associated phosphoprotein expressed on pre-B, immature, and mature B cells but not on plasma cells. It was initially developed, and then licensed, for the treatment of non-Hodgkin lymphoma. While CD20 ligation promotes B-cell activation, differentiation, and cell cycle progression, the function of CD20 is still poorly understood. The therapeutic effects of anti-CD20 are related to B-cell depletion, which can vary between patients due to antibody-dependent cell cytotoxicity, complement-mediated cell lysis, and/or triggering of intracellular pathways for apoptotic cell death.[34] The effects on serum immunoglobulin levels are modest with levels remaining in the normal range, unless patients undergo repeated cycles of B-cell depletion.

A pivotal placebo-controlled trial of rituximab therapy randomized 161 patients with RF-positive RA to compare the efficacy and safety of methotrexate alone (standard therapy) versus methotrexate plus rituximab (1000 mg on days 1 and 15), rituximab alone, or rituximab plus cyclophosphamide.[34] Up to 43% of patients receiving the combination of rituximab and methotrexate achieved 50% improvement in clinical and laboratory parameters after 24 weeks (ACR50), and this was at least as good as the rituximab/cyclophosphamide combination (41% achieving ACR50) and superior to methotrexate (13%) or rituximab alone (33%). Follow up studies indicate that B-cell repopulation occurs at a mean of 8 months after treatment, comprising immature $IgD^+CD38^+CD27^-CD5^+$ B cells, and is associated with increased serum B cell–activating factor (BAFF) levels. Early relapse is associated with reconstitution of the $CD27^+$ memory B-cell compartment.

Future Prospects for Therapy

There remain unmet needs in the treatment of RA. Principal among these are the fact that until relatively recently, treatment has been considered lifelong for the majority of patients, imposing greater risk of toxicity and, as the immune system senesces, increased risk of infection or lymphoproliferative disease. There is little doubt that early treatment with tight control offers the best outcomes, and evaluation of synthetic and biological DMARDs in at-risk subjects to establish whether they can delay or even prevent development of clinically apparent disease could be transformative. A better appreciation of the profile of RA-specific autoantigens has prompted early-phase studies to establish the safety and tolerability of peptide immunotherapy delivered by autologous DCs. From an immunological perspective, there remains a pressing need to develop immunological tools or immune biomarkers that can redefine disease subsets and measure effector and regulatory cell subsets; progress has been made with synovial pathotypes. While such tools may provide better insights into disease pathogenesis, they can also be adapted to monitor the impact of therapeutic intervention, whether this turns out to be cell-based therapy or the application of novel immune modulators. Similar approaches should be used to identify those $ACPA^+$ arthralgia patients who have a genetic predisposition to RA and are at highest risk of developing the disease.

REFERENCES

1. Thomas R, Cope AP. Pathogenesis of Rheumatoid Arthritis. In: Watts R, ed. *Oxford Textbook of Rheumatology*. : Oxford University Press; 2016:839–848.
2. Frisell T, Saevarsdottir S, Askling J. Family history of rheumatoid arthritis: an old concept with new developments. *Nat Rev Rheumatol*. 2016;12:335.
3. van der Woude D, Houwing-Duistermaat JJ, Toes RE, et al. Quantitative heritability of anti-citrullinated protein antibody-positive and anti-citrullinated protein antibody-negative rheumatoid arthritis. *Arthritis Rheum*. 2009;60:916.
4. Karlson EW, Lee IM, Cook NR, et al. A retrospective cohort study of cigarette smoking and risk of rheumatoid arthritis in female health professionals. *Arthritis Rheum*. 1999;42:910.
5. Padyukov L, Silva C, Stolt P, et al. A gene-environment interaction between smoking and shared epitope genes in HLA-DR provides a high risk of seropositive rheumatoid arthritis. *Arthritis Rheum*. 2004;50:3085.
6. Lundberg K, Wegner N, Yucel-Lindberg T, et al. Periodontitis in RA—the citrullinated enolase connection. *Nat Rev Rheumatol*. 2010;6:727.
7. Zhang X, Zhang D, Jia H, et al. The oral and gut microbiomes are perturbed in rheumatoid arthritis and partly normalized after treatment. *Nat Med*. 2015;21:895.
8. Wells PM, Adebayo AS, Bowyer RCE et al. The gut microbiota and genetic risk for rheumatoid arthritis in the absence of disease: a TwinsUK association study. Lancet Rheumatology (in press).
9. Gregersen PK, Silver J, Winchester RJ. The shared epitope hypothesis. An approach to understanding the molecular genetics of susceptibility to rheumatoid arthritis. *Arthritis Rheum*. 1987;30:1205.
10. Raychaudhuri S, Sandor C, Stahl EA, et al. Five amino acids in three HLA proteins explain most of the association between MHC and seropositive rheumatoid arthritis. *Nat Genet*. 2012;44:291.
11. Scally SW, Petersen J, Law SC, et al. A molecular basis for the association of the *HLA-DRB1* locus, citrullination and rheumatoid arthritis. *J Exp Med*. 2013;210:2569.
12. Eyre S, Bowes J, Diogo D, et al. High-density genetic mapping identifies new susceptibility loci for rheumatoid arthritis. *Nat Genet*. 2012;44:1336.
13. Burn GL, Cornish GH, Potrzebowska K, et al. Super-resolution imaging of the cytoplasmic phosphatase PTPN22 links integrin-mediated adhesion with autoimmunity. *Sci Signaling*. 2016;9:ra99.
14. Manzo A, Bombardieri M, Humby F, et al. Secondary and ectopic lymphoid tissue responses in rheumatoid arthritis: from inflammation to autoimmunity and tissue damage/remodeling. *Immunol Rev*. 2010;233:267.
15. Lewis MJ, Barnes MR, Blighe K, et al. Molecular Portraits of Early Rheumatoid Arthritis Identify Clinical and Treatment Response Phenotypes. *Cell Rep*. 2019;28:2455–2470.
16. Mizoguchi F, Slowikowski K, Wei K, et al. Functionally distinct disease-associated fibroblast subsets in rheumatoid arthritis. *Nat Commun*. 2018;9:789.
17. Croft AP, Campos J, Jansen K, et al. Distinct fibroblast subsets drive inflammation and damage in arthritis. *Nature*. 2019;570:246–251.
18. Culemann S, Grüneboom A, Nicolás-Ávila JÁ, et al. Locally renewing resident synovial macrophages provide a protective barrier for the joint. *Nature*. 2019;572:670–675.
19. McInnes IB, Schett G. Cytokines in the pathogenesis of rheumatoid arthritis. *Nat Rev Immunol*. 2007;7:429.
20. van Venrooij WJ, Vossenaar ER, Zendman AJ. Anti-CCP antibodies: the new rheumatoid factor in the serology of rheumatoid arthritis. *Autoimmun Rev*. 2004;3(Suppl. 1):S17.
21. Klareskog L, Rönnelid J, Lundberg K, et al. Immunity to citrullinated proteins in rheumatoid arthritis. *Annu Rev Immunol*. 2008;26:651.
22. Scherer HU, Huizinga TWJ, Krönke G, et al. The B cell response to citrullinated antigens in the development of rheumatoid arthritis. *Nat Rev Rheumatol*. 2018;14:157–169.
23. Rao DA, Gurish MF, Marshall JL, et al. Pathologically expanded peripheral T helper cell subset drives B cells in rheumatoid arthritis. *Nature*. 2017;542:110–114.
24. Walter GJ, Fleskens V, Frederiksen KS, et al. Phenotypic, functional, and gene expression profiling of peripheral CD45RA+ and CD45RO+ CD4+CD25+CD127(low) Treg cells in patients with chronic rheumatoid arthritis. *Arthritis Rheumatol*. 2016;68:103.
25. Calabrese LH, Calabrese C, Cappelli LC. Rheumatic immune-related adverse events from cancer immunotherapy. *Nat Rev Rheumatol*. 2018;14:569–579.
26. Harre U, Schett G. Cellular and molecular pathways of structural damage in rheumatoid arthritis. *Semin Immunopathol*. 2017;39:355–363.
27. van Steenbergen HW, Aletaha D, Beaart-van de Voorde LJ, et al. EULAR definition of arthralgia suspicious for progression to rheumatoid arthritis. *Ann Rheum Dis*. 2017;76:491–496.
28. Bos WH, Wolbink GJ, Boers M, et al. Arthritis development in patients with arthralgia is strongly associated with anti-citrullinated protein antibody status: a prospective cohort study. *Ann Rheum Dis*. 2010;69:490.
29. Goekoop-Ruiterman YP, de Vries-Bouwstra JK, Allaart CF, et al. Clinical and radiographic outcomes of four different treatment strategies in patients with early rheumatoid arthritis (the BeSt study): a randomized, controlled trial. *Arthritis Rheum*. 2005;52:3381.
30. Feldmann M, Maini RN. Anti-TNF alpha therapy of rheumatoid arthritis: what have we learned? *Annu Rev Immunol*. 2001;19:163.
31. Kishimoto T. Interleukin-6: from basic science to medicine—40 years in immunology. *Annu Rev Immunol*. 2005;23:1.
32. Schwartz DM, Kanno Y, Villarino A, Ward M, Gadina M, O'Shea JJ. JAK inhibition as a therapeutic strategy for immune and inflammatory diseases. *Nat Rev Drug Discov*. 2017;16:843–862.
33. Kremer JM, Westhovens R, Leon M, et al. Treatment of rheumatoid arthritis by selective inhibition of T-cell activation with fusion protein CTLA4Ig. *N Engl J Med*. 2003;349:1907.
34. Edwards JC, Cambridge G. B-cell targeting in rheumatoid arthritis and other autoimmune diseases. *Nat Rev Immunol*. 2006;6:394.

Juvenile Idiopathic Arthritis

Randy Q. Cron, Peter Weiser, and Timothy Beukelman

Juvenile idiopathic arthritis (JIA) is a collection of chronic inflammatory arthropathies of childhood. By definition, JIA occurs prior to the 16th birthday, has no known causes, and has evident chronic, more than 6 weeks' duration of, joint inflammation.[1]

The currently utilized 2004 revision of the International League of Associations for Rheumatology (ILAR) nomenclature and subcategorization of JIA unifies the American (juvenile rheumatoid arthritis [JRA]) and European (juvenile chronic arthritis [JCA]) classification schemes.[1] It addresses arthritis features different from those in adults and removes the word "rheumatoid" to distinguish it from adult rheumatoid arthritis (RA) (Chapter 53). The ILAR classification of JIA also includes the human leukocyte antigen HLA-B27–associated spondyloarthropathies (Chapter 58), which occur frequently during childhood (Table 54.1). There are still ongoing efforts to further re-classify JIA subcategories based on genetic research suggestive of some of the subgroups being actually identical to adult forms of chronic arthritis rather than existing as a separate childhood entity (e.g., rheumatoid factor [RF] positive polyarticular JIA and RA).[2]

Because JIA is a diagnosis of exclusion, other entities with chronic arthritis need to be ruled out first. Sarcoidosis (Chapter 72), Sjögren syndrome (SS) (Chapter 55), systemic lupus erythematosus (SLE) (Chapter 52), mixed connective tissue disease (MCTD) (Chapter 56), Lyme disease (Chapter 27), dermatomyositis (Chapter 57), and a variety of vasculitides (Chapters 59 and 60) can present with joint inflammation. While JIA is idiopathic by definition, chronic arthritis as part of inflammatory bowel disease (IBD) (Chapter 75) or associated with psoriasis (Chapter 64) is included under the JIA umbrella. When considering the seven categories of JIA together, the estimated prevalence of JIA is roughly 1 in 1000 children, with rates varying in different regions of the world, likely because of genetic risk factors and/or environmental triggers. Moreover, the relative rates of JIA categories vary in different regions of the world (e.g., oligoarticular JIA is common in Scandinavia, and enthesitis-related JIA is more typical in Latin America with the caveat of some variability in classification).

ETIOLOGY AND PATHOGENESIS

Genetic Contribution

Because JIA is an eclectic group of unique categories based largely on empirical clinical phenotypes and is far less common than RA, understanding its genetic causes is extremely difficult. Clinical examination is neither sensitive nor specific for diagnosing joint inflammation, and certain joints like in the case of the temporomandibular joint (TMJ) require imaging to screen for involvement.[3] Thus, the arbitrary distinction of subgroups based on number of involved joints can lead to misclassification, which would subsequently influence data analysis. Distinctions based on the types and location of involved joints and data from biosamples may prove useful in identifying more homogeneous subtypes and lead to better genetic understanding of etiology.[4]

Polygenic Disorder

In general, JIA is a polygenic autoimmune condition with ≥20 potential genes involved. An exception is systemic JIA (sJIA), which resembles both an autoimmune and an autoinflammatory disorder.[5] Unlike autoimmunity, which is believed to be the result of an imperfect adaptive immune system, autoinflammatory conditions are thought to result from genetic perturbations in the innate immune response, for example, as in the autoinflammatory recurrent fever syndromes.[5] Whether sJIA is the result of a single or multiple gene defects in the same immune pathway is currently unknown. However, a relatively common complication of sJIA is the life-threatening condition of macrophage activation syndrome (MAS), which resembles familial hemophagocytic lymphohistiocytosis (fHLH) (Fig. 54.1). MAS may be present in ≈50% of children with sJIA in either an overt or occult/subclinical fashion.[6] Recently, genetic polymorphisms and heterozygous (single copy) mutations in genes associated with fHLH when present as homozygous defects have been identified in children with sJIA and overt MAS.[6] Protein products of these genes, including perforin 1 and MUNC 13-4, are critical in the pathway mediating cytolysis by CD8 T cells and natural killer (NK) cells (Chapter 12). This cytolysis is crucial to the ability to shut down an immune response following control of infection. Defects in this pathway can result in a hyperinflammatory state (cytokine storm), leading to pancytopenia, coagulopathy, and multisystem organ failure.[6] As sJIA is the most common condition resulting in MAS during childhood, it might indeed be genetically related to fHLH but possibly associated with heterozygous, rather than homozygous, mutations in genes critical for cytolysis.[6]

Human Leukocyte Antigen Associations

For all other JIA categories, polygenic influences—including the major histocompatibility complex (MHC)—contribute to disease pathology. The MHC is densely packed (greater than 200 genes) with immune-associated genes (Chapter 5). It is also the most polymorphic region of the human genome and gives rise to MHC class I and class II genes, complement proteins, tumor

TABLE 54.1 Clinical and Laboratory Features of the Juvenile Idiopathic Arthritis Categories

	JUVENILE IDIOPATHIC ARTHRITIS					
Classifications	**Former Juvenile Rheumatoid Arthritis**				**Spondyloarthropathy**	
Juvenile idiopathic arthritis categories/feature	Systemic	RF⁻ polyarticular	RF⁺ polyarticular	Oligoarticular (persistent and extended)	Psoriatic (early and late onset)	Enthesitis-related arthritis (ERA) (including ankylosing spondylitis (AS) and inflammatory bowel disease [IBD])
Human leukocyte antigen (HLA)	DRB1*11	DRB1*08	DRB1*04	A2	DRB1*01	B27
Gender	Equal	F ≫ M	F ≫ M	F ≫ M	F > M	M ≫ F
Age at onset	Peak 2 years	Dual peaks	Teenage	1–3 years	Dual peaks	Teenage
Antinuclear antibody (ANA)	Rare	Yes	Yes	Yes	Yes	Rare
Uveitis	Rare	Yes	Rare	Yes	Yes	Nonsilent
Temporomandibular joint (TMJ)	Yes	Yes	Yes	Yes	Yes	Yes
Enthesitis	No	No	No	No	Older age	Yes
Arthritis	Erosive	Symmetric	Erosive	Mixed	Dactylitis	Axial

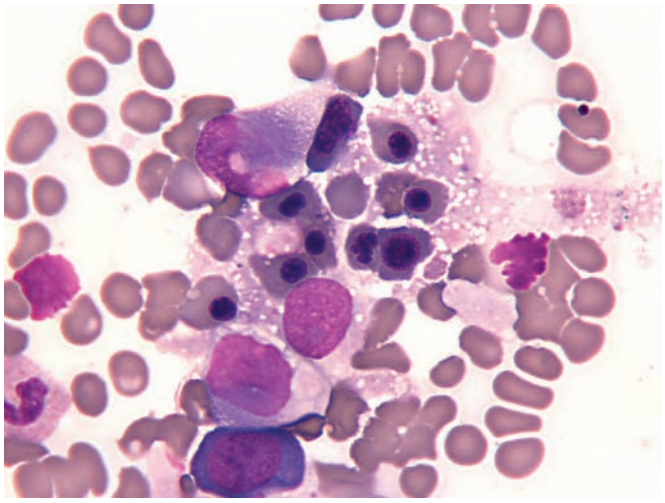


FIG. 54.1 Hemophagocytosis as Part of Macrophage Activation Syndrome. A centrally located, vacuolated histiocyte is pictured engulfing numerous nucleated immune cells and nonnucleated mature red blood cells. Wright stain at ×198 magnification. (Courtesy of Dr. David Kelly.)

necrosis factor (TNF), and others. The critical role of MHC proteins in preventing autoimmunity by shaping the T-cell repertoire (Chapter 9) likely explains why the MHC is the most consistently and strongly associated genetic locus for most JIA categories. Recently, MHC class II associations have been identified in children with sJIA, providing evidence for an autoimmune component in addition to the autoinflammatory one, as mentioned above.[7]

By comparison, enthesitis-related arthritis (ERA) and psoriatic JIA share one of the strongest known MHC associations, HLA-B27, for any described autoimmune disorder. HLA-B27 is present in 70% to 90% of Caucasian children with ERA, which follows an autosomal dominant pattern and is the only JIA category likely to have similarly afflicted first-degree relatives. The pathophysiological explanation for the HLA-B27 association remains unknown, but hypotheses vary from molecular mimicry of pathogens presented by HLA-B27 (Chapter 51) to the unfolded protein response.[8]

The remaining JIA categories have all been linked to various other MHC proteins to lesser degrees (see Table 54.1).

Nonhuman Leukocyte Antigen Associations

Genes outside of the MHC have also been linked to developing JIA. Genetic approaches have highlighted the importance of genetic differences in the development of JIA, but at most, 11% of the contribution can be linked to the MHC.[4] Recent developments in high-throughput genetic sequencing, combined with the identification of densely present restriction fragment length polymorphisms (RFLPs) throughout the human genome, have allowed for powerful new genetic approaches. These include genome-wide association studies (GWAS) designed to identify genes associated with a variety of disorders, including autoimmune diseases such as JIA.[9] To date, several genes have been reported to be associated with JIA.[4]

Strong evidence that genetic factors contribute to JIA susceptibility include twin and family studies. Data from a JIA national registry revealed that monozygotic twin pairs have a higher-than-expected proportion of twins with JIA. Moreover, siblings of children with JIA have as high as a 30-fold increased risk of JIA compared with the general population. Interestingly, siblings with JIA typically share the same JIA category, age of onset, and disease course.[4] Polymorphisms in MHC class II genes have been estimated to account for less than 20% of the recurrence risk of JIA in siblings and thus support the concept that JIA is a complex genetic trait. JIA likely shares general or common autoimmunity gene risk factors with many autoimmune disorders but may also be influenced by specific JIA risk-associated genes.[9]

Using linkage and association studies, researchers have identified a variety of potentially JIA-associated genes. However, outside of the human leukocyte antigen (HLA) genes, only a small percentage of these genes have been independently confirmed by other investigators. A few of these genes/gene products include *PTPN22* (a phosphatase involved in inhibition of T-cell activation), *WISP3* (a signaling protein), interleukin-1α (a proinflammatory cytokine), tumor necrosis factor (TNF; another proinflammatory cytokine), macrophage migratory inhibitory factor (MIF), and *SLC11A1* (a resistance factor to intracellular pathogens in macrophages).[4] Thus, gene products linked to both innate and adaptive immune responses likely contribute

to JIA disease susceptibility. Ongoing GWAS may help validate the importance of these genes and may identify other candidate gene risk factors for JIA in the future. Recently, epigenetic (modifications to DNA nucleotides rather than changes in the DNA sequence) risk factors that confer susceptibility to JIA are beginning to be explored.[10] Indeed, chromatin architecture itself is being analyzed to better understand the mechanisms of gene expression in JIA.[11]

ENVIRONMENTAL FACTORS

In addition to genetic factors, there are likely a variety of environmental triggers for developing JIA in genetically susceptible hosts. Numerous infectious agents (e.g., parvovirus, rubella virus) have been explored as risk factors for development of JIA; however, replication of these associations has been difficult. In addition, some scientists have suggested that heat shock proteins, resulting from cells undergoing environmental stress, may contribute to the development of JIA.[12] There is also a clear link between gut pathogens and potentially commensal organisms (gut microbiome) and the development of HLA-B27–associated spondyloarthropathies.[13] The role of the gut microbiome is also implicated in the pathogenesis of other subcategories of JIA.[14] Last, prior to the identification of *Borrelia* as a cause of childhood arthritis, Lyme disease–associated arthritis was difficult to distinguish from oligoarticular JIA.[15] Perhaps because of the variety of potential environmental triggers for JIA, combined with the number of different JIA categories, no single environmental trigger has been conclusively identified as contributing to the development of JIA.

IMMUNE ABNORMALITIES

Autoantibodies

As JIA is considered an autoimmune disease, except for aspects of sJIA as discussed earlier, there have been a variety of explorations into the role of various components of the immune system involved in JIA pathogenesis. The importance of the immune system in JIA pathology is highlighted by the increased incidence of childhood chronic arthritis among children with various immunodeficiencies. For example, children with immunoglobulin A (IgA) deficiency are at increased risk of developing chronic arthritis. In contrast, the presence of specific autoantibodies is associated with various forms of JIA. ANAs are present in up to 40% of patients with JIA, particularly those with oligoarticular JIA, and are associated with silent uveitis (Chapter 74).[16] Similarly, antibodies to the nuclear oncoprotein DEK have been associated with uveitis as well as joint inflammation in children with JIA. IgM RF is present in a smaller subset of children with polyarticular JIA and is associated with a more aggressive/erosive form of arthritis as in adults with rheumatoid factor–positive (RF^+) RA (Chapter 53).[1] More recently, identification of anti–cyclic citrullinated peptide (CCP) antibodies have been found in a partially overlapping subset of children with RF^+ poly-JIA.[17] Thus, a variety of autoantibodies are associated with JIA.

T-Helper Cells

T lymphocytes are also thought to play a major role in the development of JIA, as evidenced by their relative predominance among mononuclear cells in synovial fluid of chronically inflamed joints in children with JIA. CD4 T-helper (Th) cells have been categorized into a variety of cytokine-producing subsets (Chapter 11). CD4 Th1 cells, characterized by the production of interferon-γ (IFN-γ), have been identified in chronically inflamed joints in children with JIA, whereas IL-4–producing Th2 CD4 T cells are more commonly involved in the joints in oligoarticular JIA (generally less aggressive arthritis) than in polyarticular JIA. More recently, two additional CD4 Th subsets, regulatory T cells (Tregs) and Th17 cells, are being examined as potential players in JIA pathology.

$CD4^+$, $CD25^{high}$ Tregs are characterized by the transcription factor FoxP3 and by the ability to suppress immune activation (Chapter 13).[18] The ability of Tregs to suppress other T cells likely occurs by both cell-contact dependent (through the surface protein cytotoxic T-lymphocyte antigen-4 [CTLA-4]) and independent (via suppressive cytokines) mechanisms.[18] Tregs secrete the antiinflammatory cytokines IL-10 and transforming growth factor-β (TGF-β). Th17 cells, characterized by the transcription factor retinoid orphan receptor (ROR)γT, produce the proinflammatory cytokine IL-17.[19] IL-17 is thought to contribute to a variety of autoimmune disorders, including JIA. The balance between these two juxtaposed Th subsets may determine whether autoimmunity develops (Th17-dominant) or a state of immune tolerance to self (Chapter 10) persists (Treg-dominant). Indeed, recent studies have identified a predominance of Th17 cells and associated proinflammatory cytokines in the inflamed joints of children with JIA.[19] Thus, a balance of proinflammatory cytokines and suppressive cytokines may dictate the expression of autoimmunity in the form of JIA.

Cytokines

Cytokines (Chapter 14) and, particularly, their inhibition have taken a prominent status in the pathology and treatment of chronic arthritis, including JIA. The bench-to-bedside translation of anti-TNF therapy to the treatment of chronic arthritis has revolutionized the care of adults with chronic arthritis as well as children with JIA. Inhibition of this proinflammatory cytokine in the circulation (via specific monoclonal antibodies [mAb] or receptor fusion proteins) rapidly and effectively treats most forms of JIA.[1] The one exception is sJIA, which may or may not respond well to anti-TNF treatment. However, other proinflammatory cytokines, including IL-1, -6, and -18, are thought to be central to sJIA pathogenesis.[1] Indeed, serum from sJIA patients was shown to induce transcription of a variety of innate immunity genes, including IL-1, in normal peripheral blood mononuclear cells. Fortunately, novel therapies that target either IL-1 or IL-6 have proven highly successful in treating even the most severe forms of sJIA, including associated MAS.[1,6]

Macrophage Activation Syndrome

The sometimes-fatal complication MAS is most commonly seen in sJIA among rheumatic diseases. Clinically, MAS resembles many features of a sJIA disease flare-up, and it has been suggested that MAS may be inherent to sJIA disease pathology in up to half of all patients with sJIA.[6] MAS is likely part of the spectrum of HLH disorders. Primary HLH, or fHLH, typically presents in infancy following infection and results from homozygous mutations in genes involved in the cytolytic pathway employed by NK cells and CD8 T cells. Recent evidence suggests that patients with sJIA who have MAS have heterozygous defects in these same cytolytic pathway genes.[20] MAS can be triggered by a variety of infectious organisms, particularly

KEY CONCEPTS

Macrophage Activation Syndrome

- Macrophage activation syndrome (MAS) is present in up to 50% of children with systemic juvenile idiopathic arthritis (sJIA) in a subclinical (hemophagocytosis) or overt (systemic inflammatory response, 10%) form.
- MAS manifests as fever, liver dysfunction, pancytopenia, central nervous system disturbance, hyperferritinemia, hemophagocytosis, and coagulopathy.
- MAS resembles hemophagocytic lymphohistiocytosis (HLH) and is thought to result from defects in perforin-mediated cytolysis by CD8 T cells and natural killer (NK) cells.
- Patients with sJIA and MAS have been noted to have NK-cell defects and mutations in perforin-1 and MUNC13-4 cytolytic pathway genes.
- MAS can be fatal if not recognized and treated early. Mainstays of therapy include high-dose corticosteroids and cyclosporine.
- Recently, blocking interleukin-1 (IL-1) signaling with biological therapies has been found to be quickly and dramatically beneficial in treating MAS associated with sJIA.

members of the herpes virus family, but the precise role of infectious triggers of MAS in children with sJIA remains unknown. Nevertheless, the inability to effectively shut down an immune response via cytolytic mechanisms results in a "storm" of proinflammatory cytokines, such as IL-1, IL-6, IL-18, TNF, and IFN-γ. Mouse models of MAS/HLH have suggested that IFN-γ is the pivotal cytokine in MAS. In practical terms, inhibition of IL-1, and potentially IL-6, has proven rather effective at treating MAS in children with sJIA.[1,6,20] It is quite remarkable how dampening of one critical cytokine can help restore the immune imbalance of multiple proinflammatory cytokines and rapidly reverse the life-threatening clinical scenario of MAS.

JUVENILE IDIOPATHIC ARTHRITIS CLINICAL SUBTYPES

There is heterogeneity of clinical presentation and progression of the various subtypes, which is attempted to be addressed by the current ILAR classification schema (see Table 54.1), but reclassification efforts are ongoing. An elegant report by Eng et al.[21] has suggested five distinct groups of patients categorized based on clinical disease trajectories, all with subsets different from those defined by the ILAR classification. Nevertheless, the current ILAR classification is the most widely accepted for the time being.

Oligoarticular Juvenile Idiopathic Arthritis

Oligoarticular JIA is likely the most common category of JIA with involvement of one to four joints, most commonly knees, ankles, the TMJ, and fingers.[1] The majority of these children are girls of preschool-age, blonde-haired, blue-eyed with morning stiffness and swollen joints, which can be painless in 25% of cases. Thus, by the time abnormalities are noted by caregivers, the child may already have developed contractures, bony hypertrophy, and limb length discrepancy as chronic articular inflammation stimulates the osteoblasts of the nearby growth plates, resulting in temporary acceleration of growth. Decreased muscle group recruitment around the affected joint leads to muscle wasting, and can affect joint function, gait, and mobility for years to come. The oligoarticular JIA category has the highest percentage of positive ANA blood tests and is associated with potentially damaging silent/painless uveitis.[16] While involvement of a single joint is common in JIA, careful differential diagnosis consideration is necessary. Oligoarticular JIA rarely presents with isolated hip involvement on the contrary to toxic synovitis, septic hip, and malignancies. Monoarticular involvement in middle and high school aged children is more commonly associated with reactive arthritis, IBD-related arthritis, and Lyme disease. The ILAR classification has a separate subcategory for children with oligoarticular JIA who develop additional joint involvement after the first 6 months based on clinical presentation, extended oligoarticular JIA,[1] which might be a variant of the RF-negative polyarticular category. Wrist involvement is considered to be a bad prognostic factor, as is the extended oligoarticular phenotype and elevated laboratory indicators of inflammation.

Polyarticular Juvenile Idiopathic Arthritis

Inflammation of greater than four joints appear in two main JIA categories (see Table 54.1) based on the presence of serum RF, an IgM antibody against the IgG Fc receptor. RF-positive serum on two occasions at least 3 months apart is required for a child to be diagnosed with *RF⁺ polyarticular JIA*. Joint involvement is typically bilateral, symmetrical, involving the small joints of hands and feet.[1] However, large joints and cervical involvement are often present. RF⁺ polyarticular JIA usually presents in adolescent girls and is considered a form of early-onset adult RA.[1] It is a relatively infrequent category with less than 5% of all JIA patients. Antibodies to CCP are much less common in children than in adults with RA,[17] but just like RF⁺ patients, they also herald a more destructive/erosive disease calling for early aggressive therapy. Similar to its adult counterpart, this category comes with arthritis of the wrists and fingers that can lead to ulnar deviation and boutonniere and swan neck deformities. Destructive TMJ involvement is also common.[3] Extraarticular manifestations, such as low-grade fever and occasionally rheumatoid nodules over bony surfaces seen in adults with RA, are less common in children.

RF-negative polyarticular JIA usually presents as symmetrical involvement of the large and medium-sized joints, mostly knees, wrists, and ankles but also the TMJ.[3] As in psoriatic JIA, small joint involvement tends to occur later in life, and there is a bimodal distribution in preschool children and early adolescent patients.[1] There is more frequent silent uveitis in the younger age group, and this form of JIA is often difficult to distinguish from ERA in the latter.

Psoriatic Arthritis

Arthritis with concurrent psoriasis, or arthritis with two of the three criteria—dactylitis (tenosynovitis causing swelling of the digit beyond the joint capsule), psoriatic nail changes, or family history of a first-degree relative with psoriasis—comprises a separate category of JIA. As mentioned previously, there seems to be a bimodal distribution of age at onset for psoriatic JIA.[22] Preschool-aged children have mostly small joint involvement and positive ANA but may also have dactylitis, whereas middle school–aged patients have JIA that resembles ERA with enthesitis, sacroiliac joint involvement (albeit milder), and even spondylitis.[8] In general, initially there is asymmetrical involvement of the joints, and if untreated, it will progress to polyarticular joint disease. Since up to 50% of the patients develop psoriatic skin findings several years after arthritis presentation, it is often difficult to classify this condition as such at onset. Psoriatic JIA

seems to be more resistant to therapy, and approximately 40% of children have active disease into adulthood while on medications. Insidious onset of anterior uveitis is more typical for the younger age group, whereas the older—those with enthesitis—have JIA that resembles the adult type of psoriatic arthritis with an associated HLA-B27 genotype and chronic symptomatic, often painful, eye disease (see Table 54.1).

Enthesitis-Related Arthritis

In general, ERA affects boys more than girls[1] and may sometimes be a manifestation of IBD. Inflammation can present in both joints and entheses, attachments of the tendons, ligaments, or joint capsules to bone. Individual entheses can be tender even in healthy children, but three or more of them involved at the same time points to pathology (see Table 54.1). ERA often occurs in boys 8 years of age and older.[1,8,23] They report stiffness after inactivity and more pain, including back pain, than limitations in comparison to other JIA categories. Most of the time they maintain their activities, with worsening pain reported towards the end of the day. Many consider ERA a potential prelude to ankylosing spondylitis (Chapter 58), an HLA-B27–associated inflammatory condition resulting in irreversible fusion of the vertebrae and of the sacroiliac joints. Therapeutic efforts targeting TNF and IL-17 pathways are focusing on treatment of early stages of inflammation, thus preventing the calcifying hypercorrection of inflamed vertebral edges leading to the above-mentioned fusion using TNF blockade and maintaining mobility.[23] Outcomes are still under investigation but appear promising with early aggressive treatment.

Systemic Juvenile Idiopathic Arthritis

Approximately 10% of children with JIA belong to the sJIA category, also known previously as *Still disease*. The peak incidence of sJIA is ages 1 to 5 years, but it can present in adulthood as adult-onset Still disease (AOSD). The ILAR classification criteria require fever for 2 weeks, with at least three episodes of daily spiking (quotidian) fever, with at least one of the following features: fleeting salmon-colored macular rash, arthritis, lymphadenopathy, and hepatosplenomegaly.[1] When febrile, children appear rather ill, and the rash is more prominent and can be evoked by mechanical contact (Koebner phenomenon). Typically, the fever subsides, and children are visibly better in the morning hours. High levels of indicators of systemic inflammation are typical at onset and may subside later in the disease. Arthritis is often very aggressive and frequently involves wrists, ankles, and knees but also causes ankylosis of the hip and neck, leading to long-term damage and gait abnormalities. Occasionally, joint involvement begins months after fever onset, making timely sJIA diagnosis more difficult. The initial presentation mimics those of infections and malignancies, which have to be excluded. Fifty percent of children may develop (only 10% clinically overt) MAS, described previously.[6]

Laboratory Evaluation

Diagnosis of JIA is established by history and physical exam, and laboratory indicators have only a supportive role. The complete blood count (CBC) is largely normal in oligoarticular involvement, as is the erythrocyte sedimentation rate (ESR). White blood cells (WBCs) are highly elevated in sJIA but are mostly within normal limits in other groups. Anemia of chronic disease presents as normocytic and normochromic and is often found in polyarticular involvement. In those cases, ESR can also be elevated. Intermittent joint effusion of a single large joint with an elevated ESR necessitates further evaluation, and IBD should be considered, especially if there is a low serum albumin level and/or growth delay. Of note, elevated ESR on presentation predicts a worse outcome for those with the oligoarticular subtype. The platelet count, as a marker of inflammation, can be elevated in polyarticular disease and substantially so in sJIA without MAS.

Liver function tests are used for monitoring certain disease-modifying anti-rheumatic drugs (DMARDs), such as methotrexate and leflunomide. They can be elevated as a result of prolonged, and frequently concomitant, use of nonsteroidal antiinflammatory drugs (NSAIDs).

As mentioned above, in 10% of patients, sJIA can progress to overt MAS. Ferritin, an acute-phase reactant, is a very sensitive indicator of this condition. A sudden drop of at least two cell lines in the CBC, a rising C-reactive protein (CRP) level with decreasing ESR, elevated liver transaminases, prolonged prothrombin or partial thromboplastin times, high D-dimer levels, elevated triglycerides, and low fibrinogen should all alert caregivers about the likelihood of MAS in a child with sJIA.[6]

Although 75% to 85% of adult patients with RA have either a RF or CCP antibodies, often both, less than 5% of patients with JIA have the RF, and those are mostly patients with early-onset RA, often teenage girls with symmetrical small joint involvement. Serum anti-nuclear antibody, while widely used, is not suitable for screening as it is of no diagnostic utility in either making or excluding a diagnosis of JIA. ANA serves as a prognostic factor by identifying patients already diagnosed with JIA who have the highest risk for developing uveitis.[3,16] In addition, ANA levels may alert the clinician to the possibility of juvenile Sjögren disease or SLE being the etiology for the chronic arthritis.

The prevalence of the HLA-B27 antigen is 8% in the general Caucasian population but nearly 90% in the ankylosing spondylitis group.[8] HLA-B27 is useful to predict axial involvement in ERA, psoriatic JIA, and IBD-related arthritis but should be evaluated in patients with inflammatory back pain, clinically established arthritis, and/or enthesitis rather than as a routine screening test during a work-up for any back pain without morning stiffness.

Additional laboratory indicators, such as elevated serum lactate dehydrogenase and uric acid levels, may indicate malignancy. Elevated angiotensin converting enzyme (ACE) and lysozyme levels are useful when considering sarcoidosis as the etiology of childhood arthritis and uveitis in early-onset sarcoidosis cases, but non-caseating granulomas seen on histology are more diagnostic.

Imaging Evaluation

Radiography can help evaluate persistent joint pain for bony disease processes (e.g., osteochondral lesions or malignancies). Chronic subtle clinical arthritis can result in finding periarticular osteopenia. Bony erosions seen in adult RA patients or those with psoriatic arthritis are less common in children. Radiographic findings of sacroiliac joint involvement need confirmation by magnetic resonance imaging (MRI) without intravenous contrast material. The role of ultrasonography for diagnosis and monitoring of disease progression in pediatric patients is promising, but further establishment of normative data is still required. MRI, with and without intravenous contrast, may help identify synovitis, but even this imaging modality can some-

times yield false-positive results in certain joints in otherwise healthy children.[8,24,25] MRI with contrast is the gold standard for diagnosing TMJ arthritis in JIA.[3]

DIFFERENTIAL DIAGNOSIS

A wide variety of conditions can mimic the symptoms and signs of JIA. A single acutely involved joint at onset should be considered of infectious origin until proven otherwise and calls for arthrocentesis. Septic arthritis frequently has an erythematous discoloration of skin, whereas JIA does not. Infection of the joint may lead not only to rapid destruction of the joint but also to systemic dissemination of infection. Specifically, *Kingella* species infection as a differential diagnosis for oligoarticular JIA should be considered in painful monoarticular cases of young children when symptoms have sudden onset without prolonged morning stiffness.

Parainfectious arthritis, often resulting from viral disease, is usually short-lived and typically requires only NSAID treatment. In contrast, Lyme disease is characterized by a chronic extensively swollen joint(s), commonly the knee.[15] It appears several weeks to months after the usually unnoticed tick bite and should be considered in areas of high incidence (e.g., northeastern United States). Migratory arthritis is associated with malignancies, *Neisseria* sp. infection, ANCA-associated vasculitis, and rheumatic fever—the last having typically very tender, red, hot joint involvement persisting in each location for a couple of hours before moving to another.

When evaluating for sJIA, multiple other etiologies of systemic inflammation should be considered. Fever and elevated WBCs, platelets, and ESR may accompany polyarteritis nodosa, Kawasaki disease, Henoch-Schönlein purpura, and other vasculitides. Ehrlichiosis and recurrent fever syndromes (e.g., familial Mediterranean fever, TNF receptor–associated periodic fever syndrome) may also manifest as systemic inflammation, arthralgias, and rashes. A typically nonerosive symmetric polyarthritis may occur in SLE, often with cytopenias. Polyarthritis is also seen in SLE-related diseases (e.g., SS, MCTD). Evidence of specific antibodies to extractable nuclear antigens in SLE (anti-Smith, anti–double-stranded DNA [dsDNA]) and MCTD (anti-ribonuclear protein [RNP]) help distinguish these cases from JIA.

Patients presenting with joint pain only, especially after activities or toward the end of the day, without associated swelling or morning stiffness, most commonly have overuse/overload syndromes with or without benign hypermobility syndrome. Structural bony involvement resulting in joint pain occurs with little league shoulder and elbow in middle school–aged baseball players; wrist pain occurs in gymnasts frequently. Pain in the hip, without morning stiffness, may indicate Legg-Calvé-Perthes disease or slipped capital femoral epiphysis. Relentless knee pain without improvement in teenagers could be caused by osteochondritis dissecans. Other common causes of teenage knee pain include Osgood-Schlatter disease and patellofemoral syndrome.

Bone malignancies present with constant pain in the corresponding area of the skeleton. Hematological malignancies typically cause referred hip pain, as does neuroblastoma. Nighttime skeletal pain, back pain in little children, migratory character, and systemic symptoms such as fever, weight loss, and metaphyseal tenderness all should raise suspicion for oncological process. On the other end of the spectrum, unrelenting, diffuse body pain with minimal clinical findings and negative laboratory and imaging findings should be evaluated for pain amplification syndrome.

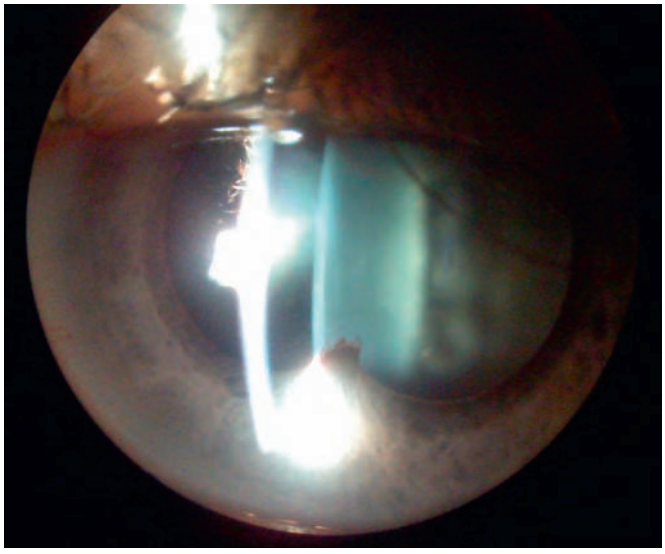


FIG. 54.2 Posterior Synechia, a Complication of Chronic Anterior Uveitis, is Associated with Several Categories of Juvenile Idiopathic Arthritis. The irregularities of the inner margins of the iris reflect fibrous adhesions between the iris and lens capsule. (Courtesy of Dr. Scott Olitsky.)

"Growing pains" or benign nocturnal limb pains of childhood is one of the most common misconceptions when evaluating joint pain. Typical "growing pains," while not associated with growth per se present at night, waking the child up with excruciating shin or leg pain; it is not joint centered and resolves with nonsteroidal (NSAID) therapy, heat, or massage. A different etiology should be investigated for when other parts of the body are involved. Rapid skeletal growth periods in the child's life can come with musculoskeletal pain, especially that of the back, when potentially delayed muscle tone and mass development cannot keep up with change of biomechanics.

Clinically Silent Complications

There are two common manifestations that present insidiously during the course of JIA and yet can cause major damage/morbidity—uveitis and TMJ arthritis. Uveitis is a relatively common complication of JIA with potentially long-term morbidity. Risk factors for uveitis among patients with JIA include young age of onset, positivity for ANA, oligoarticular disease, and female sex.[16] It usually presents as iridocyclitis, but the choroid may also be affected. Although extremely rare in sJIA, approximately 20% of patients with oligoarticular JIA and 5% of patients with RF-negative polyarticular JIA develop eye inflammation.[16] Some children with psoriatic JIA are also at risk of developing silent uveitis, especially in the ANA-positive subgroup.[22] Uveitis may lead to a great deal of morbidity, including cataracts, increased intraocular pressure, band keratopathy, and posterior synechiae (Fig. 54.2), with decreased vision developing in up to 40%.[16] The danger of most JIA-associated uveitis is its asymptomatic presentation, with the exception of symptomatic uveitis in children with ERA. Considering that the highest prevalence is in patients with oligoarticular JIA, which is most common in preschool and younger children, it is not surprising that many of the cases go unnoticed. A routine vision examination may fail

to detect uveitis, and children need to have a formal slit-lamp examination to identify inflammatory cells.[16] The most common presenting signs are synechiae (an irregular iris border resulting from adhesions to the lens), hypopyon, and band keratopathy (see Fig. 54.2). Uveitis may develop many months or years after joint symptoms, and, therefore, close follow-up is warranted. ANA positivity and young age are associated with increased incidence; thus, ANA-positive children with oligoarticular JIA need to be screened the most often, but those in other categories should be followed up on a set schedule as well.[16] Failure to do so and missed diagnosis of eye involvement may lead to the occurrence of cataracts, glaucoma, impaired vision, and even blindness.

KEY CONCEPTS

Temporomandibular Joint Arthritis

- Temporomandibular joint (TMJ) arthritis is common, present in up to 80% of children with juvenile idiopathic arthritis (JIA). It is typically asymptomatic at onset and thus requires early screening with magnetic resonance imaging (MRI) with contrast.
- TMJ arthritis can be active despite therapy with disease-modifying anti-rheumatic drugs (DMARDs) and biologicals (e.g., methotrexate plus tumor necrosis factor [TNF] inhibitors) and might require intensification of treatment.

Another frequently asymptomatic complication of JIA is TMJ arthritis. TMJ arthritis in children with JIA has been recognized increasingly in recent years as a joint inflammation, leading to silent destruction and facial deformity despite systemic therapy.[3] TMJ arthritis is quite common, with 40% to 80% of all patients with JIA patients affected.[3] The overall true prevalence is likely closer to the higher range, since not all children with JIA receive TMJ MRI screening, which frequently reveals synovial thickening (Fig. 54.3) at disease onset, and premicrognathic arthritis may be missed.[3,22] The highest rates of TMJ arthritis have been found in the extended oligoarticular JIA and RF-negative polyarticular JIA groups, as well as in children with upper extremity and neck involvement and those with an elevated ESR.[3] TMJ arthritis needs to be recognized early in children with JIA so that it can be treated prior to growth disturbance.

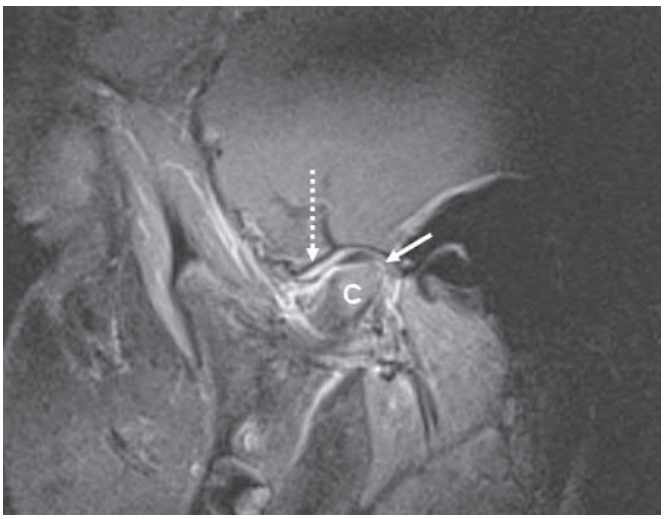


FIG. 54.3 Acute and Chronic Temporomandibular Joint Arthritis in a Child with Juvenile Idiopathic Arthritis. Synovial thickening and enhancement *(long dashed arrow)* and mandibular condyle *(C)* flattening with contour irregularity/erosion *(short arrow)* are noted in this parasagittal postcontrast T1-weighted magnetic resonance image. (Courtesy of Dr. Dan Young.)

TREATMENT

Overview

Despite significant advances in the understanding of the pathogenesis of JIA, there are currently no curative treatments. JIA frequently persists into adulthood and may result in significant morbidity, including physical disability. The objective of treatment is to prevent disability and preserve normal growth and development while providing relief of symptoms and improved quality of life by controlling the inflammatory process.[1]

Over the past 20 years, remarkable advances have been made in the treatment of JIA.[1] Chief among these advances was the advent of targeted biological therapeutic agents (Table 54.2) (Chapter 86). These agents have been shown to be quite beneficial against active disease and are generally well tolerated.[1] The early initiation of biological therapeutic agents may, in fact, alter and improve the subsequent disease course.[1] These new breakthroughs have prompted pediatric rheumatologists to "invert the treatment pyramid": that is, to rapidly incorporate more effective therapeutic agents instead of slowly progressing to them in a stepwise fashion.[1]

THERAPEUTIC PRINCIPLES

Early Aggressive Therapy

- Accumulating evidence suggests that early aggressive therapy that includes targeted biological agents near the time of clinical diagnosis (during the "window of opportunity") may improve the future disease course.

In response to the growing number of treatment options for JIA and the advent of the biological therapeutics, the American College of Rheumatology (ACR) issued Recommendations for the Treatment of JIA that were updated in 2019.[26,27] These recommendations were developed by using a rigorous methodology to produce evidence- and consensus-based guidance that reflected the current state of the field.

With new and effective therapies continuing to be introduced to the therapeutic armamentarium, treatment goals have become elevated and more stringent. The current goal is to achieve a status of clinically inactive disease[28]: that is, the absence of any significant signs or symptoms of active arthritis.

Recent advances in the treatment of adults with inflammatory arthritis have elucidated some differences in the effectiveness of specific biological agents for specific forms of arthritis. For example, many non-TNF inhibitor biologicals, such as abatacept, rituximab, and tocilizumab, are highly effective in the treatment of RA but are far less effective in the treatment of ankylosing spondylitis. In contrast, some of the more recently introduced biological agents, such as the IL-17 inhibitor secukinumab, appear effective against ankylosing spondylitis but are likely less effective for RA.[29]

Despite these recent advances, the treatment of JIA is not yet strongly influenced by the distinct categories of JIA, with the exception of sJIA. For example, there are no specific therapies currently approved for the treatment of children with psoriatic

TABLE 54.2 Biological Therapeutic Agents Used in the Treatment of Juvenile Idiopathic Arthritis

Biological Target	Name	Structure	Frequency of Administration	Route of Administration
Tumor necrosis factor (TNF)	Etanercept	TNF receptor–immunoglobulin G (IgG) fusion protein	Twice weekly to weekly	Subcutaneous (SQ) injection
	Infliximab	Monoclonal antibody (chimeric)	Every 4–8 weeks	Intravenous (IV) infusion
	Adalimumab	Monoclonal antibody (humanized)	Every 1–2 weeks	SQ injection
	Golimumab	Monoclonal antibody (humanized)	Every 4 weeks Every 8 weeks	SQ injection IV infusion
	Certolizumab pegol	Monoclonal antibody (humanized and PEGylated)	Every 2 weeks	SQ injection
CD80/86	Abatacept	Cytotoxic T-lymphocyte antigen-4 (CTLA-4–IgG fusion protein	Every 4 weeks Weekly	IV infusion SQ injection
Interleukin-1 (IL-1)	Anakinra	Receptor antagonist	Daily	SQ injection
	Canakinumab	Monoclonal antibody (humanized)	Every 8 weeks	SQ injection
	Rilonacept	Receptor-fusion protein	Weekly	SQ injection
IL-6 receptor	Tocilizumab	Monoclonal antibody (humanized)	Every 2–4 weeks Every 1–2 weeks	IV infusion SQ injection
$CD20^+$ B cells	Rituximab	Monoclonal antibody	2 infusions 2 weeks apart, repeat every 6 months	IV infusion
IL-12 and IL-23	Ustekinumab	Monoclonal antibody	Every 12 weeks	SQ injection
IL-17A	Secukinumab	Monoclonal antibody	Every 4 weeks	SQ injection

arthritis as opposed to RF-negative polyarthritis. Accordingly, the discussion of treatment in this chapter does not detail all categories of JIA but focuses rather on the "treatment groups" as defined by the 2011 ACR Recommendations for the Treatment of JIA.[30] As our knowledge about pathophysiology and response to treatment continues to expand, differential therapeutic approaches for the disparate JIA phenotypes are eagerly anticipated.

Nonsteroidal Antiinflammatory Drugs

NSAIDs formed the foundation of the treatment of JIA for decades, and numerous NSAIDs have been shown to have beneficial effects. In the absence of significant numbers of head-to-head trials, it is believed that in general all NSAIDs are similarly effective,[1] although indomethacin is considered by some to be the most effective NSAID (especially in sJIA). One NSAID may be found to be more effective than another for a particular child.

In general, NSAIDs are not considered to be disease-modifying agents: that is, they are not felt to slow the progression of disease or prevent the appearance of radiographic damage. For this reason, monotherapy with NSAIDs initially began to decline with the advent of agents such as methotrexate that have been shown to modify the disease process. NSAIDs are frequently used for symptomatic relief, but the ACR Recommendations for the Treatment of JIA strongly recommend treatment with a DMARD over NSAID monotherapy for patients with active polyarthritis.[27] Gastrointestinal (GI) discomfort is a frequent adverse effect of NSAID therapy, although frank GI bleeding appears to occur at a lower incidence than in adults. Scarring pseudoporphyria of sun-exposed skin is another risk associated with NSAID use. The long-term cardiovascular effects of NSAIDs in children have not been studied.

Glucocorticoids

Similar to many rheumatological diseases, JIA has been shown to respond to treatment with glucocorticoids (Chapter 83). Intraarticular glucocorticoid injections typically result in a near-immediate decrease in inflammation that is maintained for many months.[31] Accordingly, intraarticular injections may form the foundation of therapy for children with mild or limited oligoarthritis. In addition, intraarticular injections may be effective in children with more extensive arthritis and in those who are receiving concurrent systemic therapy. Randomized trials have shown unequivocally that injected triamcinolone hexacetonide has the longest duration of effect.[30]

JIA also may be effectively treated with systemic glucocorticoids. They are frequently used in the treatment of systemic features of JIA and may form the foundation of the treatment for the sJIA category, although biological therapies may significantly lessen the need for this practice.[28] The use of systemic glucocorticoids for the treatment of synovitis in children with JIA is not an uncommon practice. However, the risks, benefits, and appropriate use of this approach are less clear. The ACR Recommendations for the Treatment of JIA state that a limited course of systemic glucocorticoids (i.e., so-called "bridging therapy" while awaiting the benefits of newly initiated nonbiological or biological DMARDs) may be acceptable, but the use of chronic systemic glucocorticoids is strongly recommended against.[27] The anticipated adverse effects of long-term use of moderate doses of glucocorticoids includes growth failure, osteoporosis, cataract formation, glaucoma, hyperglycemia, hypertension, avascular necrosis of bone, striae, and others.[31]

Nonbiological Disease-Modifying Anti-Rheumatic Drugs

The use of DMARDs was introduced in the 1980s. The most widely used and studied is methotrexate, which has been shown in randomized clinical trials to be efficacious in treatment of JIA.[1] Following these studies, methotrexate became the cornerstone of therapy for many children with JIA. Methotrexate is typically administered weekly through either the oral or the

subcutaneous route, although studies have shown subcutaneous methotrexate to be better absorbed and more effective.[1]

Methotrexate may be associated with several typically minor adverse effects, such as nausea and fatigue. Occasionally, methotrexate may cause liver toxicity, necessitating periodic measurement of serum aminotransferase levels for routine monitoring.

Leflunomide has been shown to be slightly less efficacious than methotrexate in the treatment of JIA.[1] It may serve as an alternative therapy for children who are intolerant of methotrexate. Sulfasalazine is used, to varying degrees, by pediatric rheumatologists and may be of particular benefit to children with JIA of the ERA category.[1] Hydroxychloroquine monotherapy has been demonstrated inefficacious in treating polyarthritis in JIA.[1]

Biological Disease-Modifying Anti-Rheumatic Drugs

The use of biological DMARDs for the treatment of JIA began in the late 1990s. Etanercept, a TNF inhibitor, was the first studied and was shown to be efficacious in a randomized clinical trial.[1] Additional TNF inhibitors have been introduced and used in the treatment of JIA; adalimumab was also shown to be efficacious in a randomized clinical trial.[1] Notable differences have been discovered regarding the effectiveness of the TNF receptor fusion protein (etanercept) and the mAbs (adalimumab, infliximab, and others). Monoclonal antibody TNF inhibitors have been shown to be effective against two important JIA-associated conditions—anterior uveitis and IBD. Receptor fusion proteins have been shown to be far less effective in treating these conditions. The precise mechanism for these differences in treatment effectiveness is not clear but may be related to the ability of etanercept to bind lymphotoxin or the ability of mAb to bind surface membrane–bound TNF.[32]

Because the TNF inhibitors are large proteins, they must be administered parenterally, either by subcutaneous injection or intravenous infusion (see Table 54.2). TNF inhibitors are not generally associated with common medication adverse effects, such as headache or nausea, although they may result in injection site or infusion reactions. There appears to be a modest increase in the incidence of bacterial infections associated with TNF inhibitor use[33] and a significant risk of reactivation of latent tuberculosis.[32] Therefore, individuals should be screened for tuberculosis infection prior to initiating treatment with TNF inhibitors.[30] A possible association between TNF inhibitors and an increased rate of malignancy in JIA has been proposed, but there is no convincing evidence of a strong increased risk of overall malignancy.[34]

In addition to TNF inhibitors, the T-cell costimulation modulator abatacept has been shown in a randomized clinical trial to be efficacious in the treatment of JIA.[1] Presumably, because of the effectiveness of TNF inhibitors and their earlier availability, the use of abatacept has remained relatively limited in clinical practice.

The B-cell–depleting agent rituximab has been minimally studied in the treatment of JIA.[1] However, it appears effective in some instances, particularly in children who appear to have early-onset RA (teenagers with RF^+ and CCP-positive polyarthritis).

The IL-1 inhibitor anakinra has been shown in both uncontrolled and controlled studies to be particularly effective in the treatment of sJIA.[1,35] Similar to the experience in adults with RA, anakinra appears less effective in treating synovitis among children with the other categories of JIA. Additional IL-1 inhibitors (rilonacept, canakinumab) are also now commercially available and have been shown beneficial in clinical trials treating sJIA.[1]

Unlike other biological agents, the IL-6 inhibitor tocilizumab has been shown in randomized clinical trials to be efficacious in the treatment of both sJIA and polyarticular JIA.[1]

Treatment of Oligoarthritis (Arthritis of ≤4 Joints)

Because of fewer involved joints, oligoarthritis may generally be viewed as a milder form of JIA. However, significant disability can still occur, and children with this condition should not be assumed to have had good clinical outcomes without proper evaluation and therapy. The foundation of treatment for this JIA phenotype is intraarticular glucocorticoid injections. These injections may be administered in multiple joints concurrently and may be repeated, as needed.[31] A good response typically results in resolution of clinical signs and symptoms of arthritis for 4 to 12 months. DMARD therapy should be initiated in children who do not respond as desired to injections or who have more significant arthritis.[30] Methotrexate is typically the agent of first choice. Significant arthritis that does not respond adequately to methotrexate can be treated with TNF inhibitors.[30]

Treatment of Polyarthritis (Arthritis of ≥5 Joints)

Methotrexate is currently recommended for nearly all children with polyarthritis.[27] If a brief trial of methotrexate proves inadequate to control the arthritis, then TNF inhibitors are frequently recommended.[27] If an adequate response to the initial TNF inhibitor is not seen, then switching to a different mechanism of action (e.g., abatacept or tocilizumab) is slightly preferred over switching to a different TNF inhibitor.[27] However, the most appropriate role of non-TNF inhibitor biological DMARDs in the treatment of polyarticular JIA has yet to be clearly defined.

Treatment of Arthritis Involving Specific Joints

Arthritis involving the TMJ, hip, and sacroiliac joints may deserve special therapy. Destructive arthritis of the TMJ among children with JIA has been noted for decades. Clinical evaluation of this joint is particularly challenging, as symptoms are often absent initially, and physical examination findings may be normal. Accordingly, the optimal treatment for TMJ arthritis is unclear. Although TMJ arthritis has been known to demonstrate radiographic progression despite treatment with systemic TNF inhibitors and, in the absence of signs of active synovitis of other joints, increased systemic therapy is likely appropriate.[3,25]

The presence of hip arthritis in JIA has been shown in several studies to portend a poor prognosis.[30] Accordingly, many authors advocate early intraarticular glucocorticoid injections and increased systemic therapy when active hip arthritis is identified.

Sacroiliac arthritis is strongly associated with the development of ankylosing spondylitis. Because axial arthritis has been shown to be less responsive to methotrexate therapy, current recommendations are for the use of TNF inhibitors over methotrexate monotherapy when sacroiliac arthritis is present.[27] It is believed that early treatment with TNF inhibitors, or perhaps even newer biological agents that inhibit IL-17 or IL-12/IL-23, may be optimal for the treatment of spondyloarthritis to help prevent progression of ankylosis in children.[29]

Treatment of Erosive Arthritis

It appears that not all children with JIA have the propensity to develop an erosive arthritis that is similar to that frequently seen

in adults with RA. Children who develop erosions visualized on plain radiographs are considered to have a worse prognosis, and the current recommendation is to increase the intensity of their treatment accordingly.[27]

Treatment of Systemic Features of Systemic Arthritis

For decades, the mainstay of therapy for sJIA has been systemic glucocorticoids and NSAIDs. Nearly all children with sJIA will respond favorably to systemic glucocorticoids, if given in sufficient doses. However, often children become "steroid dependent," and efforts to decrease the glucocorticoid burden to minimize adverse effects have been unsuccessful. Presumably because of its different pathogenesis, sJIA has not been shown to respond to TNF inhibitors as favorably as the other categories of JIA. Instead, IL-1 and IL-6 appear to be key cytokines in the disease process. Accordingly, IL-1 (anakinra, canakinumab) and IL-6 (tocilizumab) inhibitors are recommended as the first line in treatments for sJIA.[35]

The appearance of clinically significant MAS generally requires directed therapy. The typical treatment approach involves increased systemic glucocorticoids. The calcineurin inhibitor cyclosporine is frequently added,[1,6] and some authors advocate IL-1 and IL-6 inhibitors for treatment of MAS.[1,6] In severe refractory cases, cytotoxic chemotherapeutic agents, such as cyclophosphamide or etoposide, may be warranted.

Treatment of Arthritis of Systemic Arthritis

Some children with sJIA will develop a chronic course of polyarthritis with a relative absence of concurrent systemic features. In general, it is recommended that these children be treated as those with polyarthritis who did not have systemic features at onset.[30] Based on clinical trial results, the IL-6 inhibitor tocilizumab may be the most effective treatment for these patients.

Treatment of Uveitis

JIA-associated anterior uveitis frequently requires directed therapy. Topical glucocorticoid eye drops—such as prednisolone acetate 1%—are frequently initiated at the time of diagnosis by the treating ophthalmologist.[16] Although effective in decreasing the inflammation of uveitis, glucocorticoid eye drops cannot be tolerated in high doses for extended periods because of the risk of cataracts and glaucoma.[16] For this reason, systemic medications are frequently employed in the treatment of JIA-associated uveitis. Methotrexate has been shown to be effective for uveitis and is the most commonly used systemic medication.[1,16] The mAb TNF inhibitor adalimumab has been shown to be highly effective in the treatment of anterior uveitis in a randomized clinical trial.[36] Other biological agents (e.g., rituximab, abatacept, tocilizumab) appear to be effective in some children with refractory uveitis, but their overall role remains unclear.

Duration of Therapy

As stated, the current goal of therapy is the attainment of clinically inactive disease. However, once this goal is reached, the appropriate next steps in management are less clear. Although none of the currently available therapies is believed to be curative, many children can successfully decrease or discontinue therapies after attaining inactive disease status without immediate recurrence of active disease. Many pediatric rheumatologists will consider decreasing the level of therapy if inactive disease status is maintained for a prolonged period, such as 12 months, although this approach is arbitrary. Further study of the appropriate management of children who attain prolonged inactive disease status will be an important future focus.[37]

TRANSLATIONAL RESEARCH

ON THE HORIZON

- Novel criteria for identifying macrophage activation syndrome (MAS) among children with systemic juvenile idiopathic arthritis (sJIA) have been developed using real patient data that are diagnostically highly sensitive and specific.
- Genetic screening for mutations in cytolytic pathway genes will help identify those individuals at risk for developing MAS.
- Murine models are paving the way for a better understanding of MAS immunopathology and potential pathways for clinical intervention.
- Clinical trials involving treatment of MAS with inhibitors of proinflammatory cytokines are underway.

The explosion in advances in immunology and genetics is leading to major breakthroughs in therapy for rheumatic diseases, including JIA. Challenges remain, however, in diagnosing and treating MAS complicating sJIA. To distinguish a sJIA disease flare-up from MAS, expert opinion and Delphi techniques are currently being used to explore novel criteria for diagnosing MAS in children with sJIA. Data collection regarding clinical, laboratory, and pathological features of children with sJIA—with and without MAS—have been used to develop new classification criteria.[38,39] Furthermore, genetic mutations and polymorphisms in genes linked to the defective cytolytic pathway in lymphocytes from patients with MAS are being explored to identify patients with sJIA with a propensity to develop MAS.[20] Mouse models of MAS are helping to better understand MAS immunopathology and the role of proinflammatory cytokines in the process.[40] Early recognition of MAS and a better understanding of the role of various cytokines in the pathogenesis of MAS will allow for improved targeted therapy for this often fatal condition.

REFERENCES

1. Stoll ML, Cron RQ. Treatment of juvenile idiopathic arthritis: a revolution in care. *Pediatr Rheumatol Online J*. 2014;12:13.
2. Nigrovic PA, Martinez-Bonet M, Thompson SD. Implications of juvenile idiopathic arthritis genetic risk variants for disease pathogenesis and classification. *Curr Opin Rheumatol*. 2019;31(5):401–410.
3. Stoll ML, Cron RQ. Temporomandibular joint arthritis in juvenile idiopathic arthritis: the last frontier. *Int J Clin Rheumatol*. 2015;10(4):273–286.
4. Hersh AO, Prahalad S. Immunogenetics of juvenile idiopathic arthritis: a comprehensive review. *J Autoimmun*. 2015;64:113–124.
5. Holzinger D, Kessel C, Omenetti A, et al. From bench to bedside and back again: translational research in autoinflammation. *Nat Rev Rheumatol*. 2015;11(10):573–585.
6. Cron RQ, Davi S, Minoia F, et al. Clinical features and correct diagnosis of macrophage activation syndrome. *Expert Rev Clin Immunol*. 2015;11(9):1043–1053.
7. Ombrello MJ, Remmers EF, Tachmazidou I, et al. HLA-DRB1*11 and variants of the MHC class II locus are strong risk factors for systemic juvenile idiopathic arthritis. *Proc Natl Acad Sci USA*. 2015;112(52):15970–15975.
8. Gmuca S, Weiss PF. Juvenile spondyloarthritis. *Curr Opin Rheumatol*. 2015;27(4):364–372.
9. Li YR, Zhao SD, Li J, et al. Genetic sharing and heritability of paediatric age of onset autoimmune diseases. *Nat Commun*. 2015;6:8442.
10. Meyer B, Chavez RA, Munro JE, et al. DNA methylation at IL32 in juvenile idiopathic arthritis. *Sci Rep*. 2015;5:11063.

11. Kessler H, Jiang K, Jarvis JN. Using chromatin architecture to understand the genetics and transcriptomics of juvenile idiopathic arthritis. *Front Immunol*. 2018;9:2964.
12. Zonneveld-Huijssoon E, Albani S, Prakken BJ, et al. Heat shock protein bystander antigens for peptide immunotherapy in autoimmune disease. *Clin Exp Immunol*. 2013;171(1):20–29.
13. Stoll ML. Gut microbes, immunity, and spondyloarthritis. *Clin Immunol*. 2015;159(2):134–142.
14. Arvonen M, Vänni P, Sarangi AN, et al. Microbial orchestra in juvenile idiopathic arthritis: Sounds of disarray? *Immunol Rev*. 2010;294(1):9–26.
15. Arvikar SL, Steere AC. Diagnosis and treatment of Lyme arthritis. *Infect Dis Clin North Am*. 2015;29(2):269–280.
16. Sen ES, Dick AD, Ramanan AV. Uveitis associated with juvenile idiopathic arthritis. *Nat Rev Rheumatol*. 2015;11(6):338–348.
17. Wang Y, Pei F, Wang X, et al. Meta-analysis: diagnostic accuracy of anti-cyclic citrullinated peptide antibody for juvenile idiopathic arthritis. *J Immunol Res*. 2015;2015:915276.
18. Miyara M, Ito Y, Sakaguchi S. TREG-cell therapies for autoimmune rheumatic diseases. *Nat Rev Rheumatol*. 2014;10(9):543–551.
19. Cimaz R, Moretti D, Pagnini I, et al. What do cytokine profiles tell us about subsets of juvenile idiopathic arthritis? *Curr Rheumatol Rep*. 2012;14(2):150–154.
20. Zhang M, Behrens EM, Atkinson TP, et al. Genetic defects in cytolysis in macrophage activation syndrome. *Curr Rheumatol Rep*. 2014;16(9):439.
21. Eng SW, Duong TT, Rosenberg AM, Morris Q, Yeung RS. The biologic basis of clinical heterogeneity in juvenile idiopathic arthritis. *Arthritis Rheumatol*. 2014;66(12):3463–3475.
22. Stoll ML, Punaro M. Psoriatic juvenile idiopathic arthritis: a tale of two subgroups. *Curr Opin Rheumatol*. 2011;23(5):437–443.
23. Dubash S, Bridgewood C, McGonagle D, Marzo-Otega H. The advent of IL-17A blockade in ankylosing spondylitis: secukinumab, ixekizumab and beyond. *Expert Rev Clin Immunol*. 2019;15(2):123–134.
24. Ording Muller LS, Humphries P, Rosendahl K. The joints in juvenile idiopathic arthritis. *Insights Imaging.*. 2015;6(3):275–284.
25. Vaid YN, Dunnavant FD, Royal SA, et al. Imaging of the temporomandibular joint in juvenile idiopathic arthritis. *Arthritis Care Res (Hoboken)*. 2014;66(1):47–54.
26. Angeles-Han ST, Ringold S, Beukelman T, et al. 2019 American College of Rheumatology/Arthritis Foundation Guideline for the Screening, Monitoring, and Treatment of Juvenile Idiopathic Arthritis-Associated Uveitis. *Arthritis Care Res (Hoboken)*. 2019;71(6):703–716.
27. Ringold S, Angeles-Han ST, Beukelman T, et al. 2019 American College of Rheumatology/Arthritis Foundation Guideline for the Treatment of Juvenile Idiopathic Arthritis: Therapeutic Approaches for Non-Systemic Polyarthritis, Sacroiliitis, and Enthesitis. *Arthritis Rheumatol*. 2019;71(6):846–863.
28. Wallace CA, Giannini EH, Huang B, et al. American College of Rheumatology provisional criteria for defining clinical inactive disease in select categories of juvenile idiopathic arthritis. *Arthritis Care Res (Hoboken)*. 2011;63(7):929–936.
29. Braun J, Kiltz U, Heldmann F, et al. Emerging drugs for the treatment of axial and peripheral spondyloarthritis. *Expert Opin Emerg Drugs*. 2015;20(1):1–14.
30. Beukelman T, Patkar NM, Saag KG, et al. 2011 American College of Rheumatology recommendations for the treatment of juvenile idiopathic arthritis: initiation and safety monitoring of therapeutic agents for the treatment of arthritis and systemic features. *Arthritis Care Res (Hoboken)*. 2011;63(4):465–482.
31. Schiappapietra B, Varnier G, Rosina S, et al. Glucocorticoids in juvenile idiopathic arthritis. *Neuroimmunomodulation*. 2015;22(1–2):112–118.
32. Jinesh S. Pharmaceutical aspects of anti-inflammatory TNF-blocking drugs. *Inflammopharmacology*. 2015;23(2–3):71–77.
33. Hurd A, Beukelman T. Infectious complications in juvenile idiopathic arthritis. *Curr Rheumatol Rep*. 2013;15(5):327.
34. Beukelman T, Xie F, Chen L, et al. Risk of malignancy associated with paediatric use of tumour necrosis factor inhibitors. *Ann Rheum Dis*. 2018;77(7):1012–1016.
35. Ringold S, Weiss PF, Beukelman T, et al. 2013 update of the 2011 American College of Rheumatology recommendations for the treatment of juvenile idiopathic arthritis: recommendations for the medical therapy of children with systemic juvenile idiopathic arthritis and tuberculosis screening among children receiving biologic medications. *Arthritis Rheum*. 2013;65(10):2499–2512.
36. Ramanan AV, Dick AD, Jones AP, et al. Adalimumab plus methotrexate for uveitis in juvenile idiopathic arthritis. *N Engl J Med*. 2017;376(17):1637–1646.
37. Lovell DJ, Johnson AL, Huang B, et al. Risk, timing, and predictors of disease flare after discontinuation of anti-tumor necrosis factor therapy in children with polyarticular forms of juvenile idiopathic arthritis with clinically inactive disease. *Arthritis Rheumatol*. 2018;70(9):1508–1518.
38. Henderson LA, Cron RQ. Macrophage activation syndrome and secondary hemophagocytic lymphohistiocytosis in childhood inflammatory disorders: diagnosis and management. *Paediatr Drugs*. 2020;22(1):29–44.
39. Ravelli A, Minoia F, Davı S, et al. 2016 Classification Criteria for Macrophage Activation Syndrome Complicating Systemic Juvenile Idiopathic Arthritis: A European League Against Rheumatism/American College of Rheumatology/Paediatric Rheumatology International Trials Organisation Collaborative Initiative. *Arthritis Rheumatol*. 2016;68(3):566–576.
40. Weaver LK, Behrens EM. Hyperinflammation, rather than hemophagocytosis, is the common link between macrophage activation syndrome and hemophagocytic lymphohistiocytosis. *Curr Opin Rheumatol*. 2014;26(5):562–569.

55

Sjögren's Syndrome

Vasiliki Koulouri, George Fragoulis*, and Clio P. Mavragani*

Sjögren's syndrome (SS) is a chronic autoimmune disease characterized by chronically inflamed exocrine tissues resulting mainly in oral and ocular dryness. Beyond the local manifestations, systemic features affecting virtually any organ system frequently occur. Malignant disease in the form of lymphoproliferation is a well-established disease complication and comorbidities such as accelerated atherosclerosis and mental health issues are also increasingly recognized. SS has been traditionally classified as primary (pSS) and secondary depending on the presence or absence of an underlying autoimmune disease. Given the wide spectrum of clinical manifestations, disease diagnosis can be often challenging. Despite the progress in understanding underlying pathogenetic mechanisms, effective therapeutic strategies remain limited for both local and systemic disease manifestations.

ETIOPATHOGENESIS

Similar to most autoimmune conditions, the exact etiology of SS remains unclear. The impact of environmental, hormonal, and/or stress factors in a genetically predisposed individual is considered the cornerstone of immune dysfunction. Aberrant activation of epithelial cells appears to be a distinctive element in SS.

KEY CONCEPTS

Pathogenesis of Sjögren's Syndrome

Etiopathogenesis	Pathophysiology
• Genetic predisposition • Epigenetic modifications • Viral infection • Endogenous viral elements • Gastrointestinal dysbiosis • Hormonal imbalance • Stress	• Epithelial cell activation • Lymphocytic infiltration • T-cell and B-cell autoreactivity • Autoantibody production—immune complex formation • "Interferon signature"

Environment

The role of viruses as the potential initiators of the immunologic cascade giving rise to SS has long been discussed. Among the offending agents, cytomegalovirus (CMV), Epstein-Barr virus (EBV), human herpes virus-6 (HHV-6), HHV-8, human T-lymphotropic virus-1 (HTLV-1), human immunodeficiency virus (HIV), hepatitis C virus (HCV), Coxsackie viruses, as well as endogenous retroviral elements have all been suspected to be implicated in disease pathogenesis. Activation of type I interferon (IFN) pathways, presence of cross-reactive autoantibodies and viral nucleic acid in patient tissues, in addition to clinical manifestations reminiscent of SS in viral illnesses, have been proposed as indicators of a viral involvement in SS origin.[1,2] Aside from viruses, evidence of irregular gut microbiota composition in SS patients displaying significant associations with disease activity have recently emerged.[3]

Hormones—Stress

High prevalence of SS in women of menopausal age is suggestive of insufficient estrogen production as a potential contributor to disease development. This hypothesis is further supported by evidence of low estrogen and androgen levels in SS patients together with the development of SS-like symptomatology in estrogen-deficient murine models. Moreover, high-stress conditions are considered to be a trigger of SS manifestations in predisposed individuals with ineffective stress management techniques and indolent hypothalamic–pituitary–adrenal axis.[2]

Genetics—Epigenetics

The significant role of genetic predisposition in SS pathogenesis is supported by the well-known presence of familial aggregation and increased prevalence of genetic polymorphisms in SS. Family members of SS patients, especially siblings, display an increased risk for development of SS or other autoimmune diseases compared with controls. Several major histocompatibility complex (MHC) class II gene alleles have been strongly associated with SS susceptibility, most notably the human leukocyte antigen (HLA) HLA-DR and HLA-DQ encoding antigens, with varying haplotypes depending on the studied population/ethnicity. Genome-wide association studies have also revealed non-HLA genetic variants as risk factors for SS. The latter pertain to genes involved in B-cell activation, nuclear factor (NF)-κB–mediated inflammatory and apoptotic processes, as well as in IFN signaling pathways.[4] The most compelling associations include polymorphisms in genes shown in Table 55.1.

Finally, epigenetic alterations, such as DNA methylation, histone modification, and posttranscriptional gene regulation mediated by noncoding ribonucleic acids (RNAs) have also been associated with a potential pathogenetic role in SS. Specifically, analyses conducted in whole blood, peripheral blood mononuclear cells (PBMCs), and salivary gland tissues of SS patients have revealed disease-associated deoxyribonucleic acid (DNA) methylation patterns, such as hypomethylation of IFN-induced genes. Demethylating mechanisms and alteration of methylating enzymes have been shown to be related to increased long interspersed nuclear element 1 (LINE-1) transcripts in salivary gland tissues.[5] Several studies have reported that alterations of microRNA (miRNA) expression, such as miR-146, miR-16, miR-200b-3p, and miR-181a, may impact immune cell regulation and thus contribute to disease pathogenesis.[6]

PATHOPHYSIOLOGY

Lymphocytic periepithelial infiltrates and B-cell hyperactivity are the key features in SS immunopathology and are associated

*equal contribution

TABLE 55.1 Genetic Factors Associated With Sjögren's Syndrome

HLA Gene Alleles Associated With SS Susceptibility	Non-HLA Genetic Variants Associated With pSS Susceptibility	
	Genes	Function
DR2, DR3,DR5, DR9	IRF5	IFN pathway
DQA1 *0501, *0201, *0301,	STAT4	
DQB1*02, *03, *0201,*0301, *0501, *0602	IL-12A	
DRB1 *03, *0301, *1501	PTPN22	
DRB3 *0101	OAS1	
DRw2, DRw3, DRw52, DRw53	BAFF	B-cell survival and proliferation
	BLK	B-cell signaling and differentiation
	CXCR5	B- and T-cell migration
	TNFAIP3	NF-κB inflammation pathway—ubiquitin
	TNIP1	NF-κB inflammation pathway Interaction with TNFAIP3
	IKZF1	Lymphocyte differentiation
	GTF2I	T-cell signaling—immunoglobulin production

BAFF, B cell–activating factor; *BLK*, B-lymphocyte kinase; *GTF2I*, general transcription factor 2I; *IKZF1*, Ikaros family zinc finger protein 1; *IL-12A*, interleukin-12A; *IRF5*, interferon-regulating factor 5; *OAS1*, 2′-5′ oligoadenylate synthetase; *PTPN22*, protein tyrosine phosphatase non-receptor 22; *STAT4*, signal transducer and activator of transcription 4; *TNFAIP3*, tumor necrosis factor (TNF)-alpha–induced protein 3; *TNIP1*, TNFAIP3-interacting protein 1. See also references 5 and 30.

with several glandular and systemic manifestations. Lymphocytes initially form periductal focal aggregates that eventually advance to the whole tissue. The composition of these infiltrates appears to be dependent on lesion severity; CD4 T cells are the main lymphocytic population of mild lesions, while more advanced ones are characterized by B-cell predominance. Imbalance in T helper 1 (Th1)/Th2 cytokine production along with a prominent Th17 lymphocyte component are the central elements of T-cell involvement. Autoreactive B lymphocytes are primarily responsible for disease-specific autoantibody (anti-Ro, anti-La) and rheumatoid factor (RF) production; together with T cells, they contribute to the formation of germinal center–like structures. Innate immunity cells (natural killers, dendritic cells, macrophages) constitute less than 10% of the infiltrates; they contribute to tissue damage and cytokine and chemokine production.

Lymphocytic accumulation around epithelial tissue and extraglandular manifestations due to epithelial damage (i.e., interstitial nephritis, primary biliary cirrhosis, bronchiolitis) are indicative of a substantial role of epithelial cells in SS pathogenesis, to the extent that the term "autoimmune epithelitis" has been suggested.[7] Epithelial cells are actively involved in immune cell stimulation and recruitment, autoantibody production, and perpetuation of inflammation in various manners:

- They supply autoantigens via increased apoptosis or exosome release (i.e., Ro [SSA], La [SSB]).
- They act as antigen-presenting cells
- They produce proinflammatory cytokines and chemokines, such as interleukin (IL)-1, IL-6, IL-8, tumor necrosis factor-α (TNF-α), CXCL13, CXCL21.
- They release B cell–activating factor (BAFF), an important agent in B-cell survival, proliferation, and activation

Antigen-driven autoantibody production generates immune complex formation, responsible for many of the systemic manifestations in SS. Immune complexes are suspected to ultimately trigger type I IFN production by plasmacytoid dendritic cells. Excess IFN production is evident by the increased expression of IFN-inducible genes in both peripheral blood and glandular tissue of SS patients, also known as the "IFN signature." Type I IFNs further stimulate BAFF production and epithelial cell activation, perpetuating a vicious cycle of aberrant immune activity. Prolonged B-cell autoreactivity as a result of chronic antigenic stimulation, immune complex formation, and long-lasting inflammatory processes in affected tissue act as a "prologue" to lymphomagenesis in SS.

Evidence of discordance between the degree of glandular dysfunction and lymphocytic infiltration in both humans and animal models is suggestive of pathogenetic processes beyond the spectrum of tissue inflammation. Ineffective neurotransmission due to anti–muscarinic receptor antibodies, deficient expression of aquaporins, and structural anomalies owing to altered protein concentrations are some of the mechanisms found to affect salivary and lachrymal secretion.[2]

DIAGNOSIS AND CLASSIFICATION

Diagnosis of SS can be achieved through a combination of detailed medical history, meticulous clinical examination, appropriate laboratory work-up, and evaluation of lachrymal and salivary gland involvement (Fig. 55.1).

Medical History and Clinical Examination

SS should be suspected in any patient presenting with persistent sicca symptomatology: dry eyes and/or mouth. Patients usually complain of burning or foreign body sensation in their eyes, difficulty swallowing, poor dental health, and recurring oral candidiasis. The patient should be questioned for previous parotid gland enlargement (PGE) and family history of SS or other systemic autoimmune disease. Extraglandular manifestations include arthralgias, Raynaud's phenomenon, peripheral neuropathy, and vasculitis, commonly presenting as purpura. Chronic fatigue is a rather common and debilitating SS symptom.[8]

Laboratory Work-Up

Laboratory work-up for evaluation of suspected SS should include complete blood count, basic metabolic panel, urinalysis, erythrocyte sedimentation rate (ESR), serum protein electrophoresis (SPEP), antinuclear antibodies (ANAs), RF, cryoglobulins, C3 and C4 levels, anti-Ro, and anti-La autoantibodies, thyroid autoantibodies, and chest x-ray to exclude sarcoidosis. Evaluation for HCV, HIV, and immunoglobulin G4 (IgG4)-related disease (IgG4-RD) should also be performed as they are included in the differential diagnosis of SS. On clinical suspicion of underlying systemic lupus erythematosus (SLE), testing for anti-dsDNA (double-stranded DNA) antibodies should be offered. Other autoimmune diseases with SS manifestations include rheumatoid

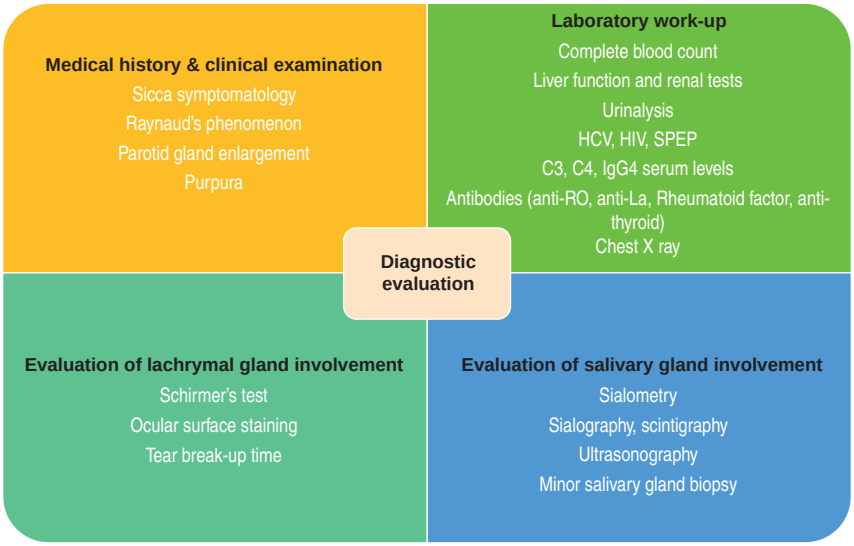


FIG. 55.1 Diagnostic Work-Up in a Patient With Suspected Sjögren's Syndrome. *HCV*, Hepatitis C virus; *HIV*, human immunodeficiency virus; *IgG4*, immunoglobulin G4; *SPEP*, serum protein electrophoresis.

arthritis (RA), primary biliary cirrhosis (PBC), and systemic sclerosis (SSc) (see Clinical Features, later). Common abnormal laboratory values in SS patients are the following:

- Cytopenias. Mainly anemia of chronic disease and leukopenia (white blood cells <4000/μL).
- Increased ESR.
- Hypergammaglobulinemia.
- Positive ANAs. ANAs by indirect immunofluorescence on the classic Hep2 cell line may be negative because Ro60 reactivity can be lost during cell preparation, whereas modified Hep2 cell line (enhanced for better Ro60 reactivity) still lacks adequate sensitivity for this antigen. If positive, the pattern is usually nuclear speckled. Anti-Ro52 positivity produces a cytoplasmic immunofluorescent pattern.[9] If there is strong suspicion of SS, a solid-phase immunoassay for anti-Ro should always be requested.
- RF positivity (approximately in 50% of SS patients) with or without cryoglobulinemia (type II or III).
- Low C4 serum levels.
- Anti-Ro/La positivity. Anti-Ro positivity is present in approximately two-thirds of SS patients. It is included in the classification criteria.
- Mild proteinuria, hyposthenuria, and alkaluria in case of interstitial nephritis/renal tubular acidosis.
- Proteinuria and hematuria in case of glomerulonephritis (GN; rare in SS—associated with vasculitis).[8,10]

Evaluation of Lachrymal Gland Involvement—Ocular Dryness

Ophthalmologic examination of dry eyes is carried out with three main tests, two of which are also items of the 2016 classification criteria for pSS:

- *Schirmer's test*: paper strips are placed in the pouch of both inferior eyelids for 5 minutes. Normally, both eyes produce enough tears to moisten more than 15 mm of paper; 5 mm or less of paper moisture in either eye is a sign of severe objective dryness and raises the possibility of SS.
- *Ocular surface staining*: it is performed with a slit lamp examination using fluorescein and lissamine green dye to assess conjunctival damage indicative of ocular dryness.
- *Tear break-up time (TBUT)*: it is not included in the SS classification criteria. Fluorescein dye is placed on the patient's eye, and the time it takes for the first dry spot to appear is measured. A TBUT less than 10 seconds is considered abnormal.[11]

Evaluation of Salivary Gland Involvement—Xerostomia/Parotid Gland Enlargement

- *Sialometry* is implemented to objectively measure the amount of saliva produced by an individual. A flow rate of 0.1 mL/min or less of unstimulated whole saliva (UWS) production constitutes a positive classification criterion for pSS.
- *Sialography* and *scintigraphy* are used to evaluate the morphology and function of salivary glands, respectively. Both of these methods tend to be replaced by less-invasive techniques.
- *Ultrasonography (US)* is a very useful tool in the assessment of salivary gland pathology in SS patients, and great efforts have been made towards the development of a universal US scoring system. Furthermore, recent evidence supports association of morphologic US abnormalities with systemic manifestations and immunologic markers of SS considered to be prognostic markers of progression to lymphoma.[12]
- *Elastography*, a modality that uses either US or magnetic resonance imaging (MRI) to evaluate the elastic properties of the affected tissue, can improve US specificity of SS diagnosis.[13]
- *Minor salivary gland biopsy (MSGB)* remains an invaluable diagnostic and prognostic tool in SS evaluation and management, although not a prerequisite for SS diagnosis (Fig. 55.2). Glandular lymphocytic infiltration is measured with the use of Focus score (FS) and tissue architectural disorganization is estimated with the Tarpley score. Positive FS is the presence of greater than or equal to 50 lymphocytes/4 mm^2 of tissue and is indicative of SS. Another notable histopathologic feature is the presence of germinal center–like structures, although their association with increased risk for lymphoma development is still a matter of controversy. Still, it is advisable that a patient with a high FS and Tarpley score

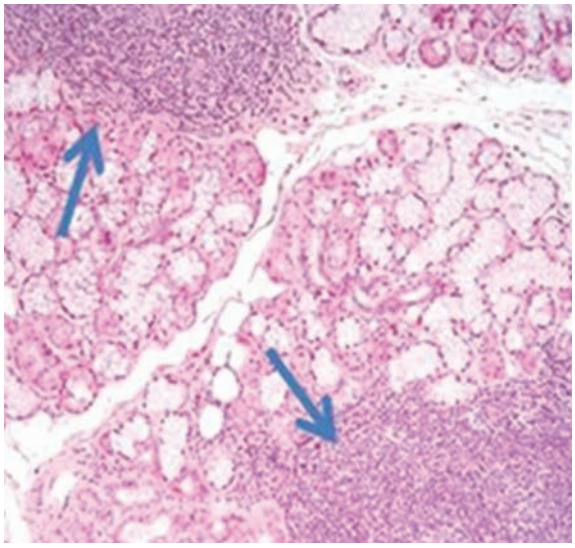


FIG. 55.2 Periepithelial mononuclear cell infiltrates *(arrows)* in labial minor salivary gland biopsy in a patient with Sjögren syndrome.

and germinal center–like structures in their MSGB be closely monitored and further evaluated for lymphoma development. A recent study revealed that SS patients with an FS of zero (essentially lacking histopathologic SS diagnosis) were less likely to have increased expression of IFN-induced genes, high IgG levels, anti-La antibodies or excessive conjunctival damage compared with those with positive FS. However, the two groups did not differ significantly in manifestation of either sicca or constitutional symptoms, asserting the significance of the immunologic criteria in SS diagnosis.[14]

Classification Criteria for Primary Sjögren Syndrome

The 2016 classification criteria for pSS (Table 55.2) were developed and published by the American College of Rheumatology (ACR) and the European League Against Rheumatism (EULAR), with a reported sensitivity of 96% and specificity of 95%. Even though they are intended as a classification tool, they are also widely used in diagnosis.[15]

CLINICAL FEATURES

SS usually displays a benign course, with nonspecific manifestations preceding the full development of the syndrome by approximately a decade. The main SS-related symptoms are dry eyes and dry mouth, reflecting lachrymal and salivary gland inflammation, respectively. Dryness of other tissues (e.g., dry skin, dry vagina leading to dyspareunia) is also commonly observed. Approximately one-third of SS patients exhibit extraglandular manifestations. These are subdivided to those linked with periepithelial mononuclear cell infiltrates (e.g., liver involvement, interstitial nephritis) and those in which vasculitis is the underlying pathogenetic mechanism (glomerulonephritis, peripheral neuritis, purpura) (Table 55.3). Importantly, 5% to 10% of SS patients develop lymphoproliferative disease (LPD), which is combined with increased mortality.[16,17]

Glandular Manifestations

As mentioned earlier, dryness of mouth and dryness of eyes are the most characteristic symptoms of SS. Ocular dryness can be expressed with many different symptoms such as blurred vision, burning, itching, or gritty sensation in the eyes. Oral dryness can present as difficulty in chewing or swallowing, abnormalities of taste and smell, or adherence of food to the buccal surfaces. In

CLINICAL MANIFESTATIONS OF SJÖGREN'S SYNDROME

- Glandular
  - Salivary glands
  - Lacrimal glands
  - Other exocrine glands (xerotrachea, bronchitis sicca, pancreatic insufficiency, dyspareunia)
- Extraglandular
  - Nonspecific (fatigue, arthralgias, Raynaud phenomenon)
  - Periepithelial (liver, lung, renal involvement [interstitial nephritis])
  - Immunocomplexes-associated disease (purpura, peripheral neuritis)
  - Lymphoproliferative disease

TABLE 55.2 2016 ACR/EULAR Classification Criteria for Primary Sjögren's Syndrome

2016 ACR/EULAR Classification Criteria for Primary Sjögren's Syndrome (pSS)

Inclusion Criteria: at least one positive answer — **To determine whether an individual has symptoms and/or signs suggestive of pSS**

1. Have you had daily, persistent, troublesome dry eyes for more than 3 months?
2. Have you had a daily feeling of dry mouth for more than 3 months?
3. Do you have a recurrent sensation of sand or gravel in the eyes?
4. Do you have to wake up at night to drink water because your mouth is so dry?
5. Do you use tear substitutes more than three times a day?
6. Do you frequently drink liquids to aid in swallowing dry food?

Exclusion Criteria: any positive answer — **Any of these conditions automatically excludes the possibility of pSS**

1. History of head and neck radiation treatment
2. Active hepatitis C infection (confirmed by PCR)
3. AIDS
4. Amyloidosis
5. Sarcoidosis
6. Graft-versus-host disease
7. IgG4-related disease

Primary Sjögren's Syndrome Criteria: A Total Score of ≥ 4

1. Labial salivary gland with focal lymphocytic sialadenitis and focus score of ≥ 1 foci/4 mm²	**3**	The histopathologic examination should be performed by a **pathologist experienced in the diagnosis of focal lymphocytic sialadenitis and focus score count**, using the **protocol described by Daniels et al**.
2. Anti-Ro/SSA positivity	**3**	
3. Ocular staining score ≥ 5 (or van Bijsterveld score ≥4) in at least one eye	**1**	Patients who are normally taking **anti-cholinergic** drugs should be evaluated for objective signs of salivary hypofunction and ocular dryness after a **sufficient interval without these medications** for these components to be a valid measure of oral and ocular dryness.
4. Schirmer test ≤ 5 mm/min in at least one eye	**1**	
5. Unstimulated whole saliva flow rate ≤ 0.1 mL/min	**1**	

ACR, American College of Rheumatology; *AIDS*, acquired immunodeficiency syndrome; *EULAR*, European League Against Rheumatism; *IgG4*, immunoglobulin G4. *PCR*, polymerase chain reaction; *SSA*, sjogren syndrome related antigen A

TABLE 55.3 Major Clinical Manifestations of Sjögren's Syndrome.

Major Clinical Manifestations of Sjögren's Syndrome (Frequency %)
1. Arthritis/arthralgia (75%)
2. Parotid gland enlargement (50%)
3. Raynaud's phenomenon (30%–40%)
4. Pulmonary involvement (20%)
5. Purpura (10%)
6. Renal involvement (usually interstitial nephritis) (10%)
7. Liver involvement (5%–10%)
8. Lymphoproliferative disease (usually mucosa-associated lymphoid tissue lymphoma) (5%–10%)
9. Peripheral neuropathy (2%–10%)
10. Central nervous system involvement (2%–10%)

Apart from dry eyes and dry mouth, a wide range of clinical symptoms can be encountered in sjögren's syndrome

addition, oral infections, mostly candidiasis, and dental caries are frequent. Various conditions are included in the differential diagnosis of sicca symptomatology and should be thoroughly excluded. These include head and neck radiation, drugs (antidepressants, parasympatholytics, neuroleptics), metabolic conditions (malnutrition, alcohol abuse, diabetes mellitus, lipoproteinemia), viral infections (HIV, HCV, HTLV), graft-versus-host disease, sarcoidosis, IgG4-RD, and other autoimmune conditions (e.g., autoimmune thyroid disease, PBC) (Table 55.4). Salivary gland (usually parotid) enlargement is observed in approximately 40% of SS patients over the course of the disease and in approximately 15% as a first symptom (Fig. 55.3). PGE is usually intermittent, bilateral, firm to palpation, and asymptomatic. It should be considered as an adverse prognostic factor for lymphoma. Furthermore, a rapid increase in the size of the enlarged parotid glands suggests a superimposed infection or development of lympho-proliferative disorder. The differential diagnosis of bilateral PGE includes diabetes mellitus, lipoproteinemia type IV and V, alcoholism, malnutrition, infection (HIV, HCV, mumps), sarcoidosis, and IgG4-RD. Unilateral PGE can be seen in salivary gland neoplasms (e.g., Warthin tumor), lymphomas, bacterial infections, and salivary duct obstruction (e.g., sialolithiasis) (Table 55.5).

Other glandular manifestations include dry skin, dry nasal mucosa, and trachea (xerotrachea) as well as dry vagina in premenopausal women, observed in approximately 10%, 20%, and 40% of the patients, respectively.[17]

TABLE 55.4 Differential Diagnosis of Dry Eyes and Dry Mouth Symptomatology

Differential diagnosis of sicca symptomatology
Radiation head and neck
Drugs
Antidepressants
Parasympatholytics
Neuroleptics
Metabolic
Malnutrition
Increased alcohol consumption
Diabetes mellitus
Lipoproteinemias
Viral
Human immunodeficiency virus
Hepatitis C virus
Human T-lymphotropic virus
Graft-versus-host disease
Sarcoidosis
IgG4-related disease
Other autoimmune conditions (RA, SLE, PBC, autoimmune thyroid disease, etc.)

IgG4, Immunoglobulin G4; *PBC*, primary biliary cirrhosis; *RA*, rheumatoid arthritis; *SLE*, systemic lupus erythematosus.

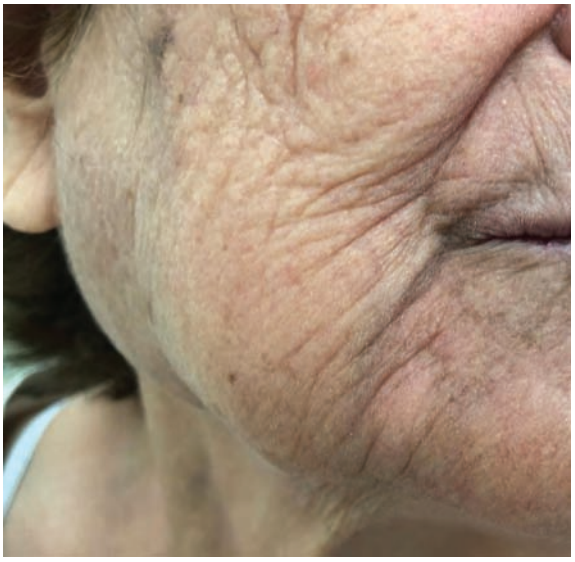


FIG. 55.3 Parotid Gland Enlargement in a Patient With Sjögren's Syndrome.

Extraglandular Manifestations

Musculoskeletal

Musculoskeletal manifestations are quite common in SS, affecting more than half of the patients. They usually present in the form of muscle aches and pain in the small joints, without causing erosive arthritis as seen in RA. Fatigue associated with functional disability is a common and devastating symptom in SS, present in up to 70% of the patients.

Raynaud's Phenomenon

Raynaud phenomenon is encountered in 30% to 50% of SS patients. It may precede sicca symptomatology, is of milder expression compared with that seen in other autoimmune diseases, and is associated with increased prevalence of extra-glandular manifestations.

Respiratory Tract Involvement

Manifestations from the respiratory tract are observed in approximately 20% of the patients and include dry cough and more rarely dyspnea. Pleurisy can occur but is more common in SS associated with systematic lupus erythematosus. In chest computed tomography, peribronchial thickening is a common finding, probably related to peribronchial and/or peribronchiolar mononuclear inflammation. In pulmonary function testing, a small airway obstructive pattern is most commonly diagnosed.[17,18] Interstitial lung disease (ILD), most commonly lymphocytic interstitial pneumonia (LIP) can also be observed and is usually of benign course. ILD in the setting of SS is associated with the presence of anti-Ro but may also be encountered in seronegative patients without sicca symptomatology.

TABLE 55.5 Differential Diagnosis of Parotid Gland Enlargement

Usually Bilateral
Diabetes mellitus
Lipoproteinemias (type IV and V)
Alcoholism
Malnutrition
Infections (human immunodeficiency virus, hepatitis C virus, mumps)
Sarcoidosis
Immunoglobulin (Ig)G4-related disease
Usually Unilateral
Warthin tumor
Lymphoma
Bacterial infection
Salivary duct obstruction

Hepatobiliary and Gastrointestinal Manifestations

Liver involvement in SS is not rare. It is usually expressed as PBC, while autoimmune hepatitis and sclerosing cholangitis are uncommon. PBC is encountered in approximately 5% to 10% of patients with SS. Cholestatic enzymes are elevated and antimitochondrial antibodies (AMAs) are usually positive. The natural course of PBC in the context of SS is benign, because the disease remains stable over time. Ursodeoxycholic acid is the anchor treatment. Finally, it is worth mentioning that dry eyes and dry mouth are common manifestations in patients diagnosed with PBC, found in approximately 50% to 70% of the latter. Indeed, PBC and SS share many common histopathologic and immunopathogenetic features, leading some investigators to coin the term "Sjögren's syndrome of the liver" for PBC.[19]

In regard to other gastrointestinal symptomatology, dysphagia and unexplained hoarseness may occur. These relate to esophageal dysmotility, decreased saliva volume, and gastroesophageal reflux, respectively.

Renal Involvement

Renal involvement in SS can be expressed by either tubulointerstitial nephritis or GN. The former is more common and is usually seen in young patients. It manifests with hypokalemia, low urine specific gravity, alkaline urine pH, and nephrocalcinosis. GN in the setting of SS is usually associated with vasculitis, hypocomplementemia, and cryoglobulinemia. Hypertension, mild proteinuria, and hematuria are the most common presenting manifestations. As pertains to histopathologic findings, membranoproliferative and membranous GN are the most common types. In addition, in contrast to the "full-house" pattern encountered in SLE, IgM and complement deposits are usually seen. In contrast to interstitial nephritis, GN is associated with poorer outcomes and survival.[20]

Vasculitis

Vasculitis is observed in approximately 15% of SS patients. Common findings are low serum complement levels and cryoglobulinemia, and it is associated with increased risk for lymphoma development and mortality. The skin (cutaneous vasculitis) is the most commonly affected organ, presenting with palpable purpura and more rarely with urticarial lesions. However, other tissues such as the glomeruli and vasa nervorum can be affected, producing GN and vasculitic peripheral neuropathy, respectively.

Neuropsychiatric Involvement

Peripheral neuropathy in SS is encountered in approximately 2% to 10% of patients, depending on the diagnostic methodology applied. Peripheral neuropathies in SS include (a) pure sensory neuropathies, (b) sensorimotor neuropathies, and (c) other rare types such as demyelinating neuropathy, mononeuritis multiplex, and autonomous neuropathy. Sensory neuropathies are expressed as distal symmetric sensory loss attributed to axonal degeneration of sensory fibers and rarely as sensory ataxia, which is due to loss of proprioceptive large fibers (ganglionopathy). Painful and burning paresthesias are also frequently observed in SS and are attributed to degeneration of cutaneous axons, representing a clinical condition called small fiber neuropathy. Importantly, physical and electrophysiologic examination is usually normal, and skin biopsy is required to make the diagnosis.[21] Sensorimotor neuropathies in SS patients, including axonal sensorimotor polyneuropathy, are associated with adverse prognostic factors for lymphoma development such as palpable purpura, cryoglobulinemia, and low serum complement levels.

The frequency and the type of central nervous system (CNS) involvement in SS are still debatable. Several manifestations such as hemiparesis, sensory deficits, seizures, aseptic meningitis, transverse myelitis, and multiple sclerosis–like lesions have been reported to occur in the context of SS. However, whether these are mechanistically linked to the index disease remains to be defined. Presence of anti-Ro antibodies has been associated with CNS involvement in SS and anti-aquaporin four antibodies have been detected in patients with SLE or SS and transverse myelitis along with optic neuritis.[10] Psychopathologic morbidities such as anxiety, distinct personality traits (e.g., neuroticism, psychoticism, and obsessiveness), and other features such as nocturia are also more commonly observed in SS patients.[22] Interestingly, the presence of certain autoantibodies has been linked to psychopathologic manifestations in SS patients.[10]

Lymphoproliferative Disease

LPD is probably the most severe complication of SS, associated with higher mortality rates.[16] Non-Hodgkin lymphoma (NHL) is the most common LPD seen in SS and is encountered in approximately 5% to 10% of this population. Risk factors for LPD in SS include: PGE, generalized lymphadenopathy, splenomegaly, palpable purpura, peripheral neuropathy and GN, cryoglobulinemia, low serum levels of C3 and/or C4, hypergammaglobulinemia, monoclonal gammopathy, lymphopenia, and presence of germinal centers in salivary gland biopsies (Table 55.6).[23] Thus close follow-up is required for these patients. Histologically, NHLs in the context of SS are usually mucosa-associated lymphoid tissue (MALT) lymphomas followed by nodal marginal zone and diffuse large B-cell lymphomas (DLBCLs). It is usually extranodal and is encountered in the salivary glands. However, other tissues/organs, such as the stomach or the lungs, can also be involved. The course of localized MALT in SS is usually benign and watch-and-wait policy is often followed. However, more aggressive treatment is needed especially for DLBCLs (see Therapy section).

Overlapping Autoimmune Entities and Comorbidities

Sicca symptomatology (dry eyes and dry mouth) is seen in other autoimmune rheumatic diseases (ARDs) as well, such as RA, SLE, and SSc. The term "secondary SS" has been adopted

TABLE 55.6 Prognostic Factors for Lymphoma Development in Sjögren's Syndrome

Clinical
Tongue atrophy
Salivary (usually parotid) gland enlargement
Lymphadenopathy
Splenomegaly
Palpable purpura
Peripheral neuropathy
Glomerulonephritis
Serologic
Cryoglobulinemia
Low levels of serum C3 and/or C4
Hypergammaglobulinemia
Monoclonal gammopathy
Lymphopenia
Histopathologic
Germinal centers in salivary gland biopsy

to describe these clinical situations. However, criticism has been raised whether this is indeed secondary to other ARD or a true overlap. In RA, sicca symptomatology is present in approximately 30% of the patients. Positive findings in MSGB are rare. It seems that the lesion composition in MSGB is different between SS and RA-sicca patients. In the context of SLE, patients with sicca symptoms have distinct clinical, serologic, pathologic, and immunogenetic characteristics that are similar to SS. Similarly, comparable antibody and histopathology profiles as well as frequencies of sicca symptomatology were found between patients with SS and those with dry eyes/dry mouth in the context of SSc.[24] Therefore the term "secondary SS" tends to be abandoned.

Like in other ARDs, patients with SS are currently recognized to have a higher burden of cardiovascular disease. A recent meta-analysis showed that SS patients compared with controls had increased risk for coronary, cerebrovascular, and thromboembolic events as well as for heart failure.[25] Along the same lines, data are accumulating that SS is associated with arterial stiffness and subclinical atherosclerosis, as measured by pulse wave velocity (PWV) and intima-media thickness (IMT).[4,26]

THERAPY

Therapeutic management of SS is complex, especially when severe extraglandular manifestations occur and often requires collaboration with other disciplines and specialties, such as dentists, ophthalmologists, hematologists, and others. Compared with other autoimmune diseases, there are not many randomized controlled trials (RCTs) assessing treatment options for SS. However, the EULAR 2019 recommendations for management of SS have been published. This paper, along with the respective systemic literature review, comprehensively addresses therapeutic options for SS.[27,28]

Treatment in SS aims mainly to alleviate clinical symptomatology and to prevent complications such as dental caries and mouth candidiasis. As outlined later, topical treatment for dry eyes and dry mouth is the first line of treatment, and immunosuppressants/immunomodulators and biologic drugs may be given in patients with active systemic disease.[28]

Glandular Manifestations

Eye lubricants containing either sodium hyaluronate or hydroxypropylmethylcellulose and preservative-free natural tears are used for dry eyes. In severe cases, referral should be made to ophthalmologists. The latter might prescribe cyclosporine drops (0.05%) or more rarely nonsteroidal antiinflammatory drugs (NSAIDs) or glucocorticoid drops. Future studies will define whether topical tacrolimus could also be an option. Immunosuppressive/immunomodulatory treatments are not recommended for ocular dryness.[27,28]

For oral dryness the following is suggested: meticulous oral hygiene, saliva substitutes, and local (sugar-free gum, citrus juice, xylitol, lozenges) and/or systematic (pilocarpine and cevimeline, both are muscarinic M3 receptor agonists) stimulation of salivary secretion. In patients who cannot tolerate or do not respond to muscarinic agonists, choleretic (e.g., anetholtrithione) or mucolytic (e.g., bromhexine, N-acetylcysteine) agents may be considered.[28] Moist heat and NSAIDs are also suggested as a treatment option in SS patients with PGE.[10]

KEY CONCEPTS

Treatment of Sjögren's Syndrome

1. Topical treatment is preferred for ocular and oral dryness
2. For systemic disease: glucocorticoids, immunosuppressants (e.g., azathioprine) and biologic drugs might be used
3. Glucocorticoids: keep at the minimum dose and length of treatment duration
4. No hard evidence in favor of a specific immunosuppressant
5. B-cell targeted treatment for manifestations in which vasculitis is the underlying pathogenetic mechanism and for lymphoproliferative disease
6. For lymphoproliferative disease: stratify according to type and stage

Extraglandular Manifestations

Arthralgias/Arthritis

For arthralgias/arthritis, evidence is limited and in some cases contradictory. Analgesics and short courses of NSAIDs are the first line of treatment. Low doses of glucocorticoids, hydroxychloroquine, and/or methotrexate (0.2 mg/kg/week) can also be administered, although data to support their use are inadequate. In refractory arthritis, treatment with anti-CD20 therapy could be considered.[27,28] In addition, as stressed in the recent EULAR 2019 recommendations, arthritis should be distinguished from noninflammatory pain and/or fibromyalgia. For the latter, exercise and/or antidepressants may be beneficial.[28]

Other Extraglandular Manifestations

No well-designed RTCs exist, but the following recommendations have been made for the management of active disease: start with glucocorticoids. Minimum dose and treatment duration are advised. In case of nonresponse or adverse events, glucocorticoid-sparing agents (e.g., azathioprine, methotrexate) and/or biologic treatment (rituximab) can be added. Treatment of NHL in the context of SS ranges from a watch-and-wait policy for low-grade lymphomas limited in exocrine glands to chemotherapy (mainly with B cell–depleting agents). Radiotherapy might also be considered for specific types of lymphomas.[27,28] Therapeutic decisions are based on the histologic type and the extent of the LPD.

Biologic Treatments

With regards to biologics, treatment with TNF inhibitors have failed in SS, both in terms of sicca symptomatology as well as re-

garding extraglandular symptomatology. Upregulation of the type I IFN/BAFF axis may account for that. So far, rituximab has only provided modest therapeutic benefit, but it appears to be more efficacious in clinical manifestations in which vasculitis is thought to be the main mechanism of action (e.g., purpura, vasculitic neuropathy, cryoglobulinemia) as well as in LPD in the context of SS.

Belimumab, a monoclonal antibody against BAFF, was tested in a phase 2 RCT, showing improvement in disease activity, fatigue, and mucosal dryness symptomatology but no effect on objective measures of oral and ocular dryness. Belimumab response was found to associate with the number of natural killer cells in the blood and in the salivary glands as well as with type I IFN activity.

In phase 2 RCTs, abatacept has been shown to benefit SS patients in terms of improving disease activity scores, histopathology, and salivary flow. Two phase 3 placebo-controlled RCTs (NCT02915159, NCT02067910) are underway to test efficacy and safety of subcutaneous abatacept in SS. Lanalumab, a monoclonal antibody targeting B cells through BAFF receptor blockade and antibody-dependent cellular cytotoxicity, showed some beneficial effects in a recent study. In addition, molecules targeting the CD40/CD40L pathway are in the pipeline. In fact, a monoclonal antibody against CD40 (iscalimab) had favorable results, significantly improving disease activity and fatigue indices. Finally, RSLV-132, which is a RNase1 fused to the Fc region of IgG1, has been tested in a phase 2 RCT, with favorable results in disease activity and fatigue indices in pilot analysis.[29] A phase 3 RCT (NCT03247686) has finished recruitment, and results are awaited.

It is anticipated that in the next few years, data derived from ongoing RCTs will enhance our therapeutic armamentarium and the new knowledge acquired through multi-effort collaborations will allow the design of novel tailored treatment approaches.

ON THE HORIZON

1. Completion of ongoing randomized trials with biologic agents in Sjögren syndrome will hopefully expand our therapeutic armamentarium.
2. Multi-effort harmonization of clinical information along with implementation of high-throughput techniques will allow the identification of distinct pathogenetic mechanisms in different disease phenotypes.
3. Novel biomarkers for early diagnosis, prognostic classification, and response to different treatment modalities will be hopefully available.
4. Design of novel therapeutic trials targeting newly identified pathogenetic pathways will be feasible.

REFERENCES

1. Mavragani CP, Sagalovskiy I, Guo Q, et al. Expression of long interspersed nuclear element 1 retroelements and induction of type I interferon in patients with systemic autoimmune disease. *Arthritis Rheumatol.* 2016;68(11):2686–2696.
2. Mavragani CP, Moutsopoulos HM. Sjögren's syndrome. *Annu Rev Pathol.* 2014;9:273–285.
3. Mandl T, Marsal J, Olsson P, et al. Severe intestinal dysbiosis is prevalent in primary Sjögren's syndrome and is associated with systemic disease activity. *Arthritis Res Ther.* 2017;19(1):237.
4. Nezos A, Evangelopoulos ME, Mavragani CP. Genetic contributors and soluble mediators in prediction of autoimmune comorbidity. *J Autoimmun.* 2019;104:102317.
5. Mavragani CP, Nezos A, Sagalovskiy I, et al. Defective regulation of L1 endogenous retroelements in primary Sjögren's syndrome and systemic lupus erythematosus: role of methylating enzymes. *J Autoimmun.* 2018;88:75–82.
6. Imgenberg-Kreuz J, Sandling JK, Nordmark G. Epigenetic alterations in primary Sjögren's syndrome—an overview. *Clin Immunol.* 2018;196:12–20.
7. Moutsopoulos HM. Sjögren's syndrome: autoimmune epithelitis. *Clin Immunol Immunopathol.* 1994;72(2):162–165.
8. Mariette X, Criswell LA. Primary Sjögren's syndrome. *N Engl J Med.* 2018;378(10):931–939.
9. Pollock W, Toh BH. Routine immunofluorescence detection of Ro/SS-A autoantibody using HEp-2 cells transfected with human 60 kDa Ro/SS-A. *J Clin Pathol.* 1999;52(9):684–687.
10. Mavragani CP, Moutsopoulos HM. Sjögren syndrome. *CMAJ.* 2014;186(15):E579–E586.
11. Beckman KA, Luchs J, Milner MS. Making the diagnosis of Sjögren's syndrome in patients with dry eye. *Clin Ophthalmol.* 2016;10:43–53.
12. Coiffier G, Martel A, Albert JD, et al. Ultrasonographic damages of major salivary glands are associated with cryoglobulinemic vasculitis and lymphoma in primary Sjögren's syndrome: are the ultrasonographic features of the salivary glands new prognostic markers in Sjögren's syndrome? *Ann Rheum Dis.* 2019
13. Cindil E, Oktar SO, Akkan K, et al. Ultrasound elastography in assessment of salivary glands involvement in primary Sjögren's syndrome. *Clin Imaging.* 2018;50:229–234.
14. Sharma R, Chaudhari KS, Kurien BT, et al. Sjögren syndrome without focal lymphocytic infiltration of the salivary glands. *J Rheumatol.* 2020;47(3):394–399.
15. Shiboski CH, Shiboski SC, Seror R, et al. American College of Rheumatology/European League Against Rheumatism Classification Criteria for primary Sjögren's syndrome: a consensus and data-driven methodology involving three international patient cohorts. *Arthritis Rheumatol.* 2017;69(1). 2016(35–45)
16. Ioannidis JP, Vassiliou VA, Moutsopoulos HM. Long-term risk of mortality and lymphoproliferative disease and predictive classification of primary Sjögren's syndrome. *Arthritis Rheum.* 2002;46(3):741–747.
17. Skopouli FN, Dafni U, Ioannidis JP, Moutsopoulos HM. Clinical evolution, and morbidity and mortality of primary Sjögren's syndrome. *Semin Arthritis Rheum.* 2000;29(5):296–304.
18. Kampolis CF, Fragkioudaki S, Mavragani CP, et al. Prevalence and spectrum of symptomatic pulmonary involvement in primary Sjögren's syndrome. *Clin Exp Rheumatol.* 2018;36(suppl 112[3]):94–101.
19. Selmi C, Meroni PL, Gershwin ME. Primary biliary cirrhosis and Sjögren's syndrome: autoimmune epithelitis. *J Autoimmun.* 2012;39(1–2):34–42.
20. Goules AV, Tatouli IP, Moutsopoulos HM, Tzioufas AG. Clinically significant renal involvement in primary Sjögren's syndrome: clinical presentation and outcome. *Arthritis Rheum.* 2013;65(11):2945–2953.
21. Pavlakis PP, Alexopoulos H, Kosmidis ML, et al. Peripheral neuropathies in Sjögren's syndrome: a critical update on clinical features and pathogenetic mechanisms. *J Autoimmun.* 2012;39(1–2):27–33.
22. Theander L, Strombeck B, Mandl T, Theander E. Sleepiness or fatigue? Can we detect treatable causes of tiredness in primary Sjögren's syndrome? *Rheumatology.* 2010;49(6):1177–1183.
23. Skarlis C, Argyriou E, Mavragani CP. Lymphoma in Sjögren's syndrome: predictors and therapeutic options. *Curr Treat Options Rheum.* 2020;6:1–17.
24. Mavragani CP, Moutsopoulos HM. Primary versus secondary Sjögren syndrome: is it time to reconsider these terms? *J Rheumatol.* 2019;46(7):665–666.
25. Beltai A, Barnetche T, Daien C, et al. Cardiovascular morbidity and mortality in primary Sjögren's syndrome: a systematic review and meta-analysis. *Arthritis Care Res.* 2020;72(1):131–139.
26. Yong WC, Sanguankeo A, Upala S. Association between primary Sjögren's syndrome, arterial stiffness, and subclinical atherosclerosis: a systematic review and meta-analysis. *Clin Rheumatol.* 2019;38(2):447–455.
27. Brito-Zeron P, Retamozo S, Kostov B, et al. Efficacy and safety of topical and systemic medications: a systematic literature review informing the EULAR recommendations for the management of Sjögren's syndrome. *RMD Open.* 2019;5(2):e001064.
28. Ramos-Casals M, Brito-Zeron P, Bombardieri S, et al. EULAR recommendations for the management of Sjögren's syndrome with topical and systemic therapies. *Ann Rheum Dis.* 2020;79(1):3–18.
29. Mavragani CP, Moutsopoulos HM. Sjögren's syndrome: old and new therapeutic targets. *J Autoimmun.* 2019:102364.
30. Imgenberg-Kreuz J, Rasmussen A, Sivils K, Nordmark G. Genetics and epigenetics in primary Sjögren's syndrome. *Rheumatology (Oxford).* 2019

56

Scleroderma–Systemic Sclerosis

John Varga and Fredrick M. Wigley

Systemic sclerosis (SSc) is an acquired chronic multisystem disease characterized by autoimmunity and inflammation, widespread functional and structural abnormalities in small blood vessels, and progressive fibrosis of the skin and visceral organs. Multiple cell types and their products interact to mediate the pathogenic processes that underlie the diverse clinical manifestations of SSc.

PREVALENCE AND EPIDEMIOLOGY

SSc is a sporadic disease with worldwide distribution. Incidence estimates in the United States range from 9 to 19 cases per million per year, and prevalence rates range from 28 to 253 cases per million. Applying the revised American College of Rheumatology classification criteria, which are more sensitive for identifying early stage of SSc, the prevalence estimates are expected to be considerably higher.[1] Age, gender, and ancestry are important factors that determine disease susceptibility as well as outcomes.[2] Like other connective tissue diseases, SSc is more prevalent in women, with the most common age of onset in the range of 40 to 60 years. Disease onset tends to occur at a younger age among patients of African ancestry than among whites. Furthermore, African ancestry patients are more likely to have diffuse skin involvement, digital ulcers, pulmonary hypertension (PH), cardiac involvement, and pulmonary fibrosis, and have a worse prognosis.

ETIOLOGY AND PATHOGENESIS

The pathogenesis of SSc involves dynamic interplay among inherited genetic risk factors, environmental exposures, and stable epigenetic modifications. SSc is a polygenic disease and is not inherited in a Mendelian fashion. Disease concordance rates among both monozygotic and dizygotic twins are relatively low (<5%). Nonetheless, family studies show that 1.6% of patients with SSc have a first-degree relative with the disease (relative risk of 13), indicating an important role for genetic background in SSc disease susceptibility. Indeed, certain human leukocyte antigen (HLA) haplotypes show striking associations with distinct SSc-specific autoantibody responses. Candidate genes shown to be associated with SSc include those implicated in interferon (IFN) signaling, T- and B-cell activation, DNA clearance, and innate immunity (Table 56.1). It is remarkable that a majority of these genes are involved in immune regulation, highlighting the potential importance of immune dysregulation in the pathogenesis of SSc.[3]

Environmental Factors

Although the etiology of SSc is unknown, microbial exposures and exposure to environmental and occupational agents, dietary factors, and drugs have been implicated as potential triggering factors. Evidence for a potentially pathogenic role for cytomegalovirus (CMV), Epstein-Barr virus (EBV), and parvovirus B19 infection or reactivation, as well as *Helicobacter pylori* infection, has been presented. Gut-microbial dysbiosis is prominent in SSc patients but its role in pathogenesis is uncertain. Several epidemic outbreaks of apparently novel multisystem illnesses with SSc-like features have been linked to environmental exposures, such as contaminated rapeseed cooking oils in Spain (the toxic oil syndrome) and gadolinium-associated nephrogenic fibrosis and L-tryptophan dietary supplements (eosinophilia–myalgia syndrome) in the United States.[4] The incidence of SSc is increased in males with occupational exposure to silica. Additional occupational exposures linked to increased risk of SSc include polyvinyl chloride, trichloroethylene, organic solvents, and heavy metals. Drugs linked to SSc-like illnesses include bleomycin, taxane, pentazocine, cocaine, and anorexigens associated with pulmonary artery hypertension. The apparent association of SSc with silicone breast implants initially raised alarm, but epidemiological investigations failed to substantiate an increased risk.[5]

Pathology

The hallmark pathological features of SSc are a noninflammatory obliterative microangiopathy in multiple vascular beds associated with fibrosis of the skin and internal organs. In early-stage disease, inflammatory cellular infiltrates may be prominent in many organs. Vascular injury is the earliest and possibly primary event in the pathogenesis of SSc. Patients show widespread vascular lesions characterized by bland intimal proliferation in the small and medium-sized arteries (Fig. 56.1). In late stages, a combination of perivascular adventitial fibrosis and generalized capillary rarefaction are prominent.

Fibrotic changes are widely distributed, but most prominent in skin, lungs, gastrointestinal (GI) tract, heart, tendon sheath, and perifascicular tissue surrounding skeletal muscle. Accumulation of collagen-rich connective tissue composed of fibulins; elastin; proteoglycans; matricellular proteins, such as tenascin-C and alternatively spliced fibronectin (EDA isoform); and other structural macromolecules in these organs causes distortion of tissue architecture, resulting in progressive functional impairment. Fibrosis of the skin causes dermal expansion with obliteration of hair follicles and sweat and sebaceous glands. Lungs show patchy infiltration with lymphocytes, plasma cells,

TABLE 56.1 Genetic Polymorphisms Associated with Systemic Sclerosis

Gene Locus	Gene Function	Genetic Polymorphisms Associated with Systemic Sclerosis
IRF5	Activation of interferon	rs2004640, others
IRF8	Monocyte differentiation	rs11642837
IRF7	Activation of interferon	rs4963128
IL-12R	Interleukin-12/T-cell signaling	rs3790567
STAT4	T-cell signaling	rs7574865, others
DNASE1L3	DNA clearance	rs35677470
ATG	Autophagosome biogenesis	rs9373839
PPARG	Adipogenesis	rs310746
CD247	T-cell receptor signaling	rs2056626
CSK	Src family tyrosine kinase	rs1378942
PTPN22	T-cell signaling phosphatase	rs2476601
BANK1	B-cell receptor signaling	rs10516487, others
BLK	B-cell receptor signaling	rs2736340
TNFAIP3/A20	Negative regulation of nuclear factor (NF)-κB inflammation	rs5029939, others
TNIP1	Negative regulation of NF-κB inflammation	rs2233287, others
TNFSF4	T cell–antigen-presenting cell interaction	rs1234314, others

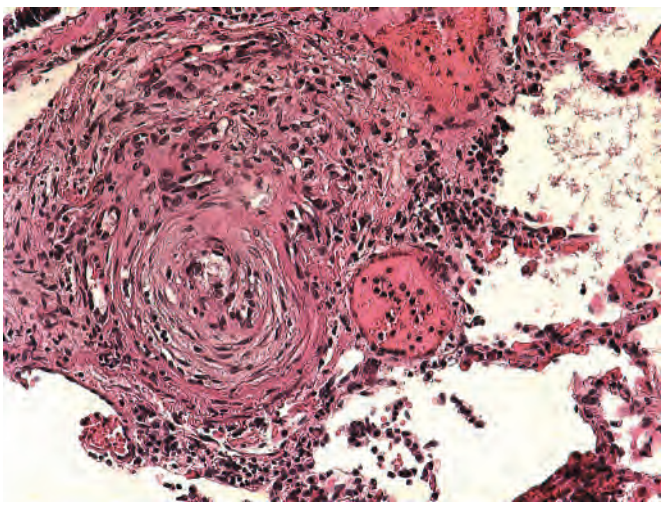


FIG. 56.1 Pulmonary Arterial Involvement. Significant intimal layer hyperplasia is seen, leading to narrowing of the vascular lumen. (Courtesy Dr. Anjana Yeldandi.)

macrophages, and eosinophils in early disease. In later stages, fibrosis and vascular damage predominate in the lungs, often coexisting in the same lesions. Intimal thickening of the pulmonary arteries, best seen with elastin stain, is a pathological hallmark of PH and at autopsy is often associated with multiple pulmonary emboli. Progressive thickening of the alveolar septae results in obliteration of the air spaces with a characteristic nonspecific interstitial pneumonia (NSIP) pattern and, less commonly, honeycombing (usual interstitial pneumonia, UIP), as well as loss of the pulmonary blood vessels. Pathological changes in the GI tract can occur at any level, from the mouth to the rectum. The esophagus shows atrophy of the lamina propria, submucosa, and muscular layers, with variable fibrosis. Replacement of the normal intestinal architecture results in disordered peristaltic activity, resulting in gastroesophageal reflux, small bowel dysmotility, and bacterial overgrowth. Chronic gastroesophageal reflux leads to esophageal inflammation, ulcerations, and stricture formation, and, in some cases, premalignant Barrett metaplasia.

The heart is frequently affected, with prominent myocardial contraction band necrosis, which reflects ischemia–reperfusion injury, and patchy areas of myocardial fibrosis. In the kidneys, vascular lesions predominate, and glomerulonephritis is rare. Scleroderma renal crisis (SRC) is an uncommon but often fulminant event associated with reduplication of elastic lamina, marked intimal proliferation, and narrowing of the lumen (onion skinning). Microangiopathic hemolysis may be accompanied by thrombocytopenia, making the differentiation from thrombotic thrombocytopenic purpura (TTP) challenging.

Pathogenesis

A comprehensive view of SSc pathogenesis must be capable of integrating the vasculopathy, immune dysregulation, and fibrosis of multiple organs, which are the hallmarks of the disease. As illustrated in Fig. 56.2, complex and dynamic interplay among these distinct pathomechanistic processes initiates, amplifies, and sustains tissue damage in SSc.[6] Animal models (Table 56.2) can be informative for delineating cell types, molecular mechanisms, and pathways contributing to SSc pathogenesis and for evaluating potential therapies.

Microangiopathy

Evidence of vascular involvement is an early, possibly initial, and widespread feature of SSc, and vascular damage in small and medium-sized vessels has major impact on the course of the disease. Vascular endothelial cell injury is initially associated with largely functional and potentially reversible alterations. Raynaud's phenomenon is characterized by abnormal blood flow response to vasomotor or cold challenge and altered production of, and responsiveness to, factors mediating vasodilatation (nitric oxide and prostacyclins) and vasoconstriction (endothelins). Microvessels show loss of pericyte coverage, increased permeability, enhanced transendothelial leukocyte migration, activation of fibrinolytic cascades, and platelet aggregation culminating in thrombosis. Endothelial–mesenchymal transition contributes to intimal and medial hypertrophy, which is coupled to fibrosis of the adventitial layers, resulting in vessel stiffening and luminal narrowing. Combined with endothelial cell apoptosis, the process culminates in the characteristically striking absence of small blood vessels seen on angiograms of the hands and kidneys in late-stage disease. Paradoxically, despite elevated circulating levels of angiogenic signals, including vascular endothelial growth factor (VEGF), reflecting pervasive tissue hypoxia, the process of revascularization appears to be defective in SSc.[7] Vascular injury in SSc is particularly prominent in the kidneys of patients with SRC, with evidence of endothelin-1 deposition and complement activation in the damaged small vessels.

Cellular and Humoral Immune Responses

In the early stages of SSc, activated T cells and monocytes/macrophages within lesional skin and the lungs secrete proinflammatory and profibrotic mediators, including transforming growth factor-β (TGF-β), cytokines, and chemokines. Because

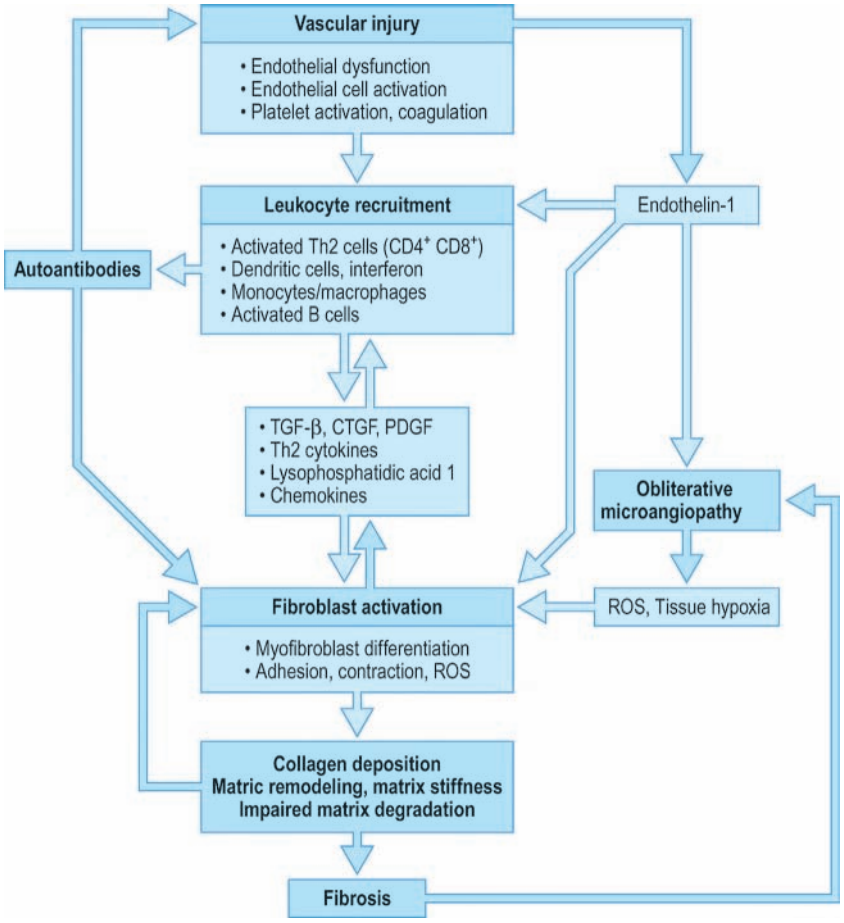


FIG. 56.2 **Pathogenesis of Systemic Sclerosis.** Interactions of cellular and molecular events triggered by injury that underlie the pathogenesis of vascular and immune dysfunction, culminating in fibrosis. *CTGF*, Connective tissue growth factor; *PDGF*, platelet-derived growth factor; *ROS*, reactive oxygen species; *TGF-β*, transforming growth factor-β.

TABLE 56.2 Selected Animal Models of Systemic Sclerosis

Animal Model	Fibrosis	Inflammation	Vascular	Affected Organ	Comment
Bleomycin (subcutaneous)	+	+	±	Skin, lungs	Fibrosis self-limited; localized
Graft-versus-host disease (GvHD)	+	+	–	Skin, lung, kidney	Complex procedure; radiation
Ang2-induced	+	+	+	Skin, lung	Skin localized
Fra2 tg	+	+	+	Skin, lung	Pulmonary hypertension (PH); early mortality
PDGFR tg	+	–	–	Multiorgan	
Wnt10b tg	+	–	–	Skin	Loss of dermal white adipose tissue (dWAT)
TSK1	+	–	–	Hypodermal lungs—emphysema	Fibrillin-1 mutation
TSK2	+	+	–	Skin	Collagen III mutation
Stiff skin syndrome	+	+	–	Skin	Fibrillin-1 mutation
Fli1-deleted	+	–	RVH	Skin, heart	
ROS	+	+	+	Skin, lung, kidney	Not widely used
Topo 1 immunization	+	+	–	Skin	Not well established
Chronic GvHD	+	+	±	Skin	Immune-driven

TGF-β, in particular, can induce its own production as well as that of other profibrotic paracrine mediators, such as connective tissue growth factor (CTGF) and platelet-derived growth factor (PDGF), an initial cytokine burst could result in amplified cytokine production and sustained autocrine and paracrine signaling. Additionally, as the extracellular matrix (ECM) of affected tissues undergoes fibrotic remodeling, it loses compliance and increases its stiffness. The mechanically altered microenvironment triggers biomechanical activation of resident stromal cells, both directly via the process of mechanotransduction and also by liberating latent matrix-bound TGF-β which then further amplifies the process. Alterations in the relative proportions and function of regulatory T cells (Treg) and T-helper 17 (Th17) cells, and innate lymphoid cells (ILCs) have been documented and may play important roles in pathogenesis. Single-cell RNA sequencing is generating a wealth of novel insights.

Virtually all patients with SSc have antinuclear and other autoantibodies in serum. Many of these autoantibodies are highly specific for SSc, associate with individual disease endophenotypes, and tend to be mutually exclusive (see below). Multiple mechanisms have been proposed to account for autoantibody generation in SSc.[8] A remarkable link between antibodies against RNA polymerase III and contemporaneous cancer has been noted, suggesting that some forms of SSc might represent a paraneoplastic syndrome. Although SSc-associated autoantibodies have well-documented clinical utility as diagnostic and prognostic disease markers, their direct pathogenic role in SSc remains uncertain.[9] Many patients with SSc have functional autoantibodies directed against fibroblasts and endothelial cells; the PDGF receptor; vascular cell receptors, such as endothelin-1 and angiotensin II receptors; and matrix metalloproteinases (MMPs). The pathogenic contribution of these potentially damaging autoantibodies is an area of intense investigation.

Fibrosis: Cellular and Molecular Components

Interstitial and vascular fibrosis, the hallmarks of SSc, are characterized by replacement of normal tissue architecture with dense and noncompliant connective tissue. Although restricted organ fibrosis is a very common sequel to any form of chronic or recurrent tissue injury, fibrosis synchronously affecting multiple organs is unique to SSc. Fibroblasts and related stromal cells of mesenchymal origin are key effector cells responsible for the development of fibrosis. Under physiological conditions, the fibroblast repair program is tightly regulated to enable optimal tissue regeneration, whereas under pathological conditions, fibroblast activation is sustained and amplified, resulting in exaggerated matrix deposition, disruption of tissue architecture, and failure of affected organs. Fibroblasts explanted from lesional SSc tissues display a variably abnormal phenotype when propagated ex vivo, with enhanced synthesis of collagen and other ECM molecules, expression of α smooth muscle actin and stress fiber formation, and spontaneous generation of reactive oxygen species (ROS). The persistent activated scleroderma phenotype reflects epigenetic modifications caused by chromatin remodeling, DNA methylation, or altered expression of noncoding regulatory microRNA and long RNAs.

The pleiotropic cytokine TGF-β is a pivotal regulator of both tissue repair and fibrosis (Chapter 9). Multiple abnormalities in TGF-β pathways have been identified in SSc, indicating a key role in pathogenesis.[10] Therapies that selectively or nonselectively block growth factor signaling triggered by growth factors, chemokines, PDGF, Wnt, and CTGF have shown some efficacy in both preclinical models and in human trials, and others are currently in clinical development.

CLINICAL FEATURES

Overview

The term *scleroderma* refers to systemic sclerosis, while *morphea*, also called *localized scleroderma*, denotes a disorder generally limited to skin or underlying tissue. SSc is highly variable in its clinical expression with subsets of patients that have unique clinical features and distinct clinical outcomes. Patients present with a seemingly common pathological disease process that can target skin, blood vessels, and the lungs, heart, GI tract, kidneys, and musculoskeletal system. Raynaud phenomenon is virtually universal in SSc, suggesting that perturbation of the terminal arteries of the circulation is a fundamental process that is common to the different subsets of the disease. Thickening of the skin distinguishes SSc from other rheumatic diseases (Table 56.3); scleroderma (hard skin) is the most specific and prominent physical finding. A small number of patients develop systemic features of SSc without appreciable skin involvement, a phenotype that is termed *systemic sclerosis sine scleroderma*. While traditionally patients are classified by the degree of clinically involved skin into two major subtypes (diffuse [dcSSc] or limited [lcSSc]), it is now recognized that the disease course and the ultimate specific organ outcome can be better predicted by a classification using a composite score encompassing the degree of skin involvement, autoantibodies, and the degree of presenting organ disease. Predictors of elevated mortality rates among patients with SSc include diffuse skin disease with a rapid rate of skin progression and internal organ involvement (especially the lungs or the kidneys, resulting in an SRC), male gender, black race, presence of malignancy, and later age of disease onset.[11,12] In some patients with SSc who have "overlap" features, SSc coexists with clinical and laboratory evidence of another autoimmune disease, such as polymyositis, autoimmune thyroid disease, Sjögren syndrome, polyarthritis, autoimmune liver disease, or systemic lupus erythematosus (SLE) (Table 56.4).

TABLE 56.3 Classification of Systemic Sclerosis

Diffuse cutaneous scleroderma—skin thickening on the trunk in addition to the face, proximal and distal extremities
Limited cutaneous scleroderma—skin thickening limited distal to the elbow and knee; may also involve the face and neck
CREST syndrome—subcutaneous *c*alcinosis; *R*aynaud phenomenon; *e*sophageal dysmotility; *s*clerodactyly; *t*elangiectasia (term no longer used)
Sine scleroderma—systemic sclerosis with no apparent skin thickening but characteristic visceral organ involvement, vascular and serological features
Overlap syndrome—criteria for systemic sclerosis (SSc), coexisting with features of systemic lupus erythematosus, rheumatoid arthritis, or inflammatory muscle disease
Mixed connective tissue disease—overlap syndrome with anti-U1 ribonucleoprotein (RNP) antibodies
Early disease—Raynaud phenomenon (RP) with puffy fingers and other clinical and/or laboratory features of SSc; specific autoantibodies, abnormal nailfold capillaroscopy, finger edema, and ischemic injury (also called very early diagnosis of systemic sclerosis, VEDOSS)

Morphea

Morphea (localized scleroderma) is now properly used to describe a group of patients with disease generally limited to fibrosing skin disorders that occurs with similar frequency in both children and adults (Table 56.5). Unlike scleroderma, in morphea internal organ involvement is unusual. Subtypes of morphea exist including circumscribed or plaques that can occur as solitary or multiple lesions (generalized morphea), linear lesions, pansclerotic, or mixed forms. Linear morphea can occur on the head, leaving severe facial deformity. When this occurs in the upper forehead, it is called *en coup de sabre*. This should be distinguished from *Parry-Romberg syndrome*, or *facial hemiatrophy*, which also affects the face, usually unilaterally, with atrophy of skin rather than fibrosis. Both en coup de sabre and Parry-Romberg syndrome can be associated with central nervous system (CNS) involvement, presenting as headaches, visual disturbance, or a seizure disorder.

TABLE 56.4 Frequency and Clinical Correlations of Systemic Sclerosis Autoantibodies

	% Frequency in SSc	Disease Subtype	Clinical Associations	Prognosis
Anticentromere	20–38	lcSSc	Pulmonary arterial hypertension	Better prognosis
Anti-topoisomerase I	15–42	dcSSc	Pulmonary fibrosis Heart involvement	Worse prognosis
Anti-RNA polymerase III	5–31	dcSSc	Renal crisis Tendon friction rubs, synovitis, myositis, joint contractures. Increased risk of cancer	Increased mortality
Anti-U3 RNP (fibrillarin)	4–10	dcSSc	Renal crisis and cardiac involvement	Poor prognosis, especially in African Americans
Anti-Th/To	1–13	lcSSc	Pulmonary fibrosis and renal crisis	Poor prognosis
Anti-U11/U12 RNP	3.2	—	RP Gastrointestinal involvement Lung fibrosis	Increased mortality
Anti-U1 RNP	2–14	lcSSc	RP, puffy fingers, arthritis, myositis, overlap syndrome (*i.e.*, MCTD)	Better prognosis
Anti-PM-Scl	4–11	Overlap with polymyositis lcSSc	RP, arthritis, myositis pulmonary involvement, calcinosis, and sicca symptoms	Better prognosis
Anti-Ku	2–4	—	Myositis, arthritis, and joint contractures	—
Anti-hUBF (NOR 90)	<5	lcSSc	Mild internal organ involvement	Better prognosis
Anti-Ro52/TRIM21	15–20	Association with other autoimmune diseases	Older age onset, pulmonary fibrosis	—

dcSSc, Diffuse cutaneous SSc; *lcSSc*, limited cutaneous SSc; *MCTD*, mixed connective tissue disease; *RNP*, ribonucleoprotein; *RP*, Raynaud phenomenon; *SSc*, systemic sclerosis; *TRIM*, tripartite motif.
From Kayser C, Fritzler MJ. Autoantibodies in systemic sclerosis: unanswered questions. *Frontiers Immunol.* 2015;6:167.

KEY CONCEPTS

Clinical Patterns of Systemic Sclerosis

Limited Cutaneous Scleroderma (lcSSc)
- Skin thickening limited to distal limbs or fingers alone (sclerodactyly)
- Increase matted telangiectasia and subcutaneous calcinosis
- Severe Raynaud phenomenon with digital ischemia
- Risk of pulmonary arterial hypertension
- Other autoimmune diseases (hypothyroid, primary biliary cholangitis, Sjögren syndrome)

Diffuse Cutaneous Scleroderma (dcSSc)
- Even skin thickening proximal limbs (upper arms, chest, abdomen)
- Increased risk of internal organ disease (ILD, SRC, heart, GI)
- Rapid disease onset
- Intense fatigue, weight loss, pruritus, arthritis, contractures
- Tendon friction rubs
- Myopathy

Systemic Sclerosis Sine Scleroderma
- No skin thickening
- Raynaud phenomenon or typical peripheral vascular disease
- One or more typical scleroderma visceral disease
- Autoantibodies

Morphea
- Localized scleroderma
- Lack of systemic disease

GI, gastrointestinal; *ILD*, interstitial lung disease; *SRC*, scleroderma renal crisis.

TABLE 56.5 Classification of Localized Scleroderma

- Plaque types of (morphea)
- Circumscribed plaques
- Guttate
- Keloid/nodular
- Bullous
- Generalized morphea
- Three or more locations
- Pansclerotic morphea
- Linear scleroderma
- Frontoparietal linear morphea (*en coup de sabre*)
- Parry-Romberg syndrome (progressive hemifacial atrophy)
- Linear streaks on limbs or trunk
- Deep morphea

Symptoms

Characteristically, the earliest symptoms of SSc are nonspecific and include fatigue, musculoskeletal distress (stiffness or pain), and feeling ill. Cold sensitivity with associated Raynaud's phenomenon is often the only early clinical clue to the presence of disease. Symptoms of esophageal dysfunction with dysphagia, heartburn, and periodic GI reflux, along with Raynaud's phenomenon, can precede other manifestations of SSc by years. In general, patients with dcSSc have a short interval between the onset of Raynaud's phenomenon and other signs and symptoms. Soft tissue swelling, intense pruritus and burning, and pitting edema on the limbs are signs of the early inflammatory or "edematous" phase of the diffuse cutaneous form of the disease. The skin of the fingers, hands, distal limbs, and face are usually affected first and more severely compared with other body areas (Fig. 56.3). Patients may note skin hyperpigmentation (patches or generalized tanning). Other early skin changes include vitiligo-like hypopigmentation ("salt and pepper" appearance), often along the scalp, dorsum of the hand, on the chest, or the back (Fig. 56.4). Escalating musculoskeletal symptoms are common and are associated with muscle weakness and decreased joint mobility.

CLINICAL PEARLS

Clinical Features of Early Systemic Sclerosis

- Definite Raynaud phenomenon
- Gastroesophageal reflux with heartburn
- Swelling of the fingers and hands
- Musculoskeletal pain and stiffness
- Dilated nailfold capillaries
- Hyper- or hypopigmentary changes of skin

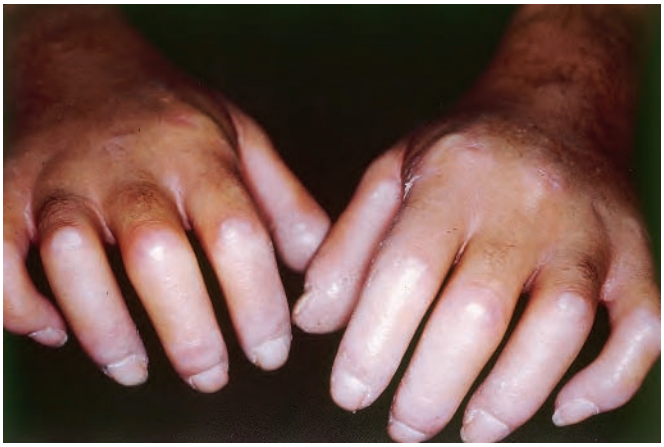


FIG. 56.3 Skin Fibrosis in Systemic Sclerosis. Severe fibrosis of the skin of the hands and forearms causing joint contractures and skin ulcerations in a woman with diffuse cutaneous systemic sclerosis.

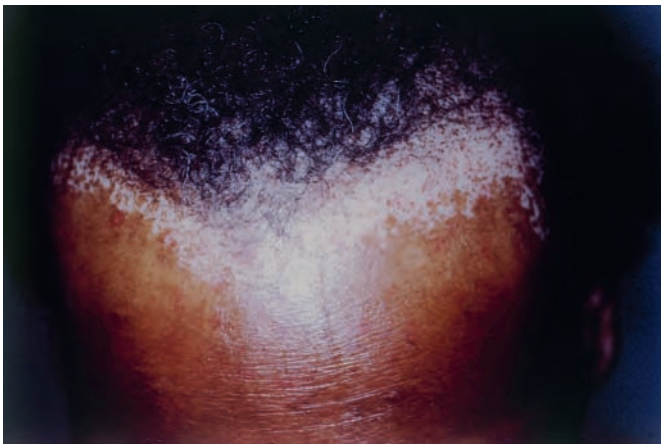


FIG. 56.4 Pigmentation Changes in the Skin. Vitiligo (salt-and-pepper) appearance of the involved skin in a patient of African descent with diffuse cutaneous systemic sclerosis.

Clinical Course

The early edematous phase of dcSSc has prominent inflammatory features, with significant skin edema and erythema, and is associated with inflammatory cell infiltration in the dermis. After a period of weeks to months, the inflammatory phase fades and the "fibrotic" phase dominates with increased collagen and extracellular material deposited causing thickening of the skin and loss of flexibility. The fibrotic process starts in the dermis and is associated with loss of body hair, reduced production of skin oils, and a decline in sweating capacity as these cutaneous structures atrophy. Gradually, the subcutaneous tissue becomes affected, with atrophy of subcutaneous fat and fibrosis extending to the underlying fascia, muscle, and other soft tissue structures. Progressive flexion contractures of the finger and other joints ensue (Fig. 56.5). Tendon friction rub is a prominent crepitation that can be felt or even heard over tendons of the lower and upper extremities. Tendon friction rubs are caused by fibrosis in the tissues surrounding the affected joints; they are associated with rapidly progressive skin disease and are linked to an overall poor prognosis.

While the degree of skin involvement is highly variable, the natural history of skin disease in dcSSc tends to be monophasic,

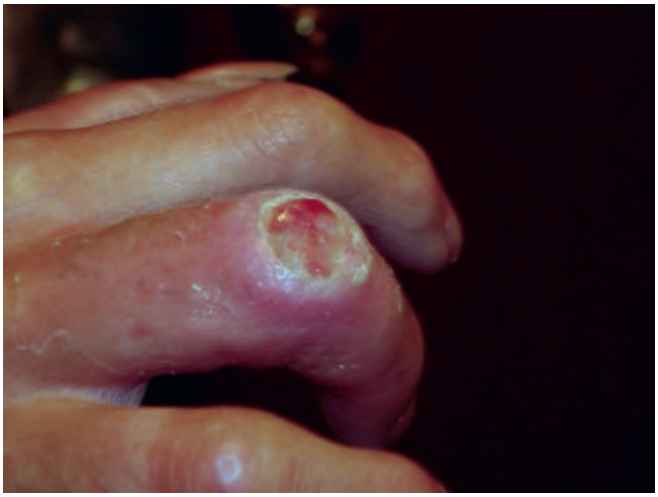


FIG. 56.5 Traumatic Digital Ulcer. Ulceration of atrophic skin over the metacarpal joint of patient with diffuse scleroderma.

and relapse is uncommon after the edematous and active fibrotic phase.[13] The duration of active skin disease from the first signs of skin involvement to its maximal extent is characteristically about 18 months; cutaneous inflammation and progressive fibrosis gradually subside, and regression of skin involvement begins. In the late stages of dcSSc, skin remodeling can be dramatic, with return to normal-appearing skin in those areas spared from severe end-stage fibrosis. Skin ulcerations or dystrophic calcification often are late disease complications of the fibrotic, atrophic, and avascular skin (see Fig. 56.5). Although the skin involvement is generally the most dramatic and visible manifestation of dcSSc, internal organ involvement occurs during the early active stage of advancing skin disease. Patients with dcSSc have a significant risk for interstitial lung disease (ILD), severe GI dysfunction, SRC, progressive heart disease, and recurrent digital ulcers during the initial active inflammatory and fibrotic phases of the skin disease. In practical terms, this means that in dcSSc, the initial 3 to 4 years is the period that the systemic process is most active; if organ failure does not occur during this period, the systemic process may stabilize without further progression.

In contrast to dcSSc, the disease course in the limited cutaneous (lcSSc) variant of SSc is more indolent and often relatively benign. After the onset of Raynaud phenomenon, several years may pass before additional symptoms or signs are recognized. The most common non-Raynaud's symptoms in patients with lcSSc are those of upper GI disease with dysphagia and gastroesophageal reflux. Dilated venules form visible erythematous vascular lesions *(telangiectasia)* early in the disease, seen most commonly on the fingertips, palms, face, lips, and inside the oral cavity (Fig. 56.6). Subcutaneous calcinosis caused by deposition of calcium hydroxyapatite crystals occurs commonly at sites of tissue ischemia and recurrent trauma, such as the fingertips, forearm, or elbow. Severe Raynaud's phenomenon with macrovascular occlusive involvement occurs more frequently in lcSSc than in dcSSc and can be associated with critical digital ischemia, ischemic ulcerations, gangrene, and amputation.

In both dcSSc and lcSSc, visible capillary abnormalities easily seen at the nailfold (dilatation and loss or dropout of capillaries) are near universal. An individual presenting with Raynaud's phenomenon, abnormal nailfold capillaries, puffy fingers, and the presence of an SSc-specific autoantibody can be suspected of having SSc even before other more obvious manifestations are noted.[14]

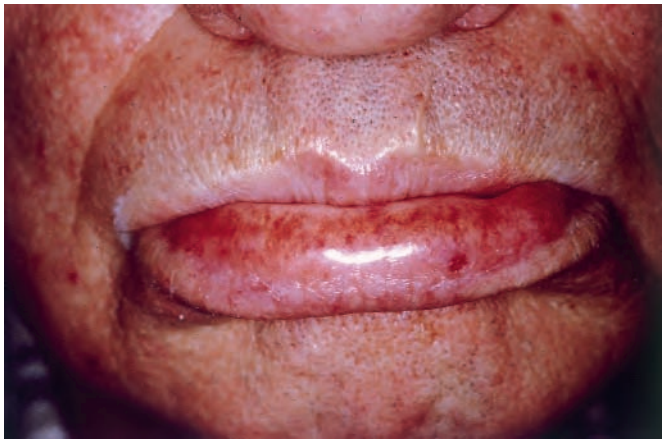


FIG. 56.6 Telangiectasia. Characteristic telangiectasia on the lip in a woman with limited cutaneous systemic sclerosis.

Raynaud's Phenomenon

Raynaud's phenomenon (RP) is clinically defined as well-demarcated color changes of the digits induced by cold or emotional stress.[15] Vascular constriction of arterio-venous anastomoses (AVAs), arterioles, and small arteries in the skin and tissues of the digits causes the pallor of the digits, the phase of complete loss of blood flow. The initial vasospasm and pallor is followed by cyanosis of the skin caused by venous pooling and low flow with deoxygenated blood. Finally, after rewarming, the recovery phase occurs with vasodilation manifested by hyperemia with blushed reddened skin as blood flow rebounds. Raynaud's phenomenon is common, affecting 3% to 5% of the general population and about 10% to 15% of selected populations (e.g., cold geographic environments). Raynaud's phenomenon is considered to be *primary RP* when there is no other associated disease state. In these cases, it is thought to represent a genetic trait with young age of onset of cold sensitivity as a result of abnormal cutaneous vessel reactivity to environmental temperatures. The underlying defect in primary RP is generally considered to be a *local fault* in the thermoregulatory vessels with a change in functional activity of the vasculature without evidence of structural alteration or injury to vessels. There is evidence that this local fault is due to enhanced sympathetic responses via increased α_2-adrenergic receptor activity on the smooth muscle cells of the involved AVAs in patients with Raynaud's phenomenon.[16]

In contrast, *secondary RP* are a variety of acquired conditions including those that cause vascular injury and can be associated with structural deterioration and/or loss of the cutaneous vessels including the nutritional microvasculature. In SSc, the disease process targets the peripheral vasculature, including AVAs or thermoregulatory blood vessels. This leads to not only the abnormal reactivity to ambient temperatures typical of Raynaud's phenomenon but also significant structural disease of blood vessels, which causes tissue ischemia as a result of compromise to blood flow within capillaries or nutritional vessels. As a consequence of both an exaggerated response to cold and structural vascular disease, patients with SSc have severe Raynaud's phenomenon with multiple, and often prolonged, daily episodes. The severe vascular disease can lead to critical ischemia with digital ulcerations or deep tissue infarction and gangrene. The vascular disease in SSc is not limited to skin but is systemic. There is evidence that abnormal vascular reactivity resulting in tissue injury occurs in the pulmonary, renal, GI, and coronary circulations.

Treatment of Raynaud's phenomenon must be individualized and adjusted according to its severity. Initial therapy should include avoidance of cold exposure and use of methods of reducing sympathetic tone such as reducing emotional stress. Many patients with primary RP do well with non-drug therapy alone. The best studied medications to treat Raynaud's phenomenon are the calcium channel blockers and they are the recommended first choice of therapy. Currently, phosphodiesterase inhibitors such as sildenafil are considered a second-line option with or without a calcium channel blocker. Evidence for use of other agents is weak.

Gastrointestinal Involvement

GI disease is a major cause of morbidity and mortality and accounts for one of the most frequent initial symptoms of scleroderma.[17] Every part of the GI tract can be involved, including the oral pharynx, esophagus, stomach, and the small and large bowels. Almost all patients with SSc demonstrate evidence of distal esophageal dysfunction. Clinically, this presents with dysphagia, heartburn, and regurgitation typical of gastroesophageal reflux disease (GERD). These symptoms are caused by peristaltic dysfunction of the distal esophagus, decreased lower esophageal sphincter (LES) pressure, and delayed gastric emptying, which is thought to be secondary to autonomic dysfunction. Pathological studies of the esophagus show that the smooth muscle (circular greater than longitudinal) of the bowel atrophies without significant fibrosis, vascular injury, or obvious inflammation. Functional and pharmacological studies have shown that a neurogenic process precedes smooth muscle dysfunction. Autoantibodies directed against enteric neurons and anti-muscarinic antibodies are found, suggesting that the initial insult is an immune process targeting neurological mechanisms of normal GI motility.

Early satiety, bloating, nausea, periodic vomiting, and decreased appetite with weight loss can occur secondary to poor gastric emptying (gastroparesis). Gastroparesis is commonly associated with retention of food and liquids in the stomach and esophageal dysfunction thus aggravating GERD. Mild to severe intestinal dysmotility can involve either or both the small and large bowels. Patients may note a change in bowel habits with episodes of either diarrhea from small disease or periods of constipation due to decreased transit in the large bowel. Persistent diarrhea with bloating and pain may be a manifestation of malabsorption of fats as an atonic and small intestinal bacterial overgrowth (SIBO); untreated this can be associated with dramatic weight loss and malnutrition. In severe cases bowel dysmotility episodes of pseudo-obstruction involving either the small or large bowel presents with severe abdominal pain, bloating, abdominal distension, and vomiting.

Treatment of the scleroderma bowel involvement is focused on conventional methods of managing GI dysmotility in that there is no solid evidence that disease-modifying therapy such as immunosuppression alters the GI disease process. GERD can be managed with careful eating habits, attention to body position after eating, blocking acid production with an H_2 blocker or a proton pump inhibitor, and in some cases the use of a pro-kinetic drug to help esophageal emptying (e.g., metoclopramide, domperidone). Constipation is treated with appropriate fiber intake, a stool softener, or agents to assist bowel movement (e.g., polyethylene glycol, linaclotide, or prucalopride), while SIBO

and diarrhea respond to dietary changes (e.g., FODMAP diet) and intermittent antibiotic therapy. Severe disease with pseudo-obstruction requires periods of bowel rest and total parenteral nutrition (TPN).

Pulmonary Involvement

Lung disease accounts for significant lifetime morbidity and is now the leading cause of death in patients with SSc. A survey of all-cause mortality reported that 35% of all SSc-related deaths were directly attributable to interstitial lung disease SSc-related ILD (SSc-ILD), and 26% of deaths resulted from pulmonary arterial hypertension (PAH) or PH with secondary heart disease. Another survey reported a high risk of mortality from ILD (hazard ratio [HR] 3.70); PAH identified on transthoracic echocardiography (TTE) (HR 7.49), and a diffusing capacity for carbon monoxide (DLCO) of less than 60% of the predicted value (HR 3.17).[18] Emerging pulmonary vascular disease leads to PAH in about 10–15% of patients with scleroderma.

Interstitial Lung Disease

While ILD is the leading cause of mortality, the clinical course is highly variable without progressive disease in the majority. NSIP is the most common histopathological pattern in patients with SSc-ILD. Inflammatory alveolitis and subsequent tissue fibrosis causes a functional restrictive ventilatory defect and an abnormal gas exchange. In early disease, SSc-ILD is often asymptomatic, detected in over 40% by either lung function or high-resolution CT scan imaging (HRCT). The presence of fibrosis and ground-glass opacities on HRCT is an indication of the activity of the disease process and correlates with subsequent decline in forced vital capacity (FVC). Aggressive life-threatening disease occurs in a subgroup of patients (10% to 20%) and is measured by a progressive fall in FVC usually within 4 years of the onset of scleroderma. It is a good prognosis if there is only minor impairment in FVC on lung function on first encounter, but a FVC of less than 70% of predicted is associated with a poor prognosis. Likewise, when the extent of disease defined by HRCT is less than 20%, the 10-year survival is reported to be 67%, whereas in the group with HRCT-defined disease extent of greater than 20%, 10-year survival is considerably poorer at 43%.[26] There are clinical features and biological biomarkers that can predict the course of ILD. The prevalence of SSc-ILD is found in 50% of those with dcSSc compared with 35% in those with lcSSc. The presence of topoisomerase-I, U3RNP, or Th/To autoantibodies are strongly linked to risk of developing SSc-ILD, and the worst outcome is often seen in African Americans and Native Americans.[27] Two multicenter clinical trials give evidence that immunosuppressive therapy can control active SSc-ILD; one compared cyclophosphamide to placebo, and the other compared cyclophosphamide to mycophenolate. Both studies demonstrated statistically significant stabilization or improvement of lung function with equivalent benefit for both agents. Nintedanib (a tyrosine kinase inhibitor) with antifibrotic properties is now FDA approved for treatment of early active SSc-ILD. Other agents are being tested in SSc-ILD, including rituximab (monoclonal antibody [mAb] to CD20 on B lymphocytes) and pirfenidone (an antifibrotic/anti-inflammatory agent used in idiopathic pulmonary fibrosis).

Pulmonary Hypertension

PH can occur in scleroderma secondary to isolated pulmonary vascular disease (PAH), in association with SSc-ILD and hypoxia or secondary to cardiac disease. Therefore, patients need to undergo a right heart cauterization (RHC) to confirm the presence of PAH or PH and to characterize its severity. PAH is now defined by RHC showing a mean pulmonary artery pressure (mPAP) greater than 20 mm Hg and a pulmonary capillary wedge pressure ≤15 mm Hg with pulmonary vascular resistance (PVR) ≥3 Wood without evidence of significant pulmonary parenchymal disease. PAH is detected by RHC in 8% to 15% of patients. The natural history of PAH is highly variable; in the majority it is clinically silent for years (9 to 12 years after the onset of disease) before it presents with exertional dyspnea and hypoxia secondary to altered gas exchange and right heart failure. PAH can also occur within 5 years from the disease onset, particularly in patients with anti-U3 RNP (fibrillarin) antibodies. Risk factors for developing SSc-PAH include disease onset at a later age, history of severe RP, numerous skin telangiectasias, abnormal nailfold capillaries, and the presence of anti-centromere, anti-U1 ribonucleoprotein (RNP), anti–U3 RNP, or anti-B23 antibodies. Early detection and intervention with combination therapy has improved survival with now a 3-year survival reported to be 56% to 75%. Current therapy for SSc-associated PAH is focused on supportive care, reduction of cardiac workload, and vasoactive drugs.[19] Treatment regimens include four classes of medications which are prostacyclin analogs, endothelin-receptor antagonists, phosphodiesterase-5 inhibitors, and guanylate cyclase stimulators. In relatively short-term clinical trials, each of these treatments has been shown to improve exercise tolerance and hemodynamics, with variable benefit on disease progression. Early intervention and combination therapy help to achieve improved survival outcomes. Lung transplantation remains an option for carefully selected patients with SSc, with survival following lung transplantation being similar to that of non-SSc patients with other lung diseases.

Cardiac Involvement

Cardiac involvement is common, but often there are no signs or symptoms until the heart has been severely affected. Clinical criteria are insensitive, but modern sensitive tools, such as echocardiography, tissue Doppler imaging, Holter monitoring, speckle-tracking echocardiography, cardiac magnetic resonance imaging, or thallium scanning, can detect disease early and suggest that 40% to 60% of patients have heart disease. When heart disease is symptomatic, the prognosis is poor with primary heart disease accounting for 15% to 30% of all SSc deaths. SSc can affect virtually any cardiac structure, including the myocardium with fibrosis or inflammatory myocarditis, the myocardial microvasculature, or the pericardium with asymptomatic small effusions or large effusions leading to cardiac tamponade. As a consequence of cardiac tissue injury, left ventricular (LV) systolic dysfunction and/or LV diastolic dysfunction and/or right ventricular (RV) failure can occur and conduction abnormalities with arrhythmias are common, but valvular heart disease is unusual. Myocardial fibrosis is thought a consequence of vasospasm of coronary arterial microcirculation, cardiomyocyte injury, and fibroblast proliferation. Involvement of a larger vessel causing coronary artery disease is not established. Cardiac disease occurs in both the limited and the diffuse subtypes but is more common in the patients with rapidly progressive diffuse skin disease, in those with anti-SCL70, anti-U3 RNP, anti-KU, or anti-Th/To antibodies, and in association with peripheral muscle disease. The ideal therapy is defined by the underlying situation (e.g., vasodilator therapy to improve heart perfusion

or anti-arrhythmia therapy for a complex arrhythmia). A disease-modifying drug that specifically targets scleroderma heart disease has not yet been discovered.

Renal Involvement

Renal disease is an important clinical problem with both direct involvement and compromise to kidney function secondary to other organ dysfunction. The most dreaded renal complication is a SRC, which affects approximately 11% of dcSSc and 4% of lcSSc subjects. A SRC classically presents with the new onset of accelerated arterial hypertension and rapidly progressive oliguric renal insufficiency, but more modest elevations in blood pressure and normotensive events can occur due to the same underlying pathology. Other causes of renal disease can occur in SSc, including interstitial nephritis, glomerulonephritis, and anti-neutrophil cytoplasmic antibody (ANCA)–associated vasculitis, which can be confused with an SRC. The exact mechanism triggering acute SRC is not known, but an autoimmune insult to the underlying SSc microvascular disease is thought to initiate a self-perpetuating process leading to intimal thickening and vascular occlusion, glomerular hypoperfusion, and a high renin state. SRC generally develops relatively early (<3 to 4 years) in the disease; in patients with rapid progression of diffuse skin disease who have anti-RNA polymerase III antibodies (about 30%) or anti-topoisomerase (about 10%); the risk is low in patients with anti-centromere antibodies. Patients with SRC who present with a creatinine greater than 3 mg/dL have a poor outcome, with increased likelihood of permanent hemodialysis, need for renal transplantation, and/or high mortality. Prompt and aggressive intervention with angiotensin-converting enzyme (ACE) inhibitors to achieve blood pressure control before evidence of renal dysfunction significantly improves the prognosis; since the introduction of ACE inhibitors survival from SRC has dramatically improved from a 1-year survival rate of 10% to a 5-year survival rate of 60%.[20] About 30% to 45% of patients whose blood pressure is normalized do not require dialysis, but despite ideal therapy, approximately 40% of patients with SSc still need long-term kidney support or have a poor survival. In a subset of patients, reasonable kidney function can resume months after renal failure and they no longer need dialysis support. Renal transplantation is still an option, with a long-term outcome similar to that of other causes of renal failure.

Musculoskeletal Complications

Musculoskeletal impairment is a leading cause of disability and poor quality of life among patients with SSc. It is seen in the majority of patients, occurs early in the disease, and commonly has a dominant effect on hand function. There is a variety of associated clinical symptoms, including arthralgias, myalgias, stiffness, pain and loss of function as a result of joint contractures, inflammatory arthritis, myopathy, nerve entrapment (e.g., carpal tunnel syndrome), fibrosis of tendons, and deconditioning from disuse. Loss of joint function is less likely to be secondary to synovitis but is usually caused by fibrosis of the overlying skin, the supporting joint structures, and the joint capsule itself. Flexion contractures associated with "friction rubs" most prominently felt over the ankles, knees, shoulders, and wrist are typical of advanced dcSSc. Muscle weakness is a common comorbid condition caused by deconditioning, muscle disuse, or malnutrition associated with weight loss. Most patients with skeletal myopathy have proximal muscle weakness on examination, and about 40% have one of the scleroderma/myopathy–associated autoantibodies (*i.e.*, anti-PM/Scl-75, anti-PM/Scl-100, anti-CD1 or anti-Ku). While the predominant features on muscle biopsies of the weak scleroderma patient are necrosis and inflammation, there is a wide range of histopathological features consistent with overlap polymyositis, dermatomyositis, necrotizing myopathy, and fibrosis alone. The small subset of SSc patients with a fibrosing myopathy alone tends to be of the diffuse skin subtype, African American race, have lower FVC, and have shorter disease duration.[21,22] When inflammation and/or necrosis are seen, immunosuppressive therapy has the potential to control the muscle disease.

Emotional Aspects

Scleroderma is a chronic disfiguring disease that can affect every aspect of a patient's life including employment, performance in the family, social interactions, and sexually. Fear, anxiety, and depression are coupled with functional disability, chronic pain, sleep disturbance, fatigue, and a poor view of appearance and body image. A major depressive disorder is a common consequence that may fluctuate in severity and have an unrecognized major negative impact on quality of life. The emotional aspects of the disease need to be addressed by both medical intervention and management of the psychological aspects of living with the disease.

Approach to Treatment

CLINICAL PEARLS

Recommended Approach to Systemic Sclerosis

- Determine scleroderma skin score to define clinical subtype
- Frequent clinical reassessment to define disease activity and emotional impact
- Careful clinical history to evaluate for gastrointestinal dysmotility
- Monitor for new-onset hypertension: prevent renal crisis
- Pulmonary function test: early detection of interstitial lung disease
- Doppler echocardiography: screen for pulmonary arterial hypertension
- Obtain serology profile to help predict clinical outcome

The clinical and biological evidence support the concept that SSc occurs in a genetically susceptible host, and that the disease process is initiated by a yet unidentified factor (e.g., cancer or environmental factor) that provokes autoimmunity with secondary inflammation, vascular injury, and, unlike other autoimmune diseases, progressive tissue fibrosis. The spectrum of the disease is highly variable with distinct clinical phenotypes ranging from rapid diffuse skin disease with serious internal organ involvement to a mild course with minimal skin fibrosis and overall good outcome. The natural course of the disease also varies in an individual patient. Inflammation occurs early in the acute phase, with signs of vascular perturbation (e.g., RP), followed by rapid tissue fibrosis and then later an indolent phase with skin remodeling with or without progressive internal organ disease. The management must target the specific situation, level of disease activity, and specific organ dysfunction. Most success occurs in treating specific organ involvement (e.g., ACE inhibitor for SRC) as outlined in the above text. Intervention with currently available drugs should be initiated early, ideally during the edematous active inflammatory phase of the disease. It is during the early stage of the disease that any disease modifying agent has the greatest potential to control disease-progression. The primary outcome measure in most clinical

trials has focused on aggressive skin disease as measured by the modified Rodnan skin score; in current studies, however, composite scores that tally the burden of disease (activity and severity) as well as patient-reported distress are included. While no disease-modifying agent is fully established, attempts to control the scleroderma disease process can be divided into immunotherapy, treatment of fibrosis, and medications directed at the vascular disease process.[6,23]

Immunotherapy

Nonselective immunotherapy is the most popular approach (methotrexate, cyclophosphamide, azathioprine, mycophenolate, intravenous immunoglobulin) used to control active skin disease, ILD, inflammatory muscle or joint disease, while few studies have supported the use of selective immunotherapy (anti-thymocyte globulin, cyclosporin A, sirolimus). Biological agents have been used to target inflammatory cytokines (anti-TNF, Il-6, IL-1) or T-cell activation (abatacept) or B-cell populations (rituximab). For patients with severe disease or who are failing low-dose immunotherapy, cell-based immunotherapy using autologous hematopoietic stem cell transplantation following myeloablation or non-myeloablative therapy has shown promise.[24,25]

Treatment of Fibrosis

While there are several agents under investigation, there is no clear evidence of an agent that directly suppresses activated tissue fibroblasts. Strategy to alter fibroblast to myofibroblast or epithelial–mesenchymal transition, shortening fibroblast survival or affecting fibroblast interactions with ECM via blocking cellular integrins are all under investigation. Antagonists of the bioactive phospholipid lysophosphatidic acid (LPA), agonist for the cannabinoid receptor (CB2) and small molecules that inhibit tyrosine kinase (imatinib) and downstream growth factors are being studied. Nintedanib has shown promise in a clinical trial demonstrating a reduction in the rate of decline of lung function in patients with SSc-ILD.[18] Pirfenidone, another antifibrotic agent, is being tested in a clinical trial for treating SSc-ILD.

Treatment of Vascular Disease

Although nonspecific vasodilation therapy (see Reynaud Phenomenon and Pulmonary Hypertension sections) is the main approach to the vascular disease, it is recognized that vascular injury is a major process that needs attention. Targeting the vascular disease process is attempted by the use of endothelin receptor antagonists, prostaglandin therapy, inhibiting platelet activation, statins, enhancing nitric oxide (phosphodiesterase inhibitors), and immunotherapy.

Treatment

It is important to carefully characterize the patient's clinical phenotype, specific organ involvement, and level of disease activity before deciding on therapy. It is also important to distinguish disease activity from disease severity or advanced cumulative organ damage. For example, a patient with late-stage dcSSc with irreversible organ damage is unlikely to benefit from aggressive immunosuppressive or antiinflammatory therapy. It is during the early edematous stage of the disease that immunosuppression and antiinflammatory and antifibrotic agents have the greatest potential to control disease progression. Clinical experience teaches that once the edematous phase of the disease shifts to the more indolent fibrotic phase, current treatments are less likely to control the progression of disease and tissue damage. Few well-designed controlled studies have been carried out for the treatment of early active disease; most reports are of anecdotal, uncontrolled experiences, complicated by investigator bias and the highly variable natural course of SSc.

Scleroderma is a multifaceted disease that includes active autoimmunity, vascular injury, and fibrosis and/or epithelial damage. Thus, the major targets for therapy include regulating the immune system, controlling tissue fibrosis, and protection or prevention of progressive vascular disease. Drugs currently used for dcSSc include methotrexate, cyclophosphamide, and mycophenolate mofetil. Targeting specific immune cells is now a novel strategy that includes eliminating B cells with rituximab and suppressing activation of T cells with abatacept. Inhibiting proinflammatory and/or profibrotic cytokines or signaling pathways that may block tissue injury and fibrosis are being explored. These include anti-chemokines (CCR2) or inhibitors of CTGFs, recombinant humanized mAb to block TGF-β_1 activity, modulation of interleukin-6 (IL-6) signaling, antagonists of the bioactive phospholipid LPA, and small-molecule inhibitors, such as tyrosine kinase inhibitors or pirfenidone. It is recognized that vascular disease also plays a fundamental role in the morbidity and mortality seen in scleroderma. Novel drugs targeting vascular disease, including endothelin receptor antagonists, agents that enhance nitric oxide production, prostacyclin analogs, antiplatelet agents, and statins, are currently being used.

THERAPEUTIC PRINCIPLES

Treatment of Systemic Sclerosis

Organ-Specific Therapy

- Vasodilator therapy for Raynaud's phenomenon
- Proton pump inhibitor for gastroesophageal reflux
- Angiotensin-converting enzyme inhibitor for scleroderma renal crisis
- Vasodilator therapy for pulmonary arterial hypertension
- Antiinflammatory therapy for arthritis
- Immunosuppressive therapy for interstitial lung disease

Disease-Modifying Agents

- Nintedanib, an orally active small-molecule multiple kinase inhibitor, is currently approved for the treatment of SSc-associated ILD
- Several immunomodulatory, antiinflammatory, vasoactive, and antifibrotic agents, as well as autologous hematopoietic stem cell transplantation, are used and/or currently under investigation.

ON THE HORIZON

Novel Therapeutic Approaches to Systemic Sclerosis Management in Clinical Trials

- Targeted individualized therapies employing precision medicine approaches to optimize therapeutic efficacy while minimizing drug-associated toxicity
- Development of antifibrotic agents, including new approaches to block fibrogenic pathways
- Use of biological agents, including anti-cytokines, anti-chemokines, and growth factor inhibitors

Other Fibrosing Diseases

Several disorders can cause skin fibrosis and mimic SSc. SSc can be distinguished from these SSc-like fibrosing conditions by characteristic clinical, pathological, and laboratory features. The

TABLE 56.6 Differential Diagnosis of Systemic Sclerosis and Scleroderma

Disorders Characterized by Similar Clinical Presentations
- Systemic lupus erythematosus
- Sjögren syndrome
- Rheumatoid arthritis
- Polymyositis/dermatomyositis
- Primary Raynaud's phenomenon

Disorders Characterized by Similar Visceral Features
- Primary pulmonary hypertension
- Primary biliary cirrhosis
- Idiopathic intestinal hypomotility
- Idiopathic pulmonary fibrosis
- Malignant hypertension

Disorders Characterized by Skin Thickening
- Scleromyxedema
- Scleredema (of Buschke), diabetic scleredema
- Nephrogenic fibrosing dermatopathy
- Eosinophilic fasciitis/diffuse fasciitis with eosinophilia
- Eosinophilia–myalgia syndrome
- Generalized morphea
- Chronic graft-versus-host disease
- POEMS syndrome (polyneuropathy, organomegaly, endocrinopathy, monoclonal protein, skin changes)
- Amyloidosis
- Carcinoid syndrome
- Pentazocine-induced scleroderma
- Diabetic digital sclerosis
- Vinyl chloride disease
- Toxic oil syndrome
- Bleomycin exposure
- Werner syndrome
- Phenylketonuria
- Porphyria cutanea tarda
- Vibration white finger syndrome
- Chronic reflex sympathetic dystrophy

most distinguishing clinical features of SSc include the presence of RP, nailfold capillary changes, the characteristic distribution, and characteristic skin changes. The skin pattern in SSc is noted for severe finger, hand, and distal limb involvement and sparing of the skin of the back of the trunk. SSc-specific serum autoantibodies also help define the presence of the SSc disease process. Conditions that mimic SSc include localized forms of scleroderma (see Table 56.5), eosinophilic fasciitis, nephrogenic systemic fibrosis, scleromyxedema (papular mucinosis), scleredema, graft-versus-host disease (GvHD), toxic oil syndrome, and eosinophilia–myalgia syndrome (Table 56.6).

REFERENCES

1. Maricq HR, Weinrich MC, Keil JE, et al. Prevalence of scleroderma spectrum disorders in the general population of South Carolina. *Arthritis Rheum*. 1989;32:998–1006.
2. Mayes MD, Lacey JV Jr, Beebe-Dimmer J, et al. Prevalence, incidence, survival, and disease characteristics of systemic sclerosis in a large US population. *Arthritis Rheum*. 2003;48:2246–2255.
3. Bossini-Castillo L, López-Isac E, Martín J. Immunogenetics of systemic sclerosis: defining heritability, functional variants and shared-autoimmunity pathways. *J Autoimmun*. 2015;64:53–65.
4. Hummers LK. The importance of recognizing scleroderma-type disorders in clinical practice. *Nat Clin Pract Rheumatol*. 2008;4:638–640.
5. Janowsky EC, Kupper LL, Hulka BS. Meta-analyses of the relation between silicone breast implants and the risk of connective-tissue diseases. *N Engl J Med*. 2000;342:781–790.
6. Bhattacharyya S, Wei J, Varga J. Understanding fibrosis in systemic sclerosis: shifting paradigms, emerging opportunities. *Nat Rev Rheumatol*. 2011;8:42–54.
7. Kuwana M, Okazaki Y, Yasuoka H, et al. Defective vasculogenesis in systemic sclerosis. *Lancet*. 2004;364:603–610.
8. Sato S, Fujimoto M, Hasegawa M, et al. Altered blood B lymphocyte homeostasis in systemic sclerosis: expanded naive B cells and diminished but activated memory B cells. *Arthritis Rheum*. 2004;50:1918–1927.
9. Harris ML, Rosen A. Autoimmunity in scleroderma: the origin, pathogenetic role, and clinical significance of autoantibodies. *Curr Opin Rheumatol*. 2003;15:778–784.
10. Mori Y, Chen SJ, Varga J. Expression and regulation of intracellular SMAD signaling in scleroderma skin fibroblasts. *Arthritis Rheum*. 2003;48:1964–1978.
11. Domsic RT, Rodriguez-Reyna T, Lucas M, et al. Skin thickness progression rate: a predictor of mortality and early internal organ involvement in diffuse scleroderma. *Ann Rheum. Dis*. 2011;70:104–109.
12. Pokeerbux MR, Giovannelli J, Dauchet L, et al. Survival and prognosis factors in systemic sclerosis: data of a French multicenter cohort, systematic review, and meta-analysis of the literature. *Arthritis Res Ther*. 2019;21(1):86.
13. Herrick AL, Peytrignet S, Lunt M, et al. Patterns and predictors of skin score change in early diffuse systemic sclerosis from the European Scleroderma Observational Study. *Ann Rheum Dis*. 2018;77(4):563–570.
14. Minier T, Guiducci S, Bellando-Randone S, et al. EUSTAR co-workers. Preliminary analysis of the very early diagnosis of systemic sclerosis (VEDOSS) EUSTAR multicenter study: evidence for puffy fingers as a pivotal sign for suspicion of systemic sclerosis. *Ann Rheum Dis*. 2014;73(12):2087–2093.
15. Herrick AL, Wigley FM. Raynaud's phenomenon. *Best Pract Res Clin Rheumatol*. 2020 Jan 29 [Epub ahead of print].
16. Flavahan NA. A vascular mechanistic approach to understanding Raynaud phenomenon. *Nat Rev Rheumatol*. 2015;11(3):146–158.
17. Miller JB, Gandhi N, Clarke J, McMahan Z. Gastrointestinal involvement in systemic sclerosis: an update. *J Clin Rheumatol*. 2018;24(6):328–337.
18. Distler O, Highland KB, Gahlemann M, et al. Nintedanib for systemic sclerosis-associated interstitial lung disease. *N Engl J Med*. 2019;380(26):2518–2528.
19. Panopoulos S, Bournia VK, Konstantonis G, et al. Predictors of morbidity and mortality in early systemic sclerosis: long-term follow-up data from a single-centre inception cohort. *Autoimmun Rev*. 2018;17(8):816–820.
20. Zanatta E, Polito P, Favaro M, et al. Therapy of scleroderma renal crisis: state of the art. *Autoimmun Rev*. 2018;17(9):882–889.
21. Paik JJ. Muscle disease in scleroderma. *Curr Opin Rheumatol*. 2018;30(6):576–580.
22. Paik JJ, Wigley FM, Shah AA, et al. Association of fibrosing myopathy in systemic sclerosis and higher mortality. *Arthritis Care Res (Hoboken)*. 2017;69(11):1764–1770.
23. Asano Y, Varga J. Rationally-based therapeutic disease modification in systemic sclerosis: Novel strategies. *Semin Cell Dev Biol*. 2019 Dec 16. [Epub ahead of print].
24. Tyndall A. Hematopoietic stem cell transplantation for systemic sclerosis: review of current status. *BioDrugs*. 2019;33(4):401–409.
25. Tyndall AJ, Bannert B, Vonk M, et al. Causes and risk factors for death in systemic sclerosis: a study from the EULAR Scleroderma Trials and Research (EUSTAR) database. *Ann Rheum Dis*. 2010;69(10):1809–1815.
26. Goh NS, Desai SR, Veeraraghavan S, et al. Interstitial lung disease in systemic sclerosis: a simple staging system. *Am J Respir Crit Care Med*. 2008;177(11):1248–1254.
27. Khanna D, Tashkin DP, Denton CP, et al. Aetiology, risk factors, and biomarkers in systemic sclerosis with interstitial lung disease. *Am J Respir Crit Care Med*. 2019 Dec 16. [Epub ahead of print].

Inflammatory Muscle Diseases

Arash H. Lahouti and Lisa Christopher-Stine

The idiopathic inflammatory myopathies (IIMs) are a group of rare systemic diseases. They consist of polymyositis (PM), dermatomyositis (DM), inclusion body myositis (IBM), and more recently described immune-mediated necrotizing myopathy. Traditionally, they are believed to be autoimmune diseases, although recent studies suggest that there is a close resemblance between IBM and other neurodegenerative diseases. Similar to other autoimmune diseases, patients with IIM often have autoantibodies in their serum. Some of these antibodies are specific for myositis and are not seen in other rheumatic disorders. Our knowledge of the myositis-specific antibodies is not comprehensive but new antibodies are being discovered constantly. Each myositis-specific antibody is closely linked to a unique clinical phenotype. Thus, these antibodies have value in classifying patients with IIMs. They can forewarn physicians of particular extramuscular manifestations and guide them toward initiating appropriate treatments.

The IIMs present with muscle weakness and elevated muscle enzymes. In patients with DM, skin manifestations may be the initial presentation. The differential diagnosis of the IIMs is broad and includes drug-induced myopathies, neuromuscular disorders, muscular dystrophies, and metabolic and endocrine myopathies. To add to the complexity of the classification, some DM patients do not develop muscle weakness, referred to as clinically amyopathic DM. The age and gender of the patient, the pattern of the weakness, the severity of the manifestations, and the associated symptoms usually aid in suspecting the correct diagnosis. For example, IBM is distinguished from other IIMs by a characteristic involvement of the finger flexor and knee extensor muscles, which is often asymmetrical. IBM is more common in older men, whereas PM and DM are commonly seen in young to middle-aged women and children. Dermatomyositis is often associated with characteristic skin findings, which are not a feature of PM and IBM. However, differentiation between the IIMs cannot be made on clinical grounds in a substantial number of patients, and further diagnostic testing is required. For example, both PM and immune-mediated necrotizing myopathy present in a same fashion with a predominant symmetrical proximal muscle weakness and elevated muscle enzymes. They can be distinguished only by pathological examination. On muscle biopsy, immune-mediated necrotizing myopathy is associated with necrosis and regeneration of muscle fibers and a characteristic sparse inflammatory infiltrate. In contrast, PM is associated with the presence of cytotoxic inflammatory cells surrounding and invading muscle fibers. Electromyography is a valuable tool for differentiating between weakness originating from muscle rather than peripheral nerves. Magnetic resonance imaging (MRI) can be extremely helpful to identify inflammatory changes in patients with subtle clinical muscle involvement. Moreover, MRI can help to better estimate the burden of the disease and differentiate acute changes (edema) from chronic changes (atrophy). The role of new imaging modalities such as whole-body MRI needs to be further investigated.

Interstitial lung disease, gastrointestinal involvement, and arthritis are among the most common extramuscular manifestations of the IIMs. Interstitial lung disease commonly occurs as part of the antisynthetase syndrome in a subset of patients who have antisynthetase autoantibodies. Gastrointestinal manifestations include dysphagia and aspiration pneumonia. Dysphagia is particularly common and can be seen in all forms of the IIM. In addition, certain forms of the IIMs, particularly DM, can be a paraneoplastic phenomenon. The most common cancers associated with myositis include gynecological (ovarian), pulmonary, gastrointestinal (pancreatic, stomach, and colorectal), and non-Hodgkin lymphoma. Also, the IIMs may be associated with other autoimmune diseases such as systemic lupus erythematosus, Sjögren syndrome, and systemic sclerosis.

Muscle biopsy is critical for the diagnosis of IIMs. The presence of perifascicular atrophy is strongly suggestive of DM, whereas the finding of rimmed vacuoles in the appropriate context suggests IBM. Muscle biopsy may also help differentiate between the IIMs and other forms of myopathy presenting as myositis clinical mimics such as drug-induced myopathies and muscular dystrophies.

Treatment of the IIMs is based largely on experience. Corticosteroids remain the mainstay of treatment. Steroid-sparing agents such as azathioprine, methotrexate, mycophenolate mofetil, and hydroxychloroquine are frequently initiated at presentation while a steroid taper is attempted. In patients with refractory disease, rituximab, intravenous immunoglobulin, and biological medications may be tried. Inclusion body myositis is the most resistant subset of the IIMs.

Finally, an iatrogenic form of myositis has recently been described, precipitated by treatment with immune checkpoint inhibitors (ICS). ICS are novel immunotherapy antineoplastic agents that take advantage of blocking negative T-cell co-stimulation to unleash potent anti-tumor immune responses. Unfortunately, removing such immunologic checkpoints can also precipitate a de novo myositis as well as the exacerbation or relapse of pre-existing myositis.

CLINICAL FEATURES

The clinical hallmark of PM, DM, and IMNM is the gradual onset of symmetrical proximal muscle weakness over weeks to months. In some cases, myalgia may be the presenting or most bothersome symptom, but more often the patient is evaluated for the physical limitations imposed by weakness: difficulty arising from a low chair or bed or combing and brushing hair. Rash is the first feature in a considerable proportion of patients who have DM, but muscle weakness usually follows within a few months. A subset of patients with DM, clinically amyopathic dermatomyositis (C-ADM), may present only with rash in the absence of muscle weakness throughout the course of their illness. These patients are also at risk for pulmonary involvement, as are those with classic DM. The prevalence of interstitial pneumonitis in C-ADM can approach 5% to 10%, compared with 40% of patients with classic DM.[1] Arthritis, Raynaud phenomenon, fever, or lung disease presenting as cough or dyspnea may dominate the clinical picture. Cardiac and gastrointestinal symptoms, other than dysphagia in severe cases, are rarely early manifestations. Renal and central nervous system (CNS) involvement are almost never a part of the IIMs.

KEY CONCEPTS

Definition and Incidence of Idiopathic Inflammatory Myopathies

- Polymyositis (PM), dermatomyositis (DM), and related inflammatory muscle diseases are called IIM.
- Indistinguishable muscle inflammation may accompany other autoimmune connective tissue diseases or limb-girdle muscular dystrophies.
- The annual incidence in the US is 5–10 cases per million. DM and PM are more common in women than in men in all age groups; inclusion body myositis (IBM) is more common in men.

Some of the rashes of DM are virtually pathognomonic; others are not disease-specific (Fig. 57.1). The heliotrope rash, a violaceous discoloration of the eyelids, is sometimes no more than a line along the margin of the upper lid, but it may also affect both upper and lower lids completely and can be associated with edema mimicking thyroid disease. A reddish, sometimes raised and/or scaly, eruption over the metacarpophalangeal joints is known as Gottron papules. In some cases, the metatarsophalangeal joints, elbows, knees, and malleoli show a similar rash. Both heliotrope and Gottron rashes can occur rarely in cases of frank systemic lupus erythematosus (SLE) without muscle involvement. Other common rashes include a flat, red blanching eruption of the upper chest (often in a V distribution), the upper back (where a shawl would touch), and sometimes the extensor surfaces of the upper arms and thighs. Another rash that mimics the malar rash of lupus on the face may be present; however, in contrast to lupus, it does not spare the nasolabial folds. Although found on sun-exposed parts of the body, these rashes are often not photosensitive in nature. As in other connective tissue diseases, nailfold capillary dilatation, infarcts, and cuticular overgrowth occur. Mechanic's hands, a roughening and cracking of the radial sides of the fingers and the palm, resembling a condition found in people who labor with their hands, is characteristic of a subset of myositis patients with the "antisynthetase syndrome" and can also be seen in patients with PM-Scl and U1-RNP autoantibodies.

CLASSIFICATION

In the past 40 years, several investigators have proposed diagnostic classification criteria for IIM. The criteria proposed by Bohan and Peter four decades ago remain the most familiar and clearest definitions of PM and DM (Table 57.1).[2,3] They combine clinical, laboratory, electrodiagnostic, and pathological features. These criteria, however, are limited by their poor specificity in distinguishing PM from other entities, including late-onset muscular dystrophies. The resultant misclassification limits the homogeneity of the patients included in previous observational and interventional studies. Additionally, the Bohan and Peter criteria completely omit the diagnosis of IBM, the most frequent type of IIM in patients over 50 years of age.

In 2004, more comprehensive criteria of IIM were proposed and approved by a group of international experts (Table 57.2). Unlike the previous criteria, these new criteria offer the advantage of classifying two rare forms of autoimmune myositis, *i.e.*, IMNM and DM *sine* dermatitis, as separate categories (see Table 57.2).[4] Separate classification criteria systems for IBM have been devised. According to most of the previous IBM criteria, characteristic muscle biopsy changes were necessary for a definite diagnosis of IBM. However, because selective involve-

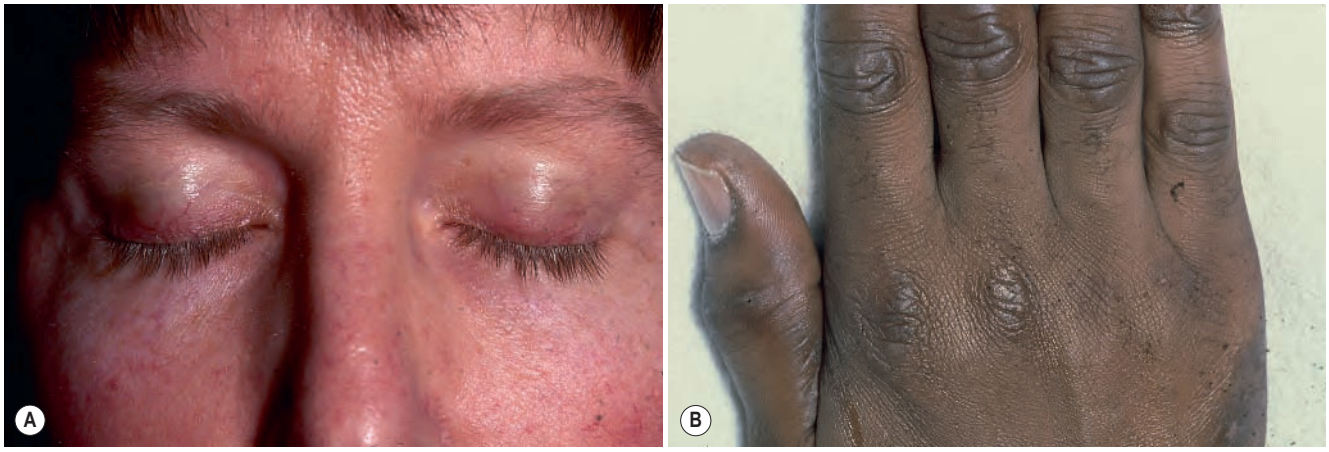


FIG. 57.1 Dermatomyositis Rash. (A) In addition to the heliotrope rash on the eyelids of this patient with dermatomyositis, there is a flat, red rash on the nose and cheeks. (B) A raised, shiny-red rash—Gottron papules—is apparent on the interphalangeal and the second and third metacarpophalangeal joints of this man with dermatomyositis.

TABLE 57.1 Idiopathic Inflammatory Myopathy: Diagnostic Criteria

Bohan and Peter Criteria

1. Symmetrical proximal muscle weakness
2. Skeletal muscle enzyme elevation
3. Abnormal EMG[a]
4. Muscle biopsy abnormalities
5. Typical skin rash of DM[b]

[a]Polyphasic, short, small motor-unit potentials; fibrillation, positive sharp waves, increased insertional irritability; bizarre, high-frequency, repetitive discharges.
[b]Gottron sign, heliotrope rash.
Possible PM = any two of the first four criteria; possible DM = criterion 5 (rash) + any two criteria.
Probable PM = any three of the first four criteria; probable DM = criterion 5 (rash) + any three criteria.
Definite PM = all four of the first four criteria; definite DM = criterion 5 (rash) + all four other criteria.
DM, Dermatomyositis; *EMG*, electromyography; *PM*, polymyositis.

ment of finger flexors and knee extensors is almost exclusive to IBM among other forms of myopathy, the newly proposed *ENMC IBM Research Diagnostic Criteria 2011* entails a separate category for clinically defined IBM (Table 57.3).[5]

It has been useful for some purposes to divide cases into groups: PM, DM, juvenile myositis, myositis associated with another connective tissue disease (usually systemic sclerosis, SLE, or Sjögren syndrome), cancer-associated myositis (usually cases in which the diagnoses are made within 6 to 12 months of one another), IBM, and a miscellaneous group that includes such rare entities as eosinophilic myositis (Table 57.4). This classification has allowed recognition of unique clinical and pathogenetic features and response to therapy. In the case of cancer-associated myositis, a more rational approach to workup based on recognition of groups at risk is now possible.

In 2017 the European League Against Rheumatism and American College of Rheumatology (EULAR ACR) Diagnostics Classification Criteria were published, becoming the acceptable gold standard criteria for classification for clinical trial purposes.[6] These criteria utilized data from 976 IIM patients (74% adults; 26% children) and 624 non-IIM patients with mimicking conditions (82% adults; 18% children). Each item was assigned a weighted score. The total score corresponds to a probability of having IIM. Subclassification is then performed using a classification tree. These new criteria addressed the addition of IBM, juvenile myositis, and amyopathic dermatomyositis. However, it had too few cases included for adjudication to add immune-mediated necrotizing myopathy (IMNM) as a separate diagnostic category from polymyositis. In addition, anti-Jo-1 is the only included myositis-specific autoantibody (MSA), and interstitial lung disease is not included among the clinically weighted items, effectively eliminating any patient with the antisynthetase syndrome in the absence of muscle or skin disease. Thus, more sophisticated big data modeling is likely to lead to more robust myositis classification criteria in the future.

KEY CONCEPTS

Characteristic Hallmarks of Inflammatory Myositis

- The clinical hallmark is proximal limb and neck weakness, rarely associated with muscle pain.
- The laboratory hallmarks are elevated serum levels of creatine kinase (CK), aldolase, lactic dehydrogenase, and the transaminases; a characteristic pattern ("irritable myopathy") is seen on electromyography (EMG). Elevated serum levels of autoantibodies are common.
- The pathological hallmarks are focal muscle necrosis, degeneration, regeneration, and inflammation.

TABLE 57.2 Proposed Diagnostic Criteria for Polymyositis and Dermatomyositis

The European Neuromuscular Centre (ENMC) criteria for the idiopathic inflammatory myopathies[4]

1. Clinical criteria
   a. Subacute onset
   b. Age >18 years (onset may be in childhood in DM and nonspecific myositis)
   c. Symmetrical proximal weakness
   d. Typical DM rash (including heliotrope, Gottron papules, Gottron sign, V sign, and shawl sign)
   e. Lack of features suggestive of IBM (asymmetry, finger flexor ≥ deltoid weakness, and knee extensors/ankle dorsiflexors ≥ hip flexors weakness), toxic myopathies, endocrine myopathies, amyloidosis, family history of muscular dystrophy or proximal motor neuropathies (*e.g.*, SMA)
2. Elevated serum creatine kinase level
3. Other laboratory criteria
   a. Abnormal electromyography
      Inclusion criteria
      I. Increased insertional and spontaneous activity in the form of fibrillation potentials, positive sharp waves, or complex repetitive discharges
      II. Morphometric analysis reveals the presence of short duration, small amplitude, polyphasic motor-unit action potentials (MUAPs)
      Exclusion criteria
      I. Myotonic discharges that would suggest proximal myotonic dystrophy or other channelopathy
      II. Morphometric analysis reveals predominantly long duration, large-amplitude MUAPs
      III. Decreased recruitment pattern of MUAPs
   b. MRI: diffuse or patchy increased signal (edema) within muscle tissue on Short T1 Inversion Recovery (STIR) images
   c. Myositis-specific antibodies detected in serum
4. Abnormal muscle biopsy
   a. Endomysial inflammatory cell infiltrate surrounding and invading nonnecrotic muscle fibers
   b. Endomysial CD8 T cells surrounding, but not definitely invading, nonnecrotic muscle fibers; or ubiquitous MHC-1 expression
   c. Perifascicular atrophy
   d. Membrane attack complex (MAC) depositions on small blood vessels, or reduced capillary density, or tubuloreticular inclusions in endothelial cells on EM, or MHC-1 expression of perifascicular fibers
   e. Perivascular, perimysial inflammatory cell infiltrate
   f. Scattered endomysial CD8 T-cell infiltrate that does not clearly surround or invade muscle fibers
   g. Many necrotic muscle fibers as the predominant abnormal histological feature. Inflammatory cells are sparse or only slight perivascular; perimysial infiltrate is not evident
   h. Rimmed vacuoles, ragged red fibers, cytochrome oxidase–negative fibers that would suggest IBM
   i. MAC deposition on the sarcolemma of nonnecrotic fibers and other indications of muscular dystrophies with immunopathology

Polymyositis—*Definite polymyositis:* (1) All clinical criteria with the exception of rash, (2) elevated serum creatine kinase (CK), (3) muscle biopsy criteria include (a) and exclude (c, d, h, and i); *probable polymyositis:* (1) All clinical criteria with the exception of rash, (2) elevated serum CK, (3) other laboratory criteria (1 of 3), (4) muscle biopsy criteria include (b) and exclude (c, d, g, h, and i). **Dermatomyositis**—*Definite dermatomyositis:* (1) All clinical criteria, (2) muscle biopsy criteria include (c); *probable dermatomyositis:* (1) All clinical criteria, (2) muscle biopsy criteria include (d) or (e), or elevated serum CK, or other laboratory criteria (1 of 3). **Amyopathic dermatomyositis**—(1) Rash typical of DM: heliotrope, periorbital edema, Gottron papules/sign, V sign, shawl sign, holster sign, (2) skin biopsy demonstrates a reduced capillary density, deposition of MAC on small blood vessels along the dermal-epidermal junction, and variable keratinocyte decoration for MAC, (3) no objective weakness, (4) normal serum CK, (5) normal electromyogram, (6) muscle biopsy, if done, does not reveal features compatible with definite or probable DM. **Possible dermatomyositis sine dermatitis**—(1) All clinical criteria with the exception of rash, (2) elevated serum CK, (3) other laboratory criteria (1 of 3), (4) muscle biopsy criteria include (c) or (d). **Nonspecific myositis**—(1) All clinical criteria with the exception of rash, (2) elevated serum CK, (3) other laboratory criteria (1 of 3), (4) muscle biopsy criteria include (e) or (f) and exclude all others. **Immune-mediated necrotizing myopathy**—(1) All clinical criteria with the exception of rash, (2) elevated serum CK, (3) other laboratory criteria (1 of 3), (4) muscle biopsy criteria include (g) and exclude all others.

TABLE 57.3 ENMC IBM Research Diagnostic Criteria 2011[5]

Clinical and Laboratory Features	Pathological Features
a. Duration >12 months	I. Endomysial inflammatory infiltrate
b. Age at onset >45 years	II. Rimmed vacuoles
c. CK no greater than 15 × ULN	III. Protein accumulation* or 15–18-nm filaments
d1. Knee extension weakness > hip flexion weakness	IV. Upregulation of MHC class I
d2. Finger flexion weakness > shoulder abduction weakness	

*Demonstration of amyloid or other protein accumulation by established methods (*e.g.*, for amyloid Congo red, crystal violet, thioflavin T/S, for other proteins p62, SMI-31, TDP-43). Current evidence favors p62 in terms of sensitivity and specificity, but the literature is limited, and further work is required.
Clinicopathologically defined IBM: Clinical criteria include (a–c) and at least one of the (d) criteria plus the first three pathological features. **Clinically defined IBM:** All clinical criteria plus one or more pathological features. **Probable IBM:** Clinical criteria include (a–c) and at least one of the (d) criteria plus one or more pathological features.
CK, Creatine kinase; *IBM*, inclusion body myositis; *ULN*, upper limit of normal.

TABLE 57.4 Traditional Classification of Idiopathic Inflammatory Myopathies

Type I	Primary idiopathic polymyositis
Type II	Primary idiopathic dermatomyositis
Type III	Dermatomyositis or polymyositis associated with malignancy
Type IV	Childhood dermatomyositis or polymyositis
Type V	Myositis associated with another connective tissue disease
Type VI	Inclusion body myositis
Type VII	Miscellaneous: eosinophilic myositis, localized nodular myositis, etc.

Several autoantibodies, called "myositis-specific autoantibodies," are unique to myositis. These have allowed a useful alternative classification (Table 57.5). For example, patients with antibodies to the aminoacyl-tRNA (transfer RNA) synthetases, of which Jo-1 is the best known, have a characteristic syndrome called antisynthetase syndrome, which usually includes interstitial lung disease, nondeforming inflammatory arthritis, fevers, mechanic's hands, and Raynaud phenomenon, in addition to myositis. Those with antibodies to the signal recognition particle (anti-SRP) have an IMNM manifested by severe disease of abrupt onset, often in the autumn. According to some studies, cardiac involvement is less common and survival is better in patients with anti-SRP than has previously been reported.[7] Those with antibodies to the nuclear antigen Mi-2 almost always have the V and shawl rashes and cuticular overgrowth in addition to myositis. A recently discovered antibody, anti-3-hydroxy-3-methylglutaryl-coenzyme A reductase (anti-HMGCR), is tightly linked to IMNM.[8,9] Interestingly, the HMGCR is the pharmacological target of statin drugs, such as atorvastatin, and a subset of patients with anti-HMGCR antibodies develop this antibody in the context of previous statin exposure, although it is also seen in statin-naïve individuals, albeit to a lesser extent. Patients with anti-HMGCR antibodies present with profound proximal muscle weakness and a marked creatine kinase (CK) elevation. Further, compared with statin-exposed individuals, statin-naïve anti-HMGCR positive patients tend to be younger, have higher CK levels, and have a blunted response to immunosuppressive medications (see Table 57.5).[2,9]

TABLE 57.5 Clinical Features Associated with Myositis-Specific Autoantibodies

Autoantibodies	Association	Characteristic Clinical Features
Anti-Jo-1 and other antisynthetases	PM, DM, IMNM	Relatively acute onset of myositis, frequent interstitial lung disease, fever, Raynaud phenomenon, arthritis, mechanic's hands, moderate response to therapy, persistent disease. Patients sometimes meet criteria for SLE or RA, but muscle disease or lung disease dominate the clinical picture and prognosis.
Anti-SRP	IMNM	Very acute onset of myositis, often in autumn, severe weakness, no rash, palpitations, females predominate, poor response to therapy
Anti-Mi-2	DM	Relatively acute onset of myositis, classic dermatomyositis rashes with V sign and shawl sign, cuticular overgrowth, good response to therapy
Anti-HMGCR	IMNM	Necrotizing myopathy, may be preceded by statin therapy, very high CK levels, minimal muscle wasting
Anti-MDA5	DM, clinically amyopathic DM	Clinically amyopathic dermatomyositis with rapidly progressive interstitial lung disease; characteristic skin findings include cutaneous ulcers and palmar papules
Anti-TIF1-γ (p155/140)	DM	Juvenile dermatomyositis and cancer-associated dermatomyositis
Anti-SAE	DM	Severe skin manifestations and dysphagia
Anti-NXP-2	DM	Associated with calcinosis and muscle contractures in children

DM, Dermatomyositis; *IMNM*, immune-mediated necrotizing myopathy; *MDA5*, melanoma differentiation–associated protein 5; *NXP-2*, nuclear matrix protein; *PM*, polymyositis; *RA*, rheumatoid arthritis; *SAE*, small ubiquitin-like modifier activating enzyme; *SLE*, systemic lupus erythematosus; *SRP*, signal recognition particle; *TIF1-γ*, transcription intermediary factor 1-γ.

IBM is different from other inflammatory myopathies. Patients with IBM rarely improve in strength with immunosuppressive therapy. They tend to be older, and in contrast to patients with PM and DM, who are predominantly women, patients with IBM are more commonly men. They have gradual, painless, asymmetrical, progressive weakness and focal atrophy that develop over years, and they may complain of frequent falls. Two decades after onset, they are frequently wheelchair-bound. The forearms of these patients exhibit a scalloped appearance, attributed to muscle atrophy. Difficulty with swallowing can occur with any inflammatory myopathy and is frequently a major problem in patients with IBM. The CK and other skeletal muscle–associated serum enzymes are normal in about one-quarter of patients with IBM and only moderately elevated in the remainder. The electromyogram in IBM frequently demonstrates both myogenic and neurogenic

features secondary to the effective denervation of some muscle cells by inflammation and necrosis.

Proposed criteria for the diagnosis of IBM rely on both pathological and clinical features. In the clinical setting of an inflammatory myopathy, the presence of the characteristic inclusions or rimmed vacuoles is diagnostic. Among the inflammatory myopathies, IBM is distinguished by substantial numbers of rimmed cytoplasmic vacuoles with tubulofilamentous material within myofibers. A variety of proteins have been found by immunohistochemistry in the muscle cells in IBM, including ubiquitin, β-amyloid precursor protein, and the transcription factor nuclear factor kappa B (NF-κB). Many of the proteins accumulating in the IBM muscle are also involved in other neurodegenerative diseases. For example, aggregation of p62, an autophagy-related protein, has been shown in IBM muscle, in Lewy bodies in Parkinson disease, and in neurofibrillary tangles in Alzheimer disease, suggesting the possibility of a degenerative process in IBM. More recently, an autoantibody to cytosolic 5′-nucleotidase 1a (c5N1A) was discovered. Although this antibody has a good specificity to distinguish IBM against other forms of autoimmune myopathy, it is positive in only 60% of IBM patients.[10] In addition, anti-c5N1A can be detected in other connective tissue diseases (such as lupus, DM, and Sjögren syndrome), as well as in healthy subjects.[11]

IBM must be distinguished from other chronic myopathies. These include acquired myopathies, such as those caused by toxins, and genetically determined myopathies, such as some muscular dystrophies and the metabolic myopathies. There are several important differences between IBM and these other myopathies. The distinction, however, is not as well defined as might be expected. Although some of the familial forms of IBM have a distinctive clinical presentation, often early in life, there have been several families with the typical late onset and inflammatory picture of the presumed sporadic cases. Several genetic loci have been identified in familial IBM, so it will be important to assess any identified mutations in familial IBM–associated genes in sporadic cases.[12]

ETIOLOGY

Immunological Clues to Origin

The implications of myositis-specific autoantibodies that bind to and inhibit the function of native human enzymes involved in the formation of new proteins are tantalizing and probably significant. These autoantibodies appear to develop before symptomatic weakness or serum CK elevation, suggesting a close link to an initiating factor. Some of the clinical features, such as fever, arthritis, and lung disease, and the apparent seasonality in onset of disease in patients with anti-SRP, are reminiscent of some viral infections. In patients with autoantibodies against one of the aminoacyl-tRNA synthetases, such as Jo-1 (histidyl-tRNA synthetase), the possible connection to a viral inciting agent appeared compelling, as certain picornaviruses, which are closely related to viruses long suspected of causing myositis, such as coxsackie viruses, can mimic tRNA in acting as a substrate for an aminoacyl-tRNA synthetase.[13] Direct proof connecting picornaviruses to human myositis, however, has not been obtained.

In individuals infected with HIV or human T-cell lymphotropic virus (HTLV)-1, the development of IBM has been reported. However, it is not always the first manifestation in these cases. There is no evidence of viral replication within the muscles, but instead the chronic infection triggers an inflammatory response.[14]

Drugs and Toxins

A large number of environmental agents have been associated with myopathies. Drug-induced myopathy should be considered particularly in cases where no other cause has been identified. Sometimes the illness strikingly resembles the spontaneous disease. D-penicillamine, for example, can induce a variety of autoimmune phenomena, including an inflammatory myopathy that closely resembles PM. A number of drugs can also produce myopathies that can be clinically confused with IIM but are histologically distinct. This large group includes 3-hydroxy-3-methylglutaryl-coenzyme A (HMG-coA) reductase inhibitors, corticosteroids, colchicine, and zidovudine (AZT).

In corticosteroid-induced myopathy, type II muscle fiber atrophy is prominent on muscle biopsy, and weakness improves when the dose is lowered. Colchicine can cause myopathy and painful neuromyopathy. The CYP3A4 system metabolizes colchicine, and taking another drug metabolized by the same pathway can result in myopathy.[15] Muscle biopsy shows autophagic vacuoles that stain for acid phosphatase. Discontinuation of colchicine usually results in improvement. AZT produces a characteristic mitochondrial myopathy. The myopathy associated with HIV infection can be distinguished from that caused by AZT by muscle biopsy, as characteristic mitochondrial abnormalities are found in the latter. Amiodarone rarely causes proximal and distal muscle weakness, or the accompanying tremor or distal sensory loss. Muscle biopsy reveals autophagic vacuoles with myeloid inclusions and debris. This is seen more commonly in patients with chronic kidney disease. The antimalarial drugs chloroquine and hydroxychloroquine produce a vacuolar myopathy, possibly by raising the intralysosomal pH so that the acid cathepsins that digest waste products in the lysosome are inoperable, so waste products accumulate in vacuoles.

Immune-checkpoint inhibitors (ICIs) are a class of novel immunotherapeutic antineoplastic agents that halt inhibitory immune-checkpoint pathways. Best known are anti-programmed death-1 (PD-1) antibodies, anti-programmed death-ligand 1 (PD-L1) antibodies, and anti-cytotoxic T-lymphocyte-associated protein 4 (CTLA-4) antibodies. While their ability to reactivate immune responses against cancer is novel and promising, inappropriate immune activation by ICIs can lead to adverse effects, termed immune-related adverse events (irAEs). The incidence of irAEs is reported to be between 17.1% and 72%, depending on the type and dose of the ICI, as well as the cancer phenotype.[16–18] While such irAEs can affect any organ, rheumatic disease complications are observed, probably a reflection of a pathogenic role of immune checkpoints in rheumatic diseases.[19,20] Because ICIs block negative T-cell co-stimulation, immunotherapy leads to an enhanced immune response, including autoimmune responses. Thus, many de novo rheumatologic manifestations as well as the exacerbation or relapse of pre-existing rheumatic diseases have been reported during and after ICI therapy.[21] A recent meta-analysis utilizing 5560 patients noted the incidence of myositis was rare at 0.22%.[22]

Bacterial and Parasitic Diseases

Certain parasitic diseases can produce an illness by direct invasion of muscle. Weakness, fever, and eosinophilia are usually present. Bacterial pyomyositis is uncommon in North America

but occurs more commonly in other parts of the world. It is attended by the signs of local infection and is often asymmetrical. A recent report, however, details two cases of pyomyositis initially diagnosed as PM. Patients presenting with disseminated pyomyositis may be difficult to distinguish clinically from those with IIM, especially in immunosuppressed individuals, who may not mount a systemic response.[23]

PATHOGENESIS

The precise mechanisms of cell damage and death in the IIMs are still unknown. T lymphocytes invading muscle fibers are predominantly lacking the costimulatory receptor CD28, indicative that they are chronically stimulated and end-differentiated. $CD28^{null}$ lymphocytes are found in both muscle tissue and peripheral blood. They produce proinflammatory cytokines and release perforin and granzyme into targeted muscle cells. B cells and plasma cells are also present in muscle-infiltrating infiltrates. The differentiation of B cells into antibody-producing plasma cells occurs within myositis muscle and is regulated by B cell–activating factor (BAFF). These antibodies are likely directed against autoantigens or cross-reacting viral antigens.

In addition to adaptive immune mechanisms, innate and nonimmune mechanisms also may be responsible for muscle cell damage and death. For example, Toll-like receptor (TLR)-3 and TLR-7 are expressed by infiltrating inflammatory cells and immature myoblasts in myositis muscle. TLR activation leads to production of inflammatory cytokines such as interleukin (IL)-6, thereby triggering or perpetuating inflammatory responses. Classic apoptosis is absent in IIM biopsies, but growing evidence suggests that autophagy could be responsible for myofiber death.

After damage, a muscle fiber, which is a syncytium, can regenerate. Initially, on light microscopy one can appreciate loss of striations leading to a homogeneous appearance, fiber size variation, atrophy, and centralization of the nuclei. In the inflammatory myopathies, cell-mediated cytotoxicity is a primary method of destruction. Macrophages and cytotoxic T cells invade the myofibers. The myocyte cytoplasm close to the invaginated cells appears vacuolated and swollen. Other regions of the same cell may show intense regeneration, as seen histologically by aggregates of nuclei with prominent nucleoli and fiber splitting. Thus degeneration and regeneration can coexist in the same fiber.

Cytokines and chemokines (Chapters 14 and 15), which form an integrated network regulating inflammation, are produced by muscle fibers, immune cells, and endothelial cells. As an example, IL-1 is consistently found in myositis muscle. IL-1 facilitates transmigration of leukocytes by increasing the expression of adhesion molecules on endothelial cells. Further, IL-1 can decrease the proliferation of myoblasts. In some myositis patients, treatment with anakinra, an IL-1 receptor antagonist, results in improvement. This finding further supports a pathogenic role of IL-1. IL-1 acts in synergy with the other proinflammatory cytokines, tumor necrosis factor (TNF) and IL-17, to induce the production of IL-6 and CCL20 by myoblasts. The chemokine CCL20 recruits Th17 lymphocytes and immature dendritic cells into the muscle. IL-6 perpetuates inflammation by suppressing $FOXP3^{+}$ regulatory T cells. More recently, IL-15 has been reported to have a pathogenic role. IL-15 is expressed in muscle cells and myoblasts derived from PM and DM patients. IL-15 regulates T-cell activation and proliferation, and a higher expression of this cytokine at baseline may be associated with a poor response to immunosuppressive therapy.[24]

The role of type I interferons (IFN-Is) in the pathogenesis of DM has been extensively studied. The IFN-Is are a family of highly related glycoproteins, including IFN-α and IFN-β, that act on the type 1 IFN receptor. In DM muscle, the IFN-Is are produced by plasmacytoid dendritic cells (pDCs).[25] After activation of the IFN receptor, a cluster of IFN-I–inducible genes is transcribed, resulting in upregulation of several proteins, also known as the IFN signature, in the muscle and peripheral blood of patients with DM. Well known is the myxovirus resistance protein A (MxA), which, similar to other IFN-I–inducible proteins, has antiviral properties. This protein is overexpressed in both the muscle and skin in DM but not in PM or IBM. Further, decreased expression of MxA mRNA in peripheral blood mononuclear cells may correlate with decreasing muscle symptoms in juvenile DM patients. Indeed, some of the IFN-induced proteins may sustain autoimmunity in DM. For example, retinoic acid-inducible gene 1 (RIG-I) and melanoma differentiation-associated gene 5 (MDA-5) are IFN-I–induced proteins that induce IFN-I production in a positive feedback fashion. Both proteins contribute to IFN induction in response to measles virus infection,[26] nurturing the suspicion for a role of viral infections in myositis. Engel and colleagues[27] have demonstrated several important pathological distinctions between DM and PM and have suggested that in DM, the capillaries are the initial point of injury. A model of disease has been proposed in which primarily capillaries are damaged and myocytes are secondarily involved. Complement and immunoglobulins may be found in the capillary walls even where the remainder of the muscle is normal. The membrane attack complex of complement deposits there, and endothelial cells are swollen and pale. Even in unaffected regions of muscle, special staining reveals a marked decrease in capillary numbers. More advanced changes include microtubular endothelial inclusions and microvacuoles. Evaluation of lymphocyte populations in DM shows a high percentage of B cells and macrophages in the perivascular regions and an increasing frequency of CD4 T cells and pDCs toward the perimysium and endomysium. In DM, especially the juvenile-onset form, inflammation and necrosis followed by atrophy appear in a perifascicular pattern. Perivascular lymphocytic infiltrates, typical of later disease, have not been described as an early change. Consistent with the importance of immune complexes, histological abnormalities in the DM skin are indistinguishable from the changes in lupus.

In contrast to DM, PM and IBM do not demonstrate marked capillary changes, the perivascular infiltrates are less pronounced, and T-cell infiltrates in the perimysial and endomysial regions are more pronounced. Nonnecrotic fibers may be surrounded by T cells and macrophages. T cells are enriched for the CD8 subset. There is negligible evidence of natural killer (NK) cell presence. Should cytotoxic T cells prove to be the major effector cells, cloned T cells could lead to relevant antigenic targets. CD8 T cells recognize their antigenic peptide in association with major histocompatibility complex (MHC) class I molecules. Although resting normal muscle has very low class I expression, it is upregulated in regenerating and degenerating fibers found in both inflammatory and noninflammatory myopathies. Interestingly, in DM, class I expression is upregulated predominantly in the perifascicular regions, around sites of atrophy, and near sites of cellular invasion. In contrast, in PM, class I expression may be diffusely upregulated even where

there is no cellular infiltrate. IBM shows a more focal class I distribution in regions of T-cell invasion. The presence of focal regions of MHC class I expression in nonnecrotic fibers at the site of activated CD8 T cells is compatible with cytotoxicity as a prime mechanism of myocyte necrosis in IBM and PM. A pivotal study using transgenic mice demonstrated that abnormal accumulation of MHC class I molecules in the endoplasmic reticulum (ER) of muscle may initiate the ER stress response.[28]

KEY CONCEPTS

Differential Histological Features of Myositis

- In dermatomyositis the earliest changes involve vessel walls, and B cells and CD4 T cells predominate in the muscle biopsy.
- In polymyositis and inclusion body myositis, the dominant pathological feature is targeting and invasion of muscle cells by CD8 cytotoxic T cells.

Proposed pathogenic mechanisms for the development of the sporadic forms of IBM are complex and include both autoimmunity and degeneration. Many proteins also found in other neurodegenerative diseases have been shown to accumulate in IBM muscle. In addition, IBM patients typically do not respond to immunosuppressive medications. These findings likely suggest that IBM is a degenerative disease. However, IBM is sometimes associated with other autoimmune diseases. In IBM biopsy specimens, inflammatory cells are predominantly composed of CD8 T cells, and a majority of patients express a circulating antibody to cytosolic 5′-nucleotidase 1 A, supporting the role of autoimmunity in the pathogenesis of IBM.

In necrotizing autoimmune myositis (also known as IMNM), there is necrosis of muscle fibers with myophagocytosis and regeneration and paucity of T-cell infiltration. Complement deposition on blood vessels has been reported. In muscle biopsies from patients with statin-associated necrotizing autoimmune myositis, MHC class-I upregulation is frequently seen.[14]

The pathogenic role of autoantibodies found in IIM patients remains uncertain. MSAs are found in 60% to 80% of patients and appear to delineate specific clinical entities; each group has a strong but not absolute human leukocyte antigen (HLA) association. In a patient with myositis and antihistidyl-tRNA synthetase (Jo-1) autoantibodies, sera available from long before the onset of symptoms or biochemical damage to muscle tissue contained the autoantibodies, suggesting that the autoantibodies were not merely a response to release of tissue antigens. The extraordinary specificity of MSAs for IIM and the lack of evidence for strong polyclonal stimulation in these diseases suggest that MSAs are related to the fundamental causative process in IIM.

The structures bound by MSAs are mostly intracellular ribonucleoproteins involved in protein synthesis, such as the aminoacyl-tRNA synthetases and the SRP. These autoantigens are found in every nucleated cell. In general, the antibodies bind to conformational epitopes and, at least in the case of the antisynthetases, block enzymatic activity. It is possible that a structural property of muscle allows these particular proteins to be presented to the immune system when the cells are damaged; alternatively, the capacity of muscle fibers to degenerate alongside intense regeneration within the same fiber may allow these proteins to be efficiently displayed.[29] Experiments have suggested that some aminoacyl-tRNA synthetases have a direct proinflammatory role through a subsidiary chemokine-like action.

A landmark study determined that cultured myoblasts express high levels of autoantigens, which are strikingly downregulated as cells differentiate into myotubes *in vitro*. These data strongly associate regenerating rather than mature muscle cells as the source of continuous autoantigen supply in autoimmune myositis.[29]

GENETICS

The IIMs do not exhibit a simple mode of inheritance, and the rare familial cases mostly reflect IBM of early onset. As noted above, there are HLA associations for particular MSAs. Specifically, HLA-DR52 has a strong association (90%) with antisynthetase-positive myositis in people of both European and African descent.[30] HLA DRB1*11:01 was recently shown to be associated with an increased risk of anti-HMGCR myopathy.[31]

NATURAL HISTORY

The prognosis for patients with IIM varies greatly with clinical type, autoantibodies, extraskeletal muscle involvement, and the interval between diagnosis and the start of treatment.

Patients with DM or myositis accompanying another connective tissue disease are likely to recover most of their strength with prompt and adequate therapy. Although recurrences are common, persistent profound weakness does not usually occur. Most patients with anti-Mi-2 autoantibodies also usually respond well to therapy. Strength usually recovers well in patients whose myositis is cancer related, but overall mortality due to the tumor is high. Indeed, an accompanying tumor remains one of the most frequent causes of death in patients with an IIM. Among the MSAs, anti-TIF-1γ and anti–NXP-2 antibodies are found with increased frequency in patients with cancer-associated DM.[32]

Patients with PM fare less well, even when those with IBM are rigorously excluded. A return to normal strength is very unusual, and each recurrence is likely to be followed by greater residual weakness, even if inflammation is fully controlled. IBM has a poorer prognosis, but it is possible that the gradual decline in strength can be halted for long periods by corticosteroid and/or cytotoxic therapy if continuing inflammation is present. Severe muscle weakness and atrophy and very high CK levels are prominent features in patients with anti-SRP autoantibodies. Patients with anti-MDA5 antibodies are at increased risk for developing progressive interstitial lung disease. Those with anti-Jo-1 autoantibodies or antibodies to another synthetase are likely to respond to therapy initially, but they typically require continuing immunosuppression to treat frequent recurrences. In this group morbidity and mortality are heavily influenced by the progression of lung involvement. Longitudinal outcome studies in DM and PM are few. Cardiac involvement, respiratory involvement, and cancer were the main causes of death in several cohort analyses. Disease course is monocyclic in approximately 20% of patients, polycyclic in 20%, and chronic in the remainder. Relapses have been noted in the initial years of therapy and after prolonged disease-free intervals; therefore, periodic surveillance is warranted for at least 2 years after remission.

PATIENT MANAGEMENT

The treatment of myositis is based on controlling skeletal muscle inflammation and damage. Immunosuppressive therapy is used

in the initial stages of the disease to reduce inflammation and muscle damage. There are very few randomized controlled trials of any of the immunosuppressive agents used; thus, therapeutic regimens and responses have remained largely anecdotal. After the initial inflammation is controlled, strengthening exercises are useful in improving functional capabilities.

Corticosteroids

Corticosteroids are the main immunosuppressive agents used in myositis treatment. An initial course of pulses of methylprednisolone may be helpful, particularly in disease of acute onset, and may also be helpful in managing disease flares. If active muscle inflammation persists or the side effects of corticosteroids are severe, other immunosuppressive treatments are employed.

Second-Line and Third-Line Immunosuppressive Therapies

The most frequently used second-line agents in the treatment of myositis are azathioprine and methotrexate. Azathioprine reduces long-term disability. Methotrexate is useful in patients with little or no response to corticosteroid therapy. Combination therapies, such as methotrexate with azathioprine, are useful even if patients have failed to respond to one of the agents alone. High-dose intravenous immunoglobulin (IVIG) is of proven benefit in DM.[33] Whereas patients with statin-associated anti-HMGCR myopathy may be particularly responsive to IVIG,[34] its usefulness in PM is less predictable. Apheresis proved ineffective in a controlled blinded study.[35] Both cyclosporine and tacrolimus have been effective in some cases, as have cyclophosphamide and chlorambucil. The most recent therapeutic options include mycophenolate and rituximab. Recently a large, randomized placebo-phase clinical trial was conducted to elucidate the role of a 44-week course of rituximab therapy in adults and children with refractory PM and DM. In general, 83% of refractory myositis patients met the Definition of Improvement (DOI), and the presence of antisynthetase and anti–Mi-2 autoantibodies was associated with a shorter time to improvement.[36,37] In refractory patients who have not responded to the above-mentioned treatments, biological agents such as etanercept, adalimumab, anakinra, abatacept, and high-dose cyclophosphamide have been tried with variable success.[24,38–41] Thus far, there has been no effective therapeutic regimen for IBM. However, IVIG has been reported to provide a transient response.[14]

Monitoring Disease Activity

Improvement in strength and normalization of serum CK activity are the best indirect measures of disease activity. A decrease in serum CK activity may herald clinical improvement, but corticosteroid treatment alone can reduce CK activity without associated clinical improvement. A lack of improvement in strength in a corticosteroid-treated patient may be due to the resistance of the inflammatory process, the presence of a corticosteroid-induced myopathy, muscle atrophy, and/or misdiagnosis. A diagnostic and therapeutic taper of the corticosteroids may then be warranted. If inflammation is present concurrently, other immunosuppressive agents are useful as the dosage of corticosteroids is lowered. If the CK value begins to rise, even if it is still within the normal range, and the symptoms of myositis are worsening in a patient whose disease has previously been controlled with corticosteroids, an increase in the dose may be warranted.

Treatment-Resistant Myositis

Some treatment-resistant PM patients have another disease. In such cases IBM or a limb-girdle muscular dystrophy should be suspected. Unlike other myositis patients, those with IBM rarely, if ever, improve in strength with immunosuppressive therapy, but stabilization of strength can be achieved in some IBM patients with immunosuppressive agents.[42,43] Patients with limb-girdle muscular dystrophies may mimic PM clinically. They may have inflammation on muscle biopsy and may occasionally have associated autoantibodies. Thus, patients with a suspected IIM who do not respond to immunosuppressive therapy should undergo further evaluation, including genetic testing, to search for a limb-girdle muscular dystrophy.

Nonskeletal Muscle Involvement

Other organs frequently involved in myositis include the skin, lungs, and joints. Such organ involvement and the systemic features of myositis (fever and weight loss) usually improve with immunosuppressive therapy that controls inflammation in the skeletal muscle. Hydroxychloroquine and other antimalarials are useful in controlling the rashes associated with myositis.

DIAGNOSTIC TOOLS, EVALUATION, AND DIFFERENTIAL DIAGNOSIS

CLINICAL PEARLS

Clinical Features That Suggest a Non-Idiopathic Inflammatory Myopathy Diagnosis

- Family history of a similar illness
- Weakness related to exercise, eating, or fasting
- Sensory, reflex, or other neurological signs
- Cranial nerve involvement
- Fasciculations
- Muscle cramping (severe)
- Myasthenia (increasing weakness with repeated contractions)
- Myotonia (difficulty relaxing a contracted muscle)
- Significant atrophy or hypertrophy early in the illness
- Marked asymmetry
- Dyspnea due to diaphragmatic weakness with normal chest x-ray

Clinical, laboratory, pathological, and electrodiagnostic findings contribute to the proper diagnosis of IIM. Even in individuals with typical clinical features, it is essential to exclude other diseases that may have similar symptoms and signs (Table 57.6). Certain clinical features should suggest a different diagnosis. These include a family history of a similar illness; sensory, reflex, or other neurological changes; fasciculations; a relationship of the weakness to exercise, food intake, or fasting; major muscle cramping, myotonia (difficulty relaxing a contracted muscle), or myasthenia (increasing weakness with repeated contractions); significant early muscle atrophy or hypertrophy; marked asymmetry; weakness in the distribution of the cranial nerves; and dyspnea due to diaphragmatic weakness rather than lung fibrosis.

The single most useful laboratory feature of muscle destruction is elevation of the serum CK, although this is nonspecific, and a small proportion of patients—probably less than 5%—have a bonafide inflammatory muscle disease without ever having an elevated CK. Elevations of the serum levels of aldolase, serum glutamic-oxaloacetic transaminase (SGOT), serum glutamate pyruvate transaminase (SGPT), and lactate dehydroge-

TABLE 57.6 Differential Diagnosis of Idiopathic Inflammatory Myopathy

Neuromuscular Disorders
Genetic muscular dystrophies
Metabolic myopathies
Disorders of carbohydrate metabolism: McArdle disease, phosphofructokinase deficiency, adult acid maltase deficiency, and others
Disorders of lipid metabolism: carnitine deficiency, carnitine palmitoyl transferase deficiency
Disorders of purine metabolism: myoadenylate deaminase deficiency
Mitochondrial myopathies
Spinal muscular atrophies
Neuropathies: Guillain–Barré and other autoimmune polyneuropathies, diabetes mellitus, porphyria
Myasthenia gravis and Eaton–Lambert syndrome
Amyotrophic lateral sclerosis
Myotonic dystrophy and other myotonias
Familial periodic paralysis

Endocrine and Electrolyte Disorders
Hypokalemia, hypercalcemia, hypocalcemia, hypomagnesemia
Hypothyroidism, hyperthyroidism
Cushing syndrome, Addison disease

Toxic Myopathies (Partial List)
Alcohol
Amiodarone
Chloroquine and hydroxychloroquine
Cocaine
Colchicine
Corticosteroids
D-penicillamine
Ipecac
Statins and other lipid-lowering agents
Zidovudine (AZT)

Infections
Viral: HIV, human T-lymphotropic virus 1 (HTLV-1), influenza
Bacterial: *Staphylococcus, Streptococcus, Clostridia*
Parasitic: toxoplasmosis, trichinosis, schistosomiasis, cysticercosis

Miscellaneous
Polymyalgia rheumatica
Vasculitis
Eosinophilia myalgia syndrome
Paraneoplastic syndromes

nase (LDH) are as frequent but less specific for muscle disease. Unlike inflammatory markers for other autoimmune inflammatory diseases, those such as the erythrocyte sedimentation rate (ESR) and C-reactive protein are often not elevated. Although some studies have shown ESR to be elevated in 50% of patients, most experts find a substantially lower proportion of IIM patients to have an elevated ESR, even with active disease.[44] Likewise, hematological abnormalities, including anemia, are uncommon and rarely related to the underlying myopathy. If a significant abnormality is found, the physician should be alert to another cause for it.

Electromyographic (EMG) abnormalities are frequently present. Although the test is useful for excluding some neurological diseases that resemble IIM, it is painful for many patients and not useful for following the course of the illness.

MRI, especially a combination of the T1 and the fat-suppressed T2 (STIR) sequences, is remarkably useful in defining the extent of involvement and in planning a biopsy (Fig. 57.2). Whole-body MRI has been shown to facilitate the characterization of inflammatory myopathy, as certain patterns of muscle and subcutaneous tissue inflammation were predictive of the IIM subset (DM, PM, or IBM). In a recent study in juvenile DM, whole-body MRI was able to reveal inflammatory changes of subcutaneous and fascial tissue that was undetectable on clinical examination.[45] MRI can also help differentiate active disease from chronic disease, with active myositis being notable for changes consistent with muscle edema on T2-weighted images, and chronic myositis revealing a decrease in muscle bulk and replacement by adipose on T1-weighted images.[46] Although not specific, the changes of inflammatory myopathy on imaging can provide considerable assistance in confusing cases, as well as help in choosing a site to biopsy.

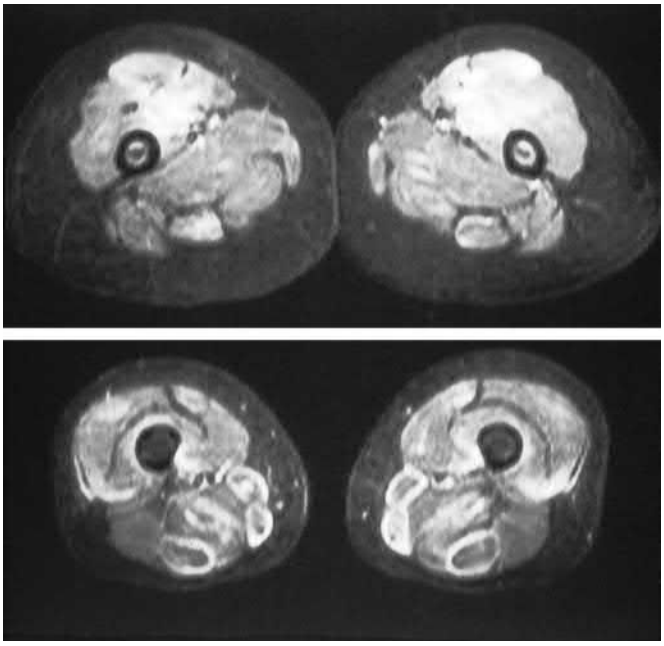


FIG. 57.2 Magnetic Resonance Images of the Upper and Lower Thighs of a Patient with Dermatomyositis Using the Fat-Suppressed T2 (STIR) Technique. With this technique inflammation appears as a bright signal; normal muscle is gray; bone, fat, fascia, and normal skin are dark. Blood vessels may appear as bright spots. Note the remarkable symmetry of the inflammation. In this patient most of the involvement is in the quadriceps in the upper thighs and around the periphery of the hamstring muscle group.

A muscle biopsy should be performed in every suspected case of myositis (Fig. 57.3). Although the patchy involvement means that the biopsy can occasionally miss inflammation, certain confounding diagnoses—for example amyloidosis, eosinophilic myositis, dystrophy, or some metabolic myopathies as well as the important variant IBM—can be diagnosed definitively only by biopsy. The identification of autoantibodies, particularly the myositis-specific autoantibodies, has distinct clinical and prognostic use.

ON THE HORIZON

- While there are newly revised and updated classification criteria for inflammatory myopathies, more precise criteria derived from big data as well as novel autoantibodies are likely in the future. Long-term studies are needed to better characterize the prognosis of idiopathic inflammatory myopathies, particularly in relation to recently discovered autoantibodies.

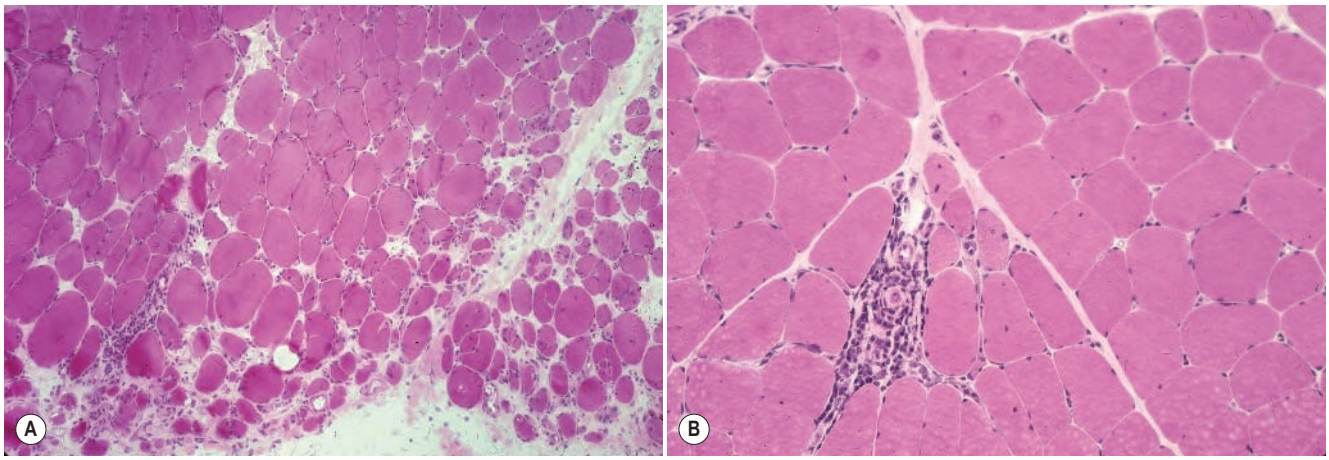


FIG. 57.3 Biopsy in Dermatomyositis. (A) Low-power (original magnification ×100) view of a muscle biopsy from a patient with dermatomyositis. Note the marked variation in fiber size and the large number of atrophic myocytes, particularly at the periphery of the fascicles. (B) High-power (original magnification ×200) view of inflammation around the vessels in the muscle biopsy of a patient with dermatomyositis. There are nearby atrophic cells and cells whose nuclei have moved away from the periphery of the cell (centralized nuclei).

PITFALLS

It is increasingly apparent that the boundary between IIM and some genetically determined myopathies cannot be cleanly drawn. Recently dystrophies with an extraordinary variety of clinical manifestations (with regard to age and distribution of weakness) have also been described.[47] Not only can inflammation be seen on biopsy in some patients, but a partial clinical response to corticosteroids also can occur. Furthermore, it is increasingly recognized that mitochondrial abnormalities can be limited to groups of skeletal muscles, leading to confusion with IIM. Toxic myopathies, of course, will continue to occur with the release of new drugs and thus will continue to be a possible source of diagnostic confusion.

Thus, not only must the history, physical examination, and biopsy be performed and interpreted with compulsiveness and care, but molecular diagnostic techniques also must be employed by clinicians in pursuit of an accurate diagnosis and appropriate therapy. The correct response to disease that persists in the face of powerful immunosuppressive therapy is a careful rethinking of the diagnosis, including, on occasion, rebiopsy and molecular consultation.

REFERENCES

1. Sontheimer RD. Dermatomyositis: an overview of recent progress with emphasis on dermatologic aspects. *Dermatol Clin*. 2002;20:387–408.
2. Bohan A, Peter JB. Polymyositis and dermatomyositis (first of two parts). *N Engl J Med*. 1975;292:344–347.
3. Bohan A, Peter JB. Polymyositis and dermatomyositis (second of two parts). *N Engl J Med*. 1975;292:403–407.
4. Hoogendijk JE, Amato AA, Lecky BR, et al. 119th ENMC international workshop: trial design in adult idiopathic inflammatory myopathies, with the exception of inclusion body myositis, 10–12 October 2003, Naarden, The Netherlands. *Neuromuscul Disord*. 2004;14:337–345.
5. Rose MR. 188th ENMC International Workshop: Inclusion Body Myositis, 2–4 December 2011, Naarden, The Netherlands. *Neuromuscul Disord*. 2013;23:1044–1055.
6. Lundberg IE. Arthritis Rheumatol 2017 Dec;69(12):2271-2282).
7. Kao AH, Lacomis D, Lucas M, et al. Anti-signal recognition particle autoantibody in patients with and patients without idiopathic inflammatory myopathy. *Arthritis Rheum*. 2004;50:209–215.
8. Christopher-Stine L, Casciola-Rosen LA, Hong G, et al. A novel autoantibody recognizing 200-kd and 100-kd proteins is associated with an immune-mediated necrotizing myopathy. *Arthritis Rheum*. 2010;62:2757–2766.
9. Mammen AL, Chung T, Christopher-Stine L, et al. Autoantibodies against 3-hydroxy-3-methylglutaryl-coenzyme A reductase in patients with statin-associated autoimmune myopathy. *Arthritis Rheum*. 2011;63:713–721.
10. Lahouti AH, Amato AA, Christopher-Stine L. Inclusion body myositis: update. *Curr Opin Rheumatol*. 2014;26:690–696.
11. Lloyd TE, Christopher-Stine L, Pinal-Fernandez I, et al. Cytosolic 5'-nucleotidase 1A as a target of circulating autoantibodies in autoimmune diseases. *Arthritis Care Res (Hoboken)*. 2016;68:66–71.
12. Argov Z, Eisenberg I, Mitrani-Rosenbaum S. Genetics of inclusion body myopathies. *Curr Opin Rheumatol*. 1998;10:543–547.
13. Mathews MB, Bernstein RM. Myositis autoantibody inhibits histidyl-tRNA synthetase: a model for autoimmunity. *Nature*. 1983;304:177–179.
14. Dalakas MC. Pathophysiology of inflammatory and autoimmune myopathies. *Presse Med*. 2011;40:e237–e247.
15. Mor A, Wortmann RL, Mitnick HJ, et al. Drugs causing muscle disease. *Rheum Dis Clin North Am*. 2011;37:219–231. vi.
16. Walji S, Rubenstein J, Shannon P, et al. Disseminated pyomyositis mimicking idiopathic inflammatory myopathy. *J Rheumatol*. 2005;32:184–187.
17. Zong M, Loell I, Lindroos E, et al. Effects of immunosuppressive treatment on interleukin-15 and interleukin-15 receptor alpha expression in muscle tissue of patients with polymyositis or dermatomyositis. *Ann Rheum Dis*. 2012;71:1055–1063.
18. Greenberg SA, Pinkus JL, Pinkus GS, et al. Interferon-alpha/beta-mediated innate immune mechanisms in dermatomyositis. *Ann Neurol*. 2005;57:664–678.
19. Bertrand A, Kostine M, Barnetche T, Truchetet ME, Schaeverbeke T. Immune related adverse events associated with anti-CTLA-4 antibodies: systematic review and meta-analysis. *BMC Med*. 2015;13:211.
20. Osta E, Coleman RE, Cortes J, Jani W. Advances in the management of HER2-positive early breast cancer. *Crit Rev Oncol Hematol*. 2017;119:1–12.
21. Wang PF, Chen Y, Song SY, Wang TJ, Ji WJ, Li SW, Liu N, Yan CX. Immune-Related Adverse Events Associated with Anti-PD-1/PD-L1 Treatment for Malignancies: A Meta-Analysis. *Front Pharmacol*. 2017;8:730.
22. Pedoeem A, Azoulay-Alfaguter I, Strazza M, Silverman GJ, Mor A. Programmed death-1 pathway in cancer and autoimmunity. *Clin Immunol*. 2014;153(1):145–152.

23. Suya Dai, Ru Jia, Xiao Zhang, et al. The PD-1/PD-Ls pathway and autoimmune diseases. *Cell Immunol*. 2014;290:72–79.
24. Cappelli LC, Gutierrez AK, Bingham CO 3rd, Shah AA. Rheumatic and Musculoskeletal Immune-Related Adverse Events Due to Immune Checkpoint Inhibitors: A Systematic Review of the Literature. *Arthritis Care Res (Hoboken)*. 2017;69(11):1751–1763.
25. Zhang S, Zhou Z, Wang L, Li M, Zhang F, Zeng X. Rheumatic immune-related adverse events associated with immune checkpoint inhibitors compared with placebo in oncologic patients: a systemic review and meta-analysis. *Ther Adv Chronic Dis*. 2021;12:2040622320976996.
26. Ikegame S, Takeda M, Ohno S, et al. Both RIG-I and MDA5 RNA helicases contribute to the induction of alpha/beta interferon in measles virus-infected human cells. *J Virol*. 2010;84:372–379.
27. Engel AG, Arahata K, Emslie-Smith A. Immune effector mechanisms in inflammatory myopathies. *Res Publ Assoc Res Nerv Ment Dis*. 1990;68:141–157.
28. Nagaraju K, Casciola-Rosen L, Lundberg I, et al. Activation of the endoplasmic reticulum stress response in autoimmune myositis: potential role in muscle fiber damage and dysfunction. *Arthritis Rheum*. 2005;52:1824–1835.
29. Casciola-Rosen L, Nagaraju K, Plotz P, et al. Enhanced autoantigen expression in regenerating muscle cells in idiopathic inflammatory myopathy. *J Exp Med*. 2005;201:591–601.
30. Goldstein R, Duvic M, Targoff IN, et al. HLA-D region genes associated with autoantibody responses to histidyl-transfer RNA synthetase (Jo-1) and other translation-related factors in myositis. *Arthritis Rheum*. 1990;33:1240–1248.
31. Mammen AL, Gaudet D, Brisson D, et al. Increased frequency of DRB1*11:01 in anti-hydroxymethylglutaryl-coenzyme A reductase-associated autoimmune myopathy. *Arthritis Care Res (Hoboken)*. 2012;64:1233–1237.
32. Fiorentino DF, Chung LS, Christopher-Stine L, et al. Most patients with cancer-associated dermatomyositis have antibodies to nuclear matrix protein NXP-2 or transcription intermediary factor 1gamma. *Arthritis Rheum*. 2013;65:2954–2962.
33. Dalakas MC, Illa I, Dambrosia JM, et al. A controlled trial of high-dose intravenous immune globulin infusions as treatment for dermatomyositis. *N Engl J Med*. 1993;329:1993–2000.
34. Mammen AL, Tiniakou E. Intravenous immune globulin for statin-triggered autoimmune myopathy. *N Engl J Med*. 2015;373:1680–1682.
35. Miller FW, Leitman SF, Cronin ME, et al. Controlled trial of plasma exchange and leukapheresis in polymyositis and dermatomyositis. *N Engl J Med*. 1992;326:1380–1384.
36. Oddis CV, Reed AM, Aggarwal R, et al. Rituximab in the treatment of refractory adult and juvenile dermatomyositis and adult polymyositis: a randomized, placebo-phase trial. *Arthritis Rheum*. 2013;65:314–324.
37. Aggarwal R, Bandos A, Reed AM, et al. Predictors of clinical improvement in rituximab-treated refractory adult and juvenile dermatomyositis and adult polymyositis. *Arthritis Rheumatol*. 2014;66:740–749.
38. The Muscle Study Group A randomized, pilot trial of etanercept in dermatomyositis. *Ann Neurol*. 2011;70:427–436.
39. Park J-K, Yoo H-G, Ahn D-S, et al. Successful treatment for conventional treatment-resistant dermatomyositis-associated interstitial lung disease with adalimumab. *Rheumatol Int*. 2012;32:3587–3590.
40. Lahouti AH, Brodsky RA, Christopher-Stine L. Idiopathic inflammatory myopathy treated with high-dose immunoablative cyclophosphamide—A long-term follow-up study. *JAMA Neurol*. 2015;72:1205–1206.
41. Kerola AM, Kauppi MJ. Abatacept as a successful therapy for myositis—a case-based review. *Clin Rheumatol*. 2015;34:609–612.
42. Leff RL, Miller FW, Hicks J, et al. The treatment of inclusion body myositis: a retrospective review and a randomized, prospective trial of immunosuppressive therapy. *Medicine (Baltimore)*. 1993;72:225–235.
43. Sayers ME, Chou SM, Calabrese LH. Inclusion body myositis: analysis of 32 cases. *J Rheumatol*. 1992;19:1385–1389.
44. Rider LG, Miller FW. Laboratory evaluation of the inflammatory myopathies. *Clin Diagn Lab Immunol*. 1995;2:1–9.
45. Malattia C, Damasio MB, Madeo A, et al. Whole-body MRI in the assessment of disease activity in juvenile dermatomyositis. *Ann Rheum Dis*. 2014;73:1083–1090.
46. Cantwell C, Ryan M, O'Connell M, et al. A comparison of inflammatory myopathies at whole-body turbo STIR MRI. *Clin Radiol*. 2005;60:261–267.
47. Emery AE. The muscular dystrophies. *BMJ*. 1998;317:991–995.

58

Spondyloarthritis

John D. Reveille and Lauren K. Ridley

The term *spondyloarthritis* (SpA) encompasses a heterogeneous group of inflammatory diseases characterized by spinal and peripheral joint arthritis, inflammation of the attachments of ligaments and tendons to bones (enthesitis), and, at times, mucocutaneous, ocular, and/or cardiac manifestations. These disorders show familial aggregation and are typically associated with genes of the major histocompatibility complex (MHC), particularly the human leukocyte antigen (HLA)-B27 (Chapter 5). The SpA spectrum includes (1) axial spondyloarthritis (AxSpA), including ankylosing spondylitis (AS) and non-radiographic axial spondyloarthritis (nr-AxSpA); (2) peripheral spondyloarthritis, which includes (a) reactive arthritis (ReA); (b) psoriatic arthritis (PsA) and/or spondylitis; (c) enteropathic arthritis and/or spondylitis associated with the inflammatory bowel diseases (IBD), ulcerative colitis (UC), or Crohn disease (CD), (d) juvenile-onset spondyloarthritis; and (e) undifferentiated SpA, which encompasses patients expressing elements of, but failing to fulfill, accepted criteria for one of the above diseases. In addition, isolated acute anterior uveitis (AAU)[1] and spondylitic heart disease (complete heart block and/or isolated aortic regurgitation)[2] associated with HLA-B27 may also be classified within the spectrum of SpA.

CLASSIFICATION OF SPONDYLOARTHRITIS

There are no diagnostic criteria for any of the SpA. Classification criteria have been developed to provide greater specificity in clinical studies. These include the Amor Criteria, proposed in 1988, the European Spondyloarthropathy Study Group (ESSG) criteria for SpA developed in 1991, and the modified New York criteria in 1984, which do not include advanced imaging techniques such as MRI (Table 58.1).[3] However, it may take up to 10 years to develop radiographic sacroiliitis after the onset of inflammatory back pain. In 2009, the Assessment of SpondyloArthritis International Society (ASAS) published new criteria that allowed (1) diagnosis of sacroiliitis based on MRI, and (2) classification of patients based solely on clinical features. This includes the use of HLA-B27 typing and C-reactive protein (CRP) determination and allows for the diagnosis of earlier disease found on MRI.[3] In recent years, more weight has been placed on the differentiation of non-radiographic axial spondyloarthritis (nr-AxSpA).

Criteria for peripheral spondyloarthritis based on the presence of peripheral arthritis, enthesitis, or dactylitis have also been developed utilizing the innovations of advanced imaging and HLA-B27 typing.[3] This resulted in a reclassification of all the various spondyloarthritides under the diagnoses of axial and peripheral SpA which aids in treatment since axial and peripheral disease respond differently.

For PsA, the Classification Criteria for Psoriatic ARthritis (CASPAR) have been established (see Table 58.1).[3] No criteria have been validated for enteropathic arthritis or for reactive arthritis.

The International League Against Rheumatism (ILAR) juvenile idiopathic arthritis (JIA) classification criteria do not recognize SpA as a distinct clinical entity. Instead, juvenile SpA patients tend to fall under three of seven subtypes of JIA in the ILAR classification: enthesitis-related arthritis (ERA), PsA, or undifferentiated arthritis. IBD-related arthritis, reactive arthritis, and juvenile AS are not addressed in the ILAR classification criteria. (ILAR; see Table 58.1).

EPIDEMIOLOGY

The frequency of SpA in general and AS in particular varies in different populations and parallels the frequency of HLA-B27. The prevalence of AS varies from 0.2% to 0.7% among people of European ancestry[4–6] and has been reported in similar frequencies in eastern Asia (Table 58.2). AS is much less frequent in persons of African and Japanese descent, in whom HLA-B27 is rare.

The prevalence of ReA has varied over time and in different regions, depending on endemic rates of sexually acquired and enteric infections that trigger it. The frequency of ReA has dramatically declined following the adoption of safer sexual practices and better sanitation, at least in economically advanced countries, in the wake of the human immunodeficiency virus (HIV) epidemic.

Psoriasis affects 1% to 3% of the general population.[5] Psoriatic arthritis occurs in about 20% to 30% of patients with psoriasis.[7] PsA incidence is similar between males and females, and it is uncommon in Asians and blacks.[8]

The prevalence of IBD is 100 to 200 per 100,000 among Caucasians, with an equal male-to-female ratio.[4,6] It is rare in people of African and Asian descent. The prevalence of extraintestinal manifestations, such as peripheral or axial disease, in IBD varies between Crohn disease and ulcerative colitis (Table 58.3).

In North American JIA patients, enthesitis-related arthritis is found in 10% to 11% of patients, psoriatic arthritis in 6% to 11%, and undifferentiated SpA in about 1% to 2% of patients.[9]

TABLE 58.1 CLASSIFICATION CRITERIA FOR SPONDYLOARTHRITIS

A. Amor Criteria for Spondyloarthropathy

Clinical symptoms or history	Points
a. Lumbar or dorsal morning stiffness or pain at night	1
b. Asymmetrical oligoarthritis	2
c. Buttock pain (1 point) or alternating buttock pain (2 points)	1 or 2
d. Dactylitis	2
e. Heel pain or other well-defined enthesopathy	2
f. Iritis or uveitis	1
g. Nongonococcal urethritis or cervicitis with 1 month before the onset of arthritis	1
h. Acute diarrhea within 1 month before the onset of arthritis	2
i. Psoriasis, balanitis, or inflammatory bowel disease (ulcerative colitis or Crohn disease)Radiographic findings	3
a. Sacroiliitis (bilateral grade 2 or unilateral grade 3)Genetic background	2
b. Presence of HLA-B27 and/or family history of ankylosing spondylitis, reactive arthritis, uveitis, psoriasis, or inflammatory bowel diseaseResponse to treatment	2
a. Clear-cut improvement within 48 h after NSAIDs intake or rapid relapse of the pain after NSAIDs discontinuation	2

If the sum is greater than or equal to 6, the patient is considered to have a spondyloarthritis.

B. European Spondyloarthropathy Study Group (ESSG) Criteria for Spondyloarthritis

1. Inflammatory back pain or synovitis (asymmetrical, lower extremity) plus one of the following:
   (a) Alternating buttock pain
   (b) Sacroiliitis
   (c) Heel pain (enthesitis)
   (d) Positive family history
   (e) Psoriasis
   (f) Crohn disease, ulcerative colitis
   (g) Urethritis or cervicitis or acute diarrhea in the preceding 4 weeks

C. The Modified New York Criteria for Ankylosing Spondylitis[4]

1. Clinical criteria:
   (a) Low-back pain and stiffness for more than 3 months which improves with exercise, but is not relieved by rest
   (b) Limitation of motion of the lumbar spine in both the sagittal and frontal planes
   (c) Limitation of chest expansion relative to normal values correlated for age and gender
2. Radiological criterion:
   (a) Sacroiliitis grade 2 bilaterally or grade 3–4 unilaterally
   (b) *Definite ankylosing spondylitis* if the radiological criterion is associated with at least one clinical criterion

D. The Classification Criteria for Psoriatic Arthritis (CASPAR) Criteria for Psoriatic Arthritis

1. Inflammatory joint disease plus at least three points from the following features:
   (a) Current psoriasis (assigned a score of 2; all others assigned a score of 1)
   (b) History of psoriasis
   (c) Family history of psoriasis
   (d) Dactylitis
   (e) Juxta-articular new bone formation
   (f) Rheumatoid factor seronegativity
   (g) Nail dystrophy

E. ASAS Criteria for Axial Spondyloarthritis in Patients With Chronic Low Back Pain for at Least 3 Months

1. Sacroiliitis on imaging (either MRI findings of inflammatory disease in the sacroiliac joints or sacroiliitis on standard pelvic X-rays by New York criteria) plus ONE of the following OR HLA-B27 positivity plus TWO of the following:
   (a) Inflammatory low back pain
   (b) Arthritis
   (c) Enthesitis
   (d) Dactylitis
   (e) Psoriasis
   (f) Crohn disease or ulcerative colitis
   (g) Good response to nonsteroidal anti-inflammatory agents
   (h) Positive family history of SpA
   (i) Presence of human leukocyte antigen (HLA)-B27
   (j) Elevated cross-reactive protein

F. ASAS Criteria for Peripheral Spondyloarthritis—Arthritis or Enthesitis or Dactylitis With At Least One of:

a. Uveitis
b. Psoriasis
c. Crohn disease or ulcerative colitis
d. Previous infection
e. Presence of HLA-B27
f. Sacroiliitis on imaging

OR

with at least two of:

a. Inflammatory low back pain
b. Arthritis
c. Enthesitis
d. Dactylitis
e. Positive family history of SpA

G. The International League Against Rheumatism (ILAR) Juvenile Idiopathic Arthritis Classification Criteria for Enthesitis-Related Arthritis (ERA)

Arthritis and enthesitis

OR

Arthritis or enthesitis with at least two of:

1. Sacroiliac joint tenderness and/or inflammatory spinal pain
2. Presence of HLA-B27
3. Family history in at least one first- or second-degree relative of medically confirmed HLA-B27-associated disease
4. Anterior uveitis that is usually associated with pain, redness, or photophobia
5. Onset of arthritis in a boy after 8 years of age

Exclusions

- Psoriasis confirmed by a dermatologist in at least one first- or second-degree relative
- Presence of systemic arthritis

H. The International League Against Rheumatism (ILAR) Juvenile Idiopathic Arthritis Classification Criteria for Psoriatic Arthritis

Arthritis and psoriasis

OR

Arthritis with at least two of the following:

1. Dactylitis
2. Nail pitting or onycholysis3.First-degree relative with psoriasis

Exclusions

- Arthritis in an HLA-B27-positive male patient that started after age 6
- Personal or family history of HLA-B27-associated disease
- RF positivity on two or more occasions at least three months apart

TABLE 58.2 The Epidemiology of Spondyloarthritis

Ethnic Group	HLA-B27 Frequency (%)	Prevalence of AS %(Criterion)	Prevalence of PsA % (Criterion)	Prevalence of SpA % (Criterion)
Europe				
Azores		n.a.	n.a.	1.6 (ESSG)
Czech Republic		0.09 (mNY)	0.05 (Vasey)	n.a.
France		0.08 (questionnaire)	0.19 (questionnaire)	0.3 (questionnaire)
Germany		0.86 (mNY)	0.29 (ESSG)	1.9 (ESSG
Greece		0.24 (mNY)	0.17 (ESSG)	0.49 (ESSG)
Iceland		0.13 (mNY)	0.14 (other)	0.49 (ESSG)
Italy		0.37 (mNY)	0.42 (mNY)	n.a.
Netherlands		0.24 (questionnaire plus Xray)	n.a.	n.a.
N. Norway(Sami)	24	1.8 (NY)	n.a.	n.a.
Norway-general	24	1.1-1.4(mNY)	0.23 (ICD)	n.a.
South Sweden		0.25 (ICD10)	0.25 (ICD10)	0.45 (ICD10)
America				
United States	6.1	0.52 (questionnaire plus X-ray)	0.16	1.4 (ESSG)
Mexico	4.6	0.02 (mNY)	0.02	n.a.
Alaska (Eskimo)		0.4 (NY)	<0.1	2.5 (ESSG)
Argentina		n.a.	0.07 (CASPAR)	n.a.
Asia				
China		0.25 (mNY)	0.02	0.78 (ESSG)
Iran		0.12 (unclear)	n.a.	0.23 (COPCORD)
Japan		0.0065 (NY)	0.001	0.0095
Pakistan		n.a.	n.a.	0.1 (COPCORD)
Siberia	37	1.1 (NY)	0.3	2.5 (ESSG)
Taiwan		0.19–0.54 (mNY)	n.a.	n.a.
Turkey		0.49 (mNY)	n.a.	1.05 (ESSG)
Vietnam		n.a.	n.a.	0.28 (COPCORD)

AS, Ankylosing spondylitis; *HLA*, human leukocyte antigen; *n.a.*, not available; *PsA*, psoriatic arthritis; *SpA*, spondyloarthritis; *NY*, New York; *mNY*, modified New York criteria for AS; *ESSG*, European Spondyloarthropathy Study Group Criteria.

TABLE 58.3 Extraintestinal Manifestations of Inflammatory Bowel Disease

Feature	Crohn Disease (%)	Ulcerative Colitis (%)
Musculoskeletal	6–46	
Peripheral arthritis	10–20	5–14
Sacroiliitis	<50	—
Concomitant axial and peripheral arthritis	3–6	
Enthesitis	7–50	
Cutaneous	2–34	
Erythema nodosum	1–20	
Ocular	4–12	
Anterior uveitis	2–5	2
Renal	—	
Nephrolithiasis	17	3
Renal insufficiency	2	Rare
Hepatobiliary	5–15	
Primary sclerosing cholangitis	<10	
Pancreatitis	2	
Vasculitis	Takayasu arteritis	Leukocytoclastic vasculitis, Bechet disease

Reference: Colia R, Corrado A, Cantatore FP. Rheumatologic and extraintestinal manifestations of inflammatory bowel diseases. *Ann Med.* 2016;48(8);577–585.

KEY CONCEPTS

The Genetic Basis of Spondyloarthritis

- Human leukocyte antigen (HLA)-B27 comprises over 80% of the overall genetic susceptibility to ankylosing spondylitis (AS) and contributes heavily to susceptibility for ReA, PsA, and enteropathic SpA.
- Additional risk comes from other major histocompatibility complex (MHC) genes, including *HLA- B*40*, among others, for AS and *C*06:02* for psoriasis.
- Genome-wide association studies (GWAS) utilizing dense single nucleotide polymorphism (SNP) mapping have located more than 130 additional genes or genetic regions in AS susceptibility.
- Over 60 genes have been identified thus far in psoriasis pathogenesis.
- GWAS utilizing dense SNP mapping examining larger sample sizes will likely continue to locate many of the remaining genes in AS susceptibility.

PATHOGENESIS

Genetics of Spondyloarthritis

Familial Aggregation

The sibling recurrence risk ratio in AS is as high as 82%, and twin-based studies estimate disease heritability exceed 90%.[10] The concordance rate for AS in identical twins has been reported to be as high as 63% compared with 23% in

nonidentical twins.[10] The concordance rate for psoriasis in monozygotic twins is 70% *versus* 15% to 30% in dizygotic twins. Recurrence risk for parents and siblings of patients with CD is 4.8% and 7%, respectively, and for UC 0.9% and 1.2%, respectively.[10]

HLA-B27 and Spondyloarthritis

HLA-B27 is found in 85% to 90% of AS patients of European ancestry, over 85% of AS patients of Hispanic ancestry, 56% to 84% of Middle Eastern or North African ancestry, and about 60% of AS patients of Asian and African American populations.[10–13] The estimated prevalence of HLA-B27 in the United States is 6.1% overall, is highest in younger individuals (7.5% before age 50 years), and falls rapidly over age 50 years (3.3%).[14]

Approximately 70% of patients with ReA have HLA-B27, except in Africa, where no association of HLA-B27 was seen in those with HIV-associated SpA.[12] HLA-B27 is found in 60% to 70% of patients with psoriatic spondylitis and in 25% of those with peripheral PsA.[11,12] Up to 70% of those with IBD-associated spondylitis have HLA-B27, although no HLA-B27 association is seen with asymptomatic sacroiliitis. Approximately 50% of patients with AAU alone are HLA-B27 positive.[11]

There are over 222 molecular subtypes of HLA-B27 that have been described as of March 2020 (http://www.ebi.ac.uk/ipd/imgt/hla/nomenclature/index.html). The most common subtypes (HLA-B*27:05, B*27:02, B*27:04, B*27:07) are clearly associated with SpA. Each subtype is found in specific populations. These subtypes evolved from the parent allele HLA-B*27:05 along five lines (Fig. 58.1) in three distinct geographical regions.

There are different hypotheses that have been extended how HLA-B27 impacts disease susceptibility. One theory suggests that SpA results from a unique set of antigenic peptides, either bacterial or self, that are bound and presented by HLA-B27 to CD8 T cells, resulting in a HLA-B27–restricted cytotoxic T-cell response found only in joints and other affected tissues (the so-called arthritogenic peptide hypothesis). However, such peptides have not been reproducibly identified to date. In fact, the data showing interaction between HLA-B27 and a gain-of-function (GOF) variant of endoplasmic reticulum (ER)–associated aminopeptidase I (ERAP1) suggest that *aberrant* antigen processing may play a central role in AS susceptibility.[15] The same has been described for HLA-C*06:02 and psoriasis.[18] An alternative concept focuses on self-association as a unique property of the HLA-B27 molecule. HLA-B27 heavy chains can form homodimers *in vitro* that are dependent on disulfide binding through their cysteine-67 residues in the extracellular α_1 domain (as well as other cysteine residues in other domains) (Fig. 58.2).[12,16] These B27 homodimers are detectable at the cell surface in patients with SpA, are capable of peptide binding, and are more abundantly expressed when the cell's antigen-presenting function is impaired. They are ligands for a number of natural killer (NK) and related cell surface receptors. Populations of synovial and peripheral blood monocytes, NK cells, and B and T lymphocytes both from patients with SpA and from controls carry receptors for HLA-B27 homodimers, including killer immunoglobulin-like receptor (KIR)3DL1 and KIR3DL2 and immunoglobulin (Ig)–like transcript 4 (ILT4).[12,15] These homodimers may act as a proinflammatory target or receptor for humoral or cell-mediated autoimmune responses. However, it is not known whether HLA-B27 homodimer formation is specific for, or even correlates with, the presence of SpA; in fact, disease does not develop in most HLA-B27–positive individuals. Alternatively, HLA-B27 misfolding and accumulation within the ER can result in a proinflammatory intracellular stress response through stimulating secretion of type 1 interferon (IFN).[16]

Other Major Histocompatibility Complex Genes and SpA Susceptibility

Other MHC genes have been implicated in AS in addition to HLA-B27 (Fig. 58.3), although their identification is complicated by the tight linkage disequilibrium found within the MHC, and many of the associations may reflect linkage to B27. These include the epithelial "stress" marker MICA, tumor necrosis factor (*TNF*), heat shock protein (*HSP*)*-70, LMP-2* and *LMP-7, HLA-DRB1*01*, and *DRB1*04* alleles.[12] However, other MHC genes have been shown to be SpA-associated independent of HLA-B27. These include *HLA-A*02:01* and *B*40* as well as negative associations with *HLA-B*07, B*15* and *B*35*, which have been described in whites, Han Chinese, and blacks with AS.[17] Also, in white Americans, *HLA-DPB1*03:01* is associated with AS.[13] *HLA-DRB1*15:01/DQB1*06:02* is negatively associated with AS but positively associated with uveitis.[13,17]

In psoriasis, the primary MHC association is with HLA-Cw6 (*HLA-C*06:02*) (though not with PsA). *HLA-B*38* has been associated with psoriatic arthritis.[17] The *HLA-B*27:05:02-C*06:02* haplotype is associated with enthesitis, dactylitis, and symmetric sacroiliitis, whereas *HLA-B*08:01-C*07:01* and its component alleles are positively associated with joint fusion and deformities, asymmetrical sacroiliitis, and dactylitis.[18] Both of these haplotypes are linked to more severe disease.[18] In contrast, *HLA-B*44* haplotypes are associated with milder disease.[18] *HLA-DRB1*01:03* has been strongly associated with enteropathic peripheral arthritis.[17]

Non–Major Histocompatibility Complex Genes in Susceptibility to Spondyloarthritis

Genome-wide association studies (GWAS) have implicated over 130 genes in AS susceptibility (see Fig. 58.3). These genes fall into different functional networks (Table 58.4). The most consistent (and best replicated) association is with *ERAP1*, which acts as a molecular ruler in the ER in trimming peptides processed in proteasomes to optimal length of nine amino acids for MHC class I binding and presentation. HLA-B27–negative patients with AS lack an association with *ERAP1*,[15] although an association with *ERAP2* is seen.

A number of genes whose products are operative in the Th17 pathway have also been associated with AS susceptibility: the IL-23 receptor (*IL23R*), which pairs with the *IL12RB1* gene product to confer IL-23 (but not IL-12) responsiveness on cells expressing both subunits; prostaglandin E receptor 4 (*PTGER4*), which stimulates dendritic cell (DC) production of IL-23 and, in turn, Th17 expansion and is overexpressed in SpA synovium; signal transducer and activator of transcription 3 (*STAT3*), a key regulatory factor in Th17 responses; and *Interleukin-12β* (*IL12B*), which encodes the IL12p40 protein, a component of both IL-12 and IL-23. *Jak2* plays a role in the downstream signaling pathway of IL-23R.[19]

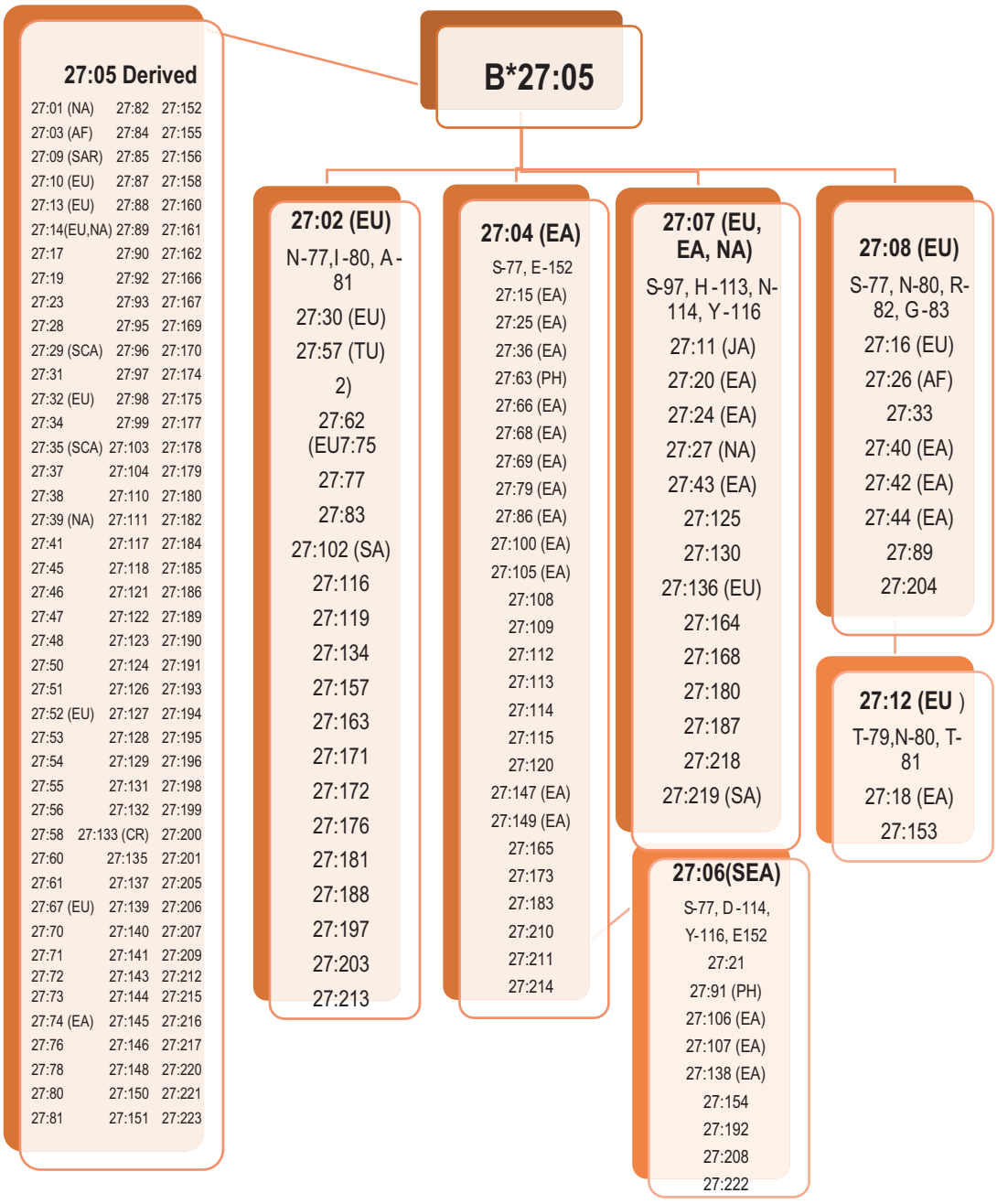


FIG. 58.1 The six major families of human leukocyte antigen (HLA)-B27 subtypes (HLA-B*27:59, B*27:64, B*27:65, and B*27:94 represent truncated genes with deletions of most of exons 2 and 3 and are not included above) are denoted in relationship to the "parent" subtype HLA-B*27:05. The ethnic or geographic origin of the subtypes is indicated in parentheses next to the allele: EU, European; NA, North American; SA, South American; EA, East Asian; SEA, Southeast Asian; NEA, Northeast Asian; ME, Middle East; Afr, African; UK, British Isles; Tur, Turkish; Mex, Mexican mestizo. Where no parentheses are given, the origin of the cell line from which the sequence came is unknown. Most B27 subtypes have evolved through five patterns of evolution along geographic lines. The first group, with HLA-B*27:02-B*27:23 (and B*27:30), appears to have evolved in Europe and the Middle East, entails anywhere from one to seven amino acid substitutions in the first (α_1) domain, and has the second (α_2) domain identical to B*27:05. The second group, including HLA-B*27:04 as the most common allele, evolved in Eastern Asia and includes a uniform amino acid substitution in the α_1 domain and anywhere from one to seven substitutions in the α_2 domain. The third and fourth groups seem to have evolved directly from HLA-B*27:05, from mutations in either exon 2 or 3, respectively. The fifth group, including HLA-B*27:07 as its leading member, evolved in southern and eastern Asia, the Middle East, and Sardinia with patterned amino acid substitutions in the α_2 domain. Finally, the sixth group, with HLA-B*27:08 as its common member evolved largely in East Europe, has different characteristic pattern of amino acid substitutions in the α_2 domain. Notable exceptions include HLA-B*27:13, B*27:109 and B*27:112-27:115, which have amino acid substitution outside α_1 and α_2 in the membrane proximal (α_3) domain. *AF*, African, *EU*, European, *EA*, Eastern Asian; *JA*, Japanese; *NA*, North American; *SEA*, Southeast Asian; *PH*, Philipino; *SA*, South American; *SCA*, Scandinavian; *TUR*, Turkic. (https://www.ebi.ac.uk/ipd/imgt/hla/allele.html).[13] (Data are derived from http://www.ebi.ac.uk/cgi-bin/imgt/hla/allele.cgi.)

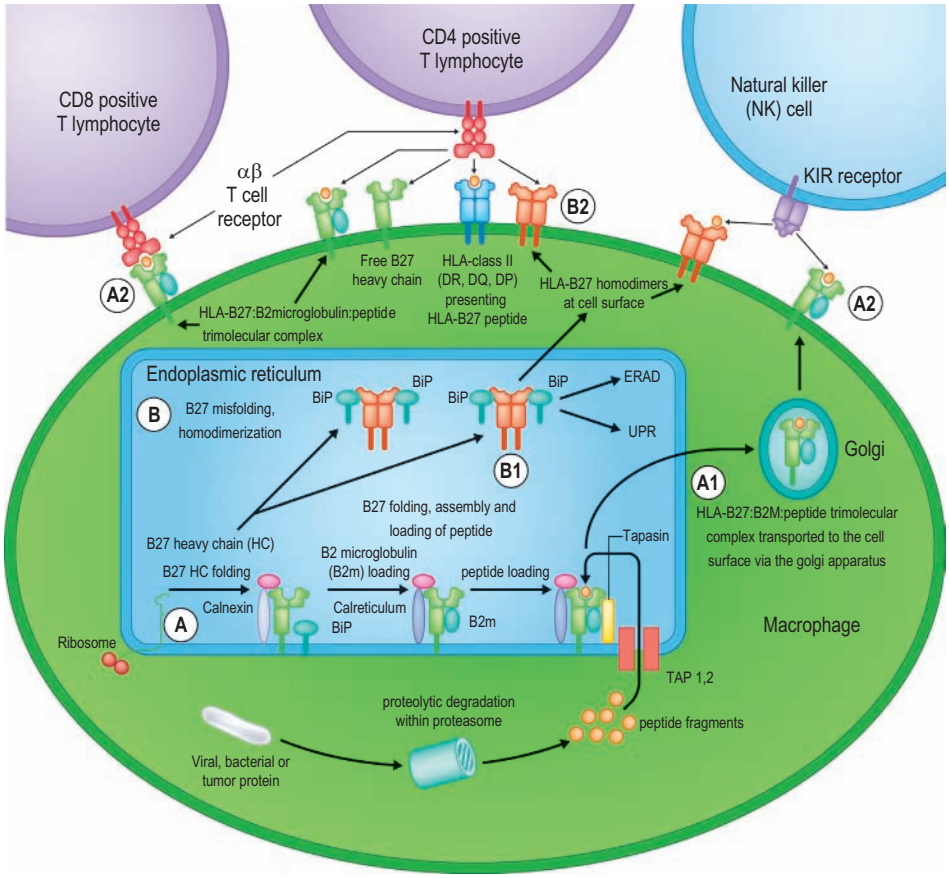


FIG. 58.2 After transcription of the human leukocyte antigen (HLA)-B27 heavy chain on ribosomes, it is inserted into the endoplasmic reticulum (ER), glycosylated, and two pathways ensue. (A) The B27 heavy chain is retained through binding with calnexin and ERp57, folded into its tertiary structure, and bound to β_2 microglobulin. After that calnexin releases the complex and it is associated with calreticulum, which, in turn, chaperones the formation of the peptide loading on to the complex of heavy chain, β_2 microglobulin and antigenic peptide, via the TAP proteins and tapasin. The antigenic peptide is derived from intracellular proteins from viruses, bacteria, tumors, and so on, that have been degraded in proteasomes, and then the peptides are trimmed for optimal length for peptide loading by endoplasmic reticulum-associated aminopeptidases (ERAP1 and ERAP2). Then the trimolecular peptide complex (HLA-B27 heavy chain, β_2 microglobulin and peptide) travels through the Golgi apparatus (A1) to the cell surface, where the antigenic peptide is presented either to the α: β T-cell receptor on CD8 T lymphocytes or to the killer immunoglobulin (KIR) receptor on natural killer (NK) cells (A2); or (B) the HLA-B27 heavy chain misfolds in the ER, forming B27 homodimers and other misfoldings which are bound to the ER chaperone BiP. Then, they either (B1) accumulate there, causing either ER-associated degradation (ERAD) or a proinflammatory ER unfolded protein response (UPR); or (B2) the B27 homodimers migrate to the cell surface, where they either become antigenic themselves or present peptide to receptors on T cells and NK cells.

GWAS in psoriasis and PsA have identified over 60 genes implicated in disease susceptibility outside the MHC.[20] These include genes in common with those for AS (*IL23R, IL12B, CDKAL1*, and *ERAP1*, the last showing interaction with HLA-Cw6 similar to that seen with HLA-B27) (see Fig. 58.3).

The first gene to be implicated in IBD susceptibility was *NOD2/CARD15*, which specifically leads to susceptibility to Crohn disease and whose protein product serves as a receptor for bacterial products in monocytes that transduces signals leading to NF-κB activation. A number of GWAS have implicated over 240 genes in IBD susceptibility.[21] The IL-23/IL-17 signaling pathway also appears involved with IBD pathophysiology with IL-23 shown to be important in development and maintenance of mucosal cells. Some of the genes implicated include *JAK2, IL23R*, and *TYK2*, among others. What is especially striking is the number of genes shared with the various types of SpA,[13] suggesting a common pathogenesis (see Fig. 58.3).

Genes and Severity of SpA

Disease severity in AS also has a hereditary component. By defining severity on the basis of radiographic involvement, two genes have been identified in a large candidate gene study.[22] One is the gene for cyclooxygenase I (COX-I)—the target of nonsteroidal anti-inflammatory drugs (NSAIDs)—the other for the receptor activator of NF-κB (RANK). No evidence was found for an MHC contribution to radiographic severity.

Infection

A role for triggering infections has been better documented in SpA than in most other rheumatic diseases. In developed countries, the most frequent type of ReA occurs following urogenital infections with *Chlamydia trachomatis* (endemic ReA). Postdysenteric ReA occurs after various *Shigella* and *Salmonella* (especially *S. typhimurium* and *S. enteriditis*), *Campylobacter jejuni* and *C. fetus*, and, in Europe, *Yersinia enterocolitica* species.

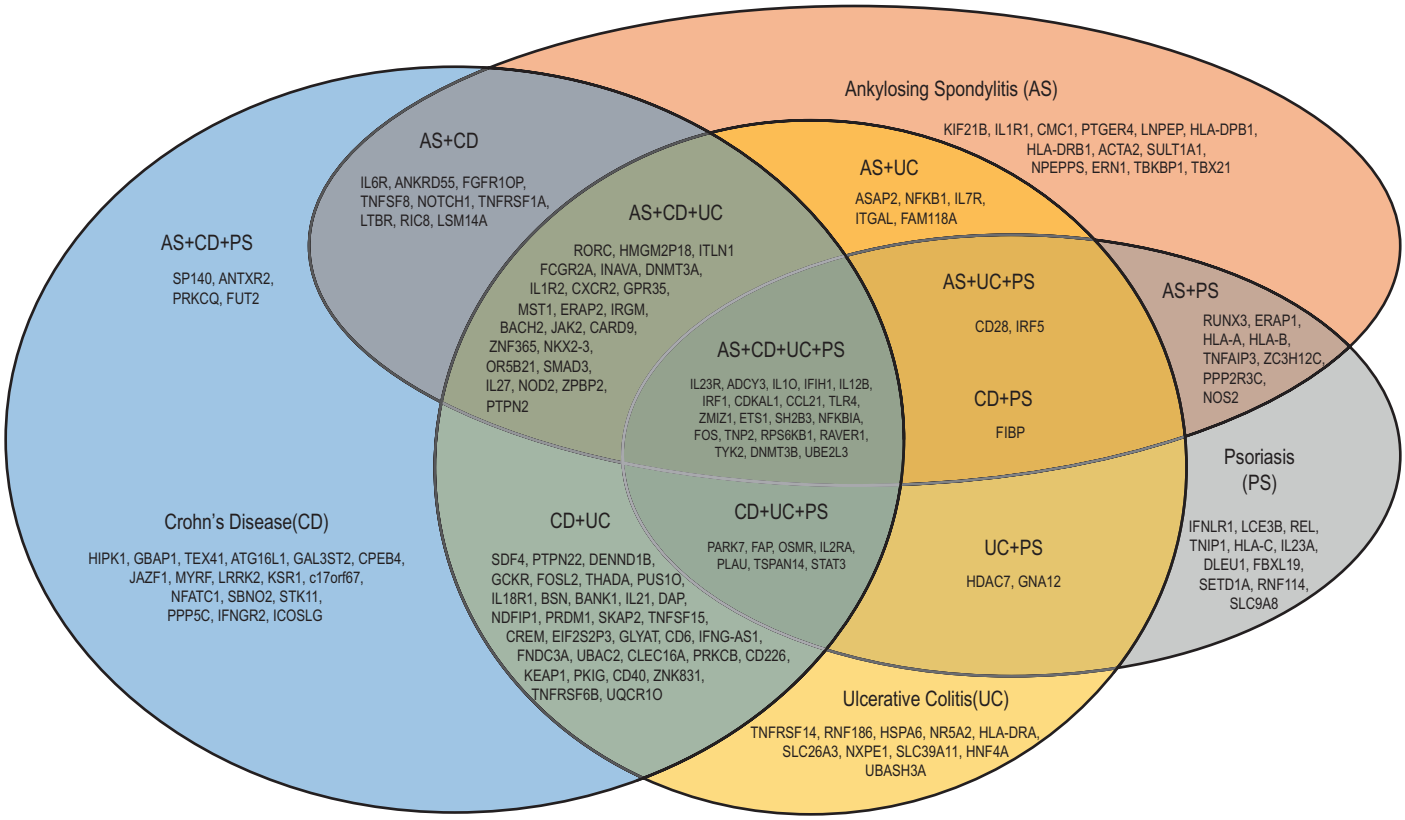


FIG. 58.3 Genetic factors implicated in spondyloarthritis. ACTA2-Actin, Alpha 2, Smooth Muscle, Aorta; ADCY3-Adenylate cyclase 3; ANKRD55-ankyrin repeat domain-containing protein 55; ANTXR2-anthrax toxin receptor 2; ASAP2-ArfGAP with SH3 domain, ankyrin repeat and PH domain 2; ATG16L1-Autophagy16-like 1; BACH2-BTB and CNC homology 2; BANK1-B cell scaffold protein with ankyrin repeats-1; BSN-Bassoon Mouse, Homolog of (Zinc Finger 231); C17orf67-chromosome 17 open reading frame 67; CARD9-caspase recruitment domain family member 9; CCL21- C-C motif chemokine ligand 21; CD6-CD6 molecule; CD28-CD28 molecule; CD40-CD40 molecule; CD226-CD226 molecule; CDKAL1-CDK5 regulatory subunit associated protein 1 like 1; CLEC16A- C-type lectin domain containing 16A; CMC1- C-X9-C motif containing 1; CPEB4- cytoplasmic polyadenylation element binding protein 4; CREM-cAMP responsive element modulator; CXCR2- C-X-C motif chemokine receptor 2; DAP-death-associated protein; DENND1B- DENN domain containing 1B; DLEU1-deleted in lymphocyte leukemia 1; DNMT3A-DNA methyltransferase 3 alpha; DNMT3B- DNA methyltransferase 3 beta; EIF2S2P3- eukaryotic translation initiation factor 2 subunit 2 beta pseudogene 3; ERAP1- endoplasmic reticulum aminopeptidase 1; ERAP2- endoplasmic reticulum aminopeptidase 2; ERN1- endoplasmic reticulum to nucleus signaling 1; ETS1- ETS proto-oncogene 1, transcription factor; FAM118A-family with sequence similarity 118 member A; FAP-fibroblast activation protein alpha; FBXL19- F-box and leucine rich repeat protein 19; FCGR2A- Fc fragment of IgG receptor IIa; FGFR1OP- FGFR1 oncogene partner; FIBP- FGF1 intracellular binding protein; FNDC3A-fibronectin type III domain containing 3A; FOS-V-Finkel-Biskis-Jinkins Murine Osteosarcoma Viral Oncogene Homolog; FOSL2- FOS-related antigen 2; FUT2-fucosyltransferase 2; GAL3ST2- galactose-3-O-sulfotransferase 2; GBAP1- glucosylceramidase beta pseudogene 1; GCKR-glucokinase regulator; GLYAT-glycine-N-acyltransferase; GNA12-G protein subunit alpha 12; GPR35-G protein-coupled receptor 35; HDAC7- histone deacetylase 7; HIPK1- homeodomain interacting protein kinase 1; HLA-A-human leukocyte antigen A; HLA-B-human leukocyte antigen B; HLA-C-human leukocyte antigen C; HLA-DP-human leukocyte antigen DP; HLA-DRA-human leukocyte antigen DR alpha; HLA-DRB1-humn leukocyte antigen DR beta 1; HMGN2P18- high mobility group nucleosomal binding domain 2 pseudogene 18; HNF4A-hepatocyte nuclear factor 4-alpha; HSPA6-heat shock 70kd protein 6;ICOSLG -Inducible T-cell Costimulator Ligand; IFIH1-interferon-induced helicase C domain-containing protein 1; IFNG-AS1-interferon gamma; IFNGR2-interferon gamma receptor 2; IFNLR1-interferon lambda receptor 1; IL1R1-interleukin 1 receptor type 1; IL1R2-interleukin 1 receptor type 2; IL2RA-interleukin 2 receptor alpha; IL6R-interleukin 6 receptor; IL7R-Interleukin 7 Receptor; IL10-interleukin 10; IL12B-interleukin 12B (IL12 p40 subunit); IL18R1-interleukin 18 receptor 1 (alpha chain); IL21-interleukin 21;IL23A-interleukin 23 alpha (IL12 p19 subunit); IL23R-interleukin 23 receptor; IL27-interleukin 27; IRF5-interferon regulatory factor 5; INAVA-Innate Immunity Activator; IRF1-interferon Regulatory Factor 1; IRF5-interferon regulatory factor 5; IRGM-immunity-related GTPase family, M; ITGAL-integrin alpha-L; ITLN1-intelectin 1; JAK2-Janus kinase 2; JAZF1- JAZF zinc finger 1;KEAP1- kelch like ECH associated protein 1; KIF21B-Kinesin Family Member 21B; KSR1- kinase suppressor of ras 1; LCE3B- late cornified envelope 3B; LNPEP-Leucyl-Cystinyl Aminopeptidase; LRRK2-leucine rich repeat kinase 2; LSM14A-Sm like Protein LSM14A mRNA Processing Body Assembly Factor; Ltbr-Lymphotoxin B Receptor; Mst1-Macrophage Stimulating 1; Myrf-Myelin Regulatory Factor; NDFIP1- Neural Precursor Cell Expressed, Developmentally Downregulated 4 Family-Interacting Protein 1; NFATC1-Nuclear Factor of Activated T Cells, Cytoplasmic, Calcineurin-Dependent 1; NFIP1; NFKB1- Nuclear Factor Kappa-B, Subunit 1; NFKBIA-nuclear Factor of Kappa Light Chain Gene Enhancer in B Cells Inhibitor, Alpha; NKX2-3-NK2 homeobox 3; NOD2-Nucleotide-Binding Oligomerization Domain Protein 2; NOS2-Nitric Oxide Synthase 2A; NOTCH1-NOTCH, Drosophila, Homolog of, 1; NPEPPS-aminopeptidase, puromycin-sensitive;NR5A2-nuclear

receptor subfamily 5 group A member 2; NXPE1-neurexophilin and PC-esterase domain family member 1; OR5B21-Olfactory receptor family 5 subfamily B member 21; OSMR-Oncostatin M receptor; PARK7-Parkinsonism associated deglycase; PKIG-cAMP-dependent protein kinase inhibitor gamma; PLAU-plasminogen activator, urokinase; PPP2R3C-protein phosphatase 2 regulatory subunit B double prime gamma; PPP5C- protein phosphatase 5 catalytic subunit; PRDM1-PR domain-containing protein 1; PRKCB-Protein kinase C beta; PRKCQ-Protein kinase C theta;PTGER4-Prostaglandin E Receptor 4, EP4 subtype; PTPN2-protein tyrosine phosphatase, non-receptor type 2;PTPN22-protein tyrosine phosphatase, non-receptor type 22;PUS10-Pseudouridylate Synthase 10; RAVER1-RAVER1, mouse, homolog of; RIC8B- RIC8, C. Elegans, Homolog Of, B; RNF114-Ring Finger Protein 114; RNF186-Ring Finger Protein 186; RORC-RAR-Related Orphan Receptor C; RPS6KB1-Ribosomal Protein S6 Kinase, 70-Kd, 1; RUNX3-Runt-Related Transcription Factor 3; SBNO2-Strawberry Notch, Drosophila, Homolog Of, 2; SDF4-stromal cell derived factor 4; SETD1A-SET Domain-Containing Protein 1A; SH2B3-SH2B Adaptor protein 3; SKAP2-SRC Kinase-Associated Phosphoprotein 2; SLC9A8-Solute Carrier Family 9 (Zinc Transporter), Member 8; SLC26A3-Solute Carrier Family 26 (Zinc Transporter), Member 26; SLC39A11-Solute Carrier Family 39 (Zinc Transporter), Member 11; SMAD3- Mothers Against Decapentaplegic, Drosophila, Homolog Of, 3; SP140-nuclear body protein SP140; STAT3- signal transducer and activator of transcription 3; STK11-Serine/Threonine Protein Kinase 11; SULT1A2-Sulfotransferase Family 1A, Cytosolic-Phenol Preferring Member 2; TBKBP1-Tank Binding Kinase Binding Protein 1; TBX21-T-Box 21; TEX41-testis expressed 41; THADA-thyroid adenoma associated gene; TLR4-Toll Receptor 4; TNFAIP3- Tumor Necrosis Factor alpha-induced protein 3; TNFRSF1A-Tumor Necrosis Factor Receptor Superfamily, Member 1A; TNFRSF6B-Tumor Necrosis Factor Receptor Superfamily, Member 6B; TNFRSF14-Tumor Necrosis Factor Receptor Superfamily, Member 14; TNFSF8- Tumor Necrosis Factor Ligand Superfamily, Member 8;TNFSF15- Tumor Necrosis Factor Ligand Superfamily, Member 15; TNIP1- TNFAIP3-Interacting Protein 1; TNP2-transition protein 2; TSPAN14-tetraspanin 14; TYK2-tyrosine kinase 2; UBAC2-UBA domain-containing protein 2; UBASH3A-ubiquitin-assoicated and SH3 domain-containing protein A; UBE2L3-ubiquitin conjugating enzyme 2EL 3; UQCR10-ubiquinol-cytochrome c reductase complex, 7.2 kd subunit; ZPBP2-zona pellucida-binding protein 2; ZC3H12C-zinc finger CCCH domain-containing protein 12C; ZMIZ1-Zinc finger MIZ-domain containing 1; ZNF365-zinc finger protein 365; ZNF831-melanoma, cutaneous malignant-susceptiblity to 1.[13]

TABLE 58.4 Functional Networks of Non-MHC Genes in the Susceptibility to Spondyloarthritis

Functional Network	IL-17-Mediated Immunity/IL-23 Pathway	CD8 T-Cell Function	Peptide Processing and Presentation	Microbial Sensing	Nuclear Factor (NF)-κB (NF-κB) Activation	Others
Genes	*IL23R* *TYK2* *IL6R* *IL7R* *IL27* *IL1R2/IL1R1* *IL12B* *JAK2* *RORC* *PTGER4*	*RUNX3* *EOMES* *IL7R* *ITGAL*	*HLA-B* *ERAP1* *UBE2L3* *NPEPPS* *ERN1* *ASAP2*	*CARD9* *NOS2* *NOD2* *IRGM*	*TLR4* *NOD2* *NOTCH1* *TNFAIP3*	*ZMIZ1* *FCGR2A* *KIF21B* *SH2B3* *TNFRSF1A* *GPR65* *SULT1A1* *GPR35* *BACH2* *ICOSLG* *NKX2-3*

The Gut and Spondyloarthritis

Many studies have shown subclinical gut inflammation in SpA patients. Indeed, between 30% and 60% of patients with ReA, undifferentiated SpA and AS had evidence of histological gut inflammation.[23] These observations have raised the speculation that the inciting event in the SpA may be a breakdown of the gut–blood barrier to intestinal bacteria. Patients with AS and their relatives have increased intestinal permeability compared with healthy controls.

The bacterial population inhabiting human intestines, referred to as the gut microbiome (Chapter 23), is a vast microbial community. There is a beneficial relationship between the human system and the gut microbiome. The human gut provides the nutrients that the microbiome metabolizes to provide metabolites such as Vitamin K and short chain fatty acids, which have immunologic capacities and possibly affect the education of the adaptive immune system. The gut microbiome in patients with SpA is different compared to non-SpA populations. However, older studies implicating a role for gut *Klebsiella pneumonia* in the pathogenesis of SpA have been refuted.[13]

There does appear to be an interplay between host genetics (*e.g.*, HLA-B27) and intestinal bacterial composition. This suggests a link between dysbiosis and SpA, upsetting the homeostasis between the microbiome and the host immune systems.[24]

PATHOLOGY OF SPA

Few data exist from synovial or spinal tissues from early disease, and the difficulty with tissue access further complicates this.[25,26] The synovium in SpA has diminished lymphoid aggregates and displays a tortuous vascular morphology compared with the rheumatoid synovium. This may be caused by vascular endothelial growth factor (VEGF) and angiogenic growth factor Ang2, the messenger RNAs (mRNAs) of which have been observed at higher levels in the synovium in PsA compared with RA. VEGF is particularly interesting because it can synergize with RANK ligand (RANKL) to induce bone resorption and also synergize with bone morphogenetic proteins to trigger bone formation, both processes typical of the altered bone remodeling seen in PsA and SpA.[25,26]

Increased production of the scavenger receptor CD163 by macrophages in both the lining and sublining layers is seen in SpA compared with RA.[24] Local production of soluble CD163 inhibits synovial T-cell activation, and levels of synovial CD163 fall with effective treatment. Increased expression of Toll-like receptors 2 and 4 (TLR2, TLR4) has been shown in SpA on $CD163^+$ peripheral blood mononuclear cells in patients with synovitis, which decreases with TNF-α blockade. The interplay between the gut microbiome via the innate immune system appears to play a role in development of SpA. A number of mouse models have shown that SpA does not develop in germ-free environments. Bacteria invade the altered gut barrier, there is activation of TLRs and NF-κB, leading to release of cytokines such as IL-23, IL-1, and IL-6. This, in combination with CD4 T cells. appears to lead to pathogenic Th17 effector cells at sites of inflammation. Joint tissues in patients with SpA have shown lymphocytic infiltrates with IL-17A predominate expression.[27]

Pathological examination of enthesitis in AS demonstrates local inflammation, fibrosis, erosion, and ossification. Immunohistochemical staining for phosphorylated SMAD1/5 in entheseal biopsies of patients with SpA reveals active bone morphogenetic protein signaling.[25] Entheseal T cells are activated by systemic IL-23, leading again to increased IL-17 production.

The pathology of psoriasis consists of an inflammatory cell infiltration in the dermis, with localized increased cytokine production and hyperproliferation of keratinocytes (Chapter 65). CD4 cells are prominent in the dermis, CD8 in the epidermis; both mature T cells respond to peptides presented by antigen-presenting cells (APCs). Differentiation of type 1 (Th1) and type 17 helper T (Th17) cells is triggered in the dermis and multiple other sites, further triggering release of other chemokines. Th17 cells, osteoclast precursors (OCPs), and dendritic cells infiltrate joints from adjacent entheses or the bloodstream. These infiltrating cells express TNF, IL-17, and RANKL and in combination with the increased expression of RANKL by synoviocytes in the lining, leads to the differentiation of OCPs into osteoclasts. This leads to the pathological bone formation in PsA.[8] In addition, $CD14^+$ monocytes that are committed to becoming osteoclasts or osteoclast precursors are increased in the circulation of patients with PsA compared with healthy controls and decline rapidly following treatment with TNF antagonists.

Intracellular and extracellular signaling pathways are being discovered that are important in pathogenesis and therefore are targets for therapy. The Janus kinase (JAK) family of intracellular protein tyrosine kinases mediate signaling pathways of extracellular cytokines and growth factors involved in SpA (Fig. 58.4).

CLINICAL FEATURES

Ankylosing Spondylitis

Musculoskeletal Symptoms

KEY CONCEPTS

Clinical Features of Inflammatory Back Pain

- Low-back pain that is present every day for at least 3 months
- Age of onset <45 years
- Morning stiffness in the back lasting at least 30 min
- Pain that is relieved by exercise and worsened by rest
- Alternating buttock pain
- Relief with nonsteroidal anti-inflammatory drugs

The first symptoms of AS usually appear in adolescence or early adulthood. The hallmark of AS is the presence of inflammatory back pain.[3] Occasionally, the first symptom of AS comes from extraspinal sources, such as AAU, peripheral joint or hip arthritis, or enthesitis, especially in patients with disease onset in childhood.

In patients with AS, the most commonly affected joints outside the spine are the hips. Peripheral arthritis other than in the hips and shoulders is uncommonly seen in patients with AS but, when present, is typical of that seen in other types of SpA, with an asymmetric oligoarthritis presenting predominantly in the lower extremities.

Chest pain, often pleuritic, can be seen in patients with AS because of involvement of the costovertebral joints, with loss of chest expansion and a restrictive ventilatory defect.

Enthesitis is a classic feature of SpA (Fig. 58.5). The most common (and most disabling) sites for enthesitis are in the foot, at the insertion of the Achilles tendon, and of the plantar fascia onto the calcaneus.[28]

Physical measurements that have been validated and recommended by an ASAS Working Group as useful for evaluating patients with AS/AxSpA specifically and with inflammatory back pain in general include forward lumbar spinal flexion (the Schober test), lateral lumbar bending, chest wall expansion, and occiput-to-wall distance.

Extraarticular Manifestations

Uveitis. The anterior portion of the uvea consists of the *iris* and *ciliary body*, and the posterior portion is known as the *choroid*. Inflammation of the anterior uveal tract is known as *anterior uveitis* or *iritis* (Chapter 75). AAU represents the typical uveitis found in SpA, occurring in up to a third of SpA patients, depending on type of SpA and disease duration, most common in AS (20% to 30%).[29] Typically, AAU presents unilaterally with sudden onset, is self-limited, and tends to be recurrent. Symptoms may include redness, eye pain, blurred vision, increased lacrimation, photophobia, and miosis. The diagnosis is confirmed by slit-lamp examination, which is also useful in monitoring treatment responses.

Prognosis is favorable in AAU, with resolution of symptoms typically in 4 to 6 weeks. Although AAU is the most common uveitis associated with SpA, posterior uveitis has been reported, especially in those with coexistent IBD.[1]

Cardiac Manifestations. The characteristic cardiac abnormalities in AS are aortic regurgitation and conduction abnormalities. Less commonly associated cardiac conditions include pericarditis, ascending aortic aneurysm, aortitis, cardiomyopathy, and mitral valve disease. HLA-B27 is an important genetic risk factor for these cardiac conditions. Aortic insufficiency is estimated to occur in up to 34% of patients with AS, especially in greater disease duration.[30]

Cardiac conduction abnormalities, including atrioventricular and intraventricular blocks, are the most common cardiac complication in patients with AS, occurring in up to 9%.[30]

Pulmonary Manifestations. The prevalence of pulmonary involvement is highest (40% to 80%) with high-resolution computed tomography (CT) compared to x-ray or lung spirometry studies, and higher with longer disease duration,[30] although the patient may be asymptomatic. The most frequently recognized manifestations are upper-lobe fibrosis, mycetoma formation, and pleural thickening. The upper-lobe fibrosis may be progressive. Another common finding is the presence of bilateral symmetric apical

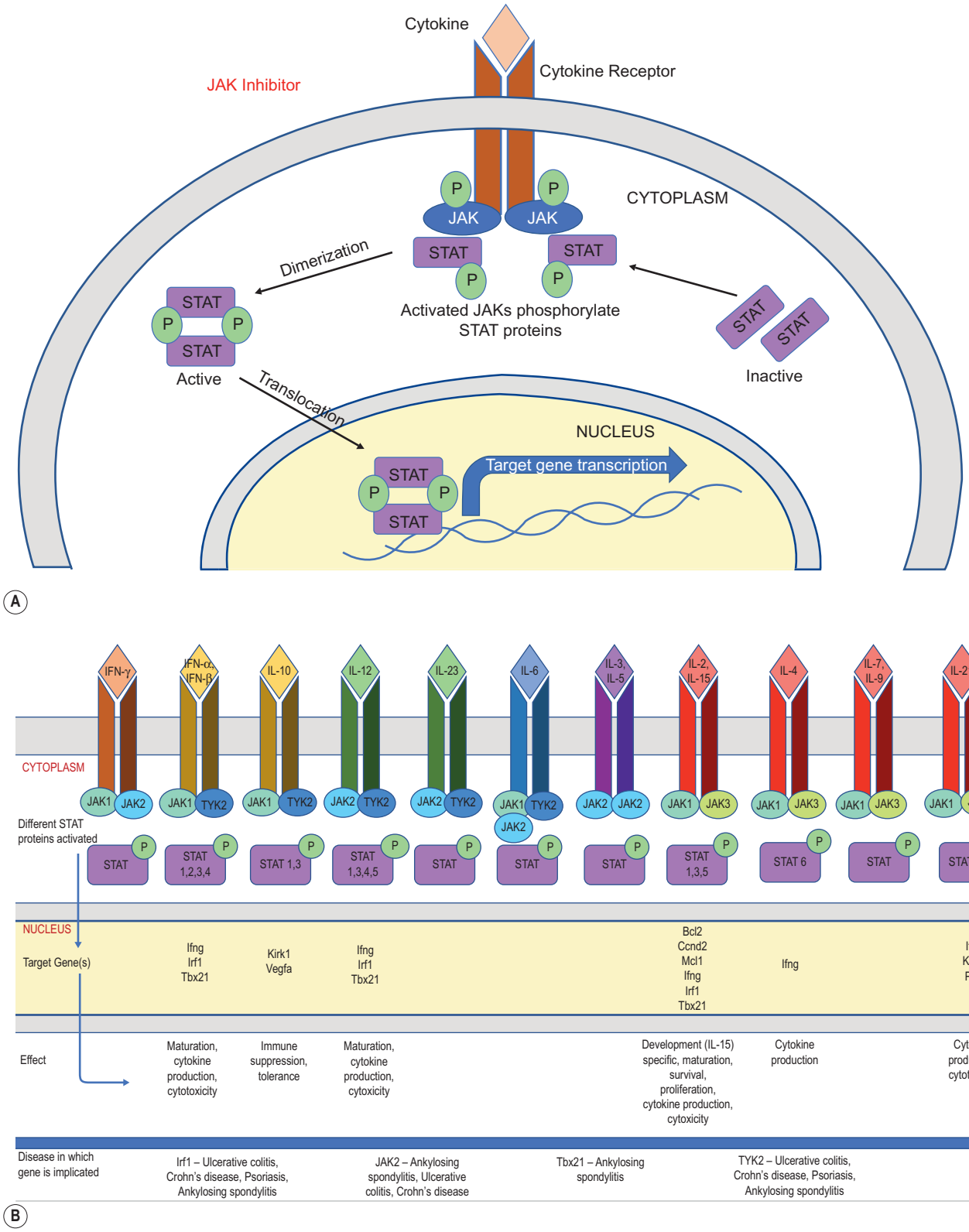


FIG. 58.4 JAK-STAT Pathway. (A) JAK-STAT signaling pathway: Receptor-associated JAKs that are activated by cytokines then autophosphorylate at the receptor subunit. This allows docking of STAT to the JAK, which is followed by phosphorylation of the STAT. The phosphorylated STATs then hetero- or homodimerize, translocate to the nucleus, bind promoter elements, and regulate transcription of target genes. (B) Cytokines signal through certain JAKs, then activating specific STATs. Genes are expressed depending on the JAK-STAT pathway, which then leads to effects. Some of the genes associated with SpAs are listed. (Adapted from Gotthardt D, Trifinopoulos J, Sexl V, et al. JAK/STAT cytokine signaling at the crossroad of NK cell development and maturation. *Front Immunol*. 2019.)

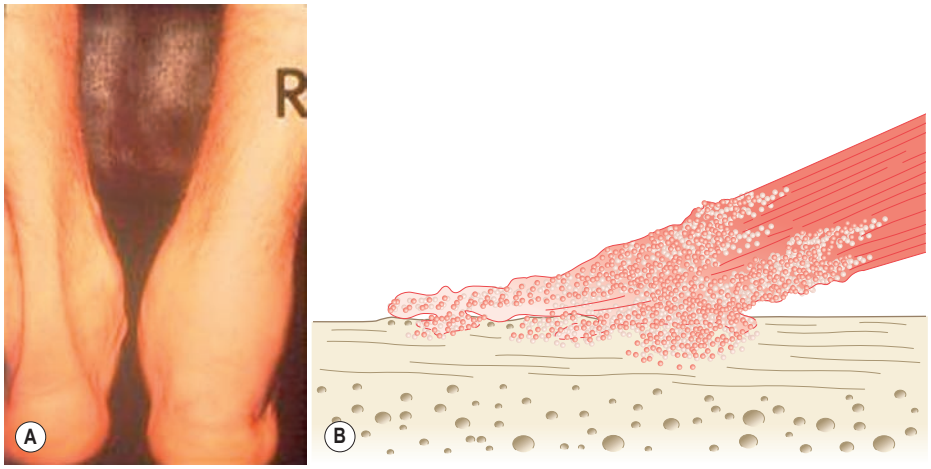


FIG. 58.5 (A) Achilles tendinitis/enthesitis in a patient with reactive arthritis. (B) Schematic drawing of enthesitis, showing periosteal new bone formation, and subchondral bone inflammation and resorption.

pleural thickening. Several recent studies have also shown an increased association with obstructive sleep apnea.[31]

Renal Manifestations. SpA patients have increased risk of renal complications. The most common is immunoglobulin A nephropathy, but other renal complications include secondary renal amyloidosis, NSAID nephropathy, nephrolithiasis, and glomerulonephritis.[32]

Osteoporosis. The prevalence of osteoporosis in AS varies between 11% and 34%.[33] Measuring bone mineral density in patients with spondylitis is complicated by false increases in spinal density from dense syndesmophyte formation, leading some to recommend quantitative CT over standard dual-energy x-ray absorptiometry (DEXA) for bone mineral density measurements, although the consensus is still to screen by DEXA.[46] Trabecular bone score (TBS) obtained from the spinal DEXA scan predicts major osteoporotic and spinal fractures in AS independent of Fracture Risk Assessment Tool (FRAX) score.[34] This has potential use for fracture prediction.[34]

Spondylodiscitis and Spinal Fractures. An uncommon but well-recognized complication of AS is spondylodiscitis, a destructive discovertebral lesion also called *Andersson lesion*. Typically, these lesions are confined to the thoracolumbar spine, sometimes with multiple-level involvement. Spondylodiscitis usually occurs at an advanced stage of AS under the form of an erosive condition related to both inflammatory, mechanical factors, and osteoporosis. The patient may or may not have a history of preceding trauma.

Even trivial falls can be catastrophic for AS patients, who are at risk for spinal fractures because of spinal rigidity and osteoporosis. The estimated prevalence of fractures in AS varies from 11% to 25%.[33] Fractures through the disk space, the weakest point in the ankylosed spine, are most common. The cervical spine is the most frequently affected region, followed by the thoracolumbar junction, and may or may not be complicated by spinal cord injury.

Neurological Manifestations. Neurological involvement in AS is most often related to spinal fracture or cauda equina syndrome. The cauda equina syndrome in AS is characterized by a slow insidious progression and a high incidence of dural ectasia, although a rapid onset secondary to a traumatic event has been reported. It tends to be a late manifestation of AS.[35] Patients with cauda equina syndrome present with a prodrome of sensory, motor, or reflex loss before progression to bowel or bladder incontinence. About half the patients have pain in the rectum or lower limbs that is presumably neurogenic in origin.

Fatigue and Psychosocial Manifestations. Fatigue and sleep disturbance are common in AS. Sleep disturbance is multifactorial, with age, anxiety, depression, nocturnal back pain, extra-spinal disease, and duration of delay in diagnosis all being important.[36] The prevalence of depression in AxSpA varies widely depending on the criteria used, ranging from 11% to 64%.[37] Depression is associated with worse disease activity, greater functional impairment, long-term smoking, and poor global health scores.[37]

AS in Women. There tends to be a greater delay in the diagnosis of AS in women. This may be in part due to higher incidence of nr-AxSpA in women,[13] with less radiographic damage and radiographic progression.[38] Women have a higher frequency of enthesitis, psoriasis, and IBD.[38] Studies have also shown higher disease activity but poorer response to TNFi treatment in women.[38] Women tend to have less severe involvement of the spine but more peripheral joint involvement. A large review of the impact of AS on reproductive events concluded that AS did not adversely affect the ability to conceive, pregnancy outcome, or neonatal health.[39]

Reactive Arthritis

The classic triad of arthritis, urethritis, and conjunctivitis, representing what was formerly known as *Reiter syndrome*, is a presenting feature of only a minority of patients with ReA (comprising only a third of the cases in some series). In ReA, the clinical features are viewed more as a spectrum ranging from the classic triad to undifferentiated SpA.

Typically, the features start 1 to 4 weeks after a triggering event, frequently an enteric or urogenital infection, but often the event passes unnoticed without any specific symptoms. The syndrome starts with constitutional symptoms, such as fatigue, malaise, and fever, followed by a sterile, asymmetrical, additive lower-extremity oligoarticular or monoarticular inflammatory arthritis. Upper-extremity involvement is less common. Dactylitis occurs in the toes or fingers, resulting in "sausage digits," which represent inflammation not only of the interphalangeal joints but also of the surrounding soft tissue structures, including tendons and subcutaneous tissue.

Sacroiliitis and spondylitis are less common than peripheral arthritis, although inflammatory back pain does occur.

Unilateral and bilateral sacroiliac involvement and even spondylitis occur, especially in those with chronic or long-standing disease. As with AxSpA, enthesitis is most common in the Achilles tendon and plantar fascia insertions, although tenderness over the symphysis pubis, iliac crest, ischial tuberosity, greater trochanters, and costochondral junctions may also occur.

Mucocutaneous lesions may be difficult to distinguish from psoriasis, especially circinate balanitis and keratoderma blennorrhagica. Circinate balanitis is an ulcerative mucosal lesion over the glans or shaft of the penis that is demarcated by a serpiginous erythematous border. Keratoderma blennorrhagica is a painless desquamative psoriatic-like papulosquamous eruption and is sometimes referred to as *pustulosis palmoplantaris* and occurs on the palms and soles of the feet. Oral lesions present as shallow, painless ulcers or patches on the palate and tongue or mucositis of the soft palate and uvula. Conjunctivitis and AAU also occur, as described in AS. Unilateral or bilateral conjunctivitis is usually an early feature manifesting with irritation, erythema, and lacrimation.

Juvenile Spondyloarthritis

JSpA usually manifests as enthesis and/or arthritis, otherwise known as seronegative enthesopathy and arthropathy (SEA) syndrome or enthesitis-related arthritis (ERA). Enthesitis is more common and can affect more sites than in adult SpA. Less commonly, JSpA can manifest as juvenile AS, psoriatic arthritis, reactive arthritis, and arthritis associated with IBD. Unlike adult SpA, spine or sacroilliac (SI) joint involvement is uncommon at presentation. Arthritis is typically more peripheral, apart from the hip, and the lower extremities tend to be more commonly affected. Tarsitis is unique to juvenile SpA and is inflammation of the intertarsal bones and overlying tissues, which causes midfoot pain and swelling. Involvement of the spine or SI joint can develop over 5 to 10 years.[40]

Psoriatic Arthritis

Skin involvement exhibits five clinical patterns (Table 58.5). Usually psoriatic arthritis occurs after onset of psoriasis, with a latency of 10 years on average,[7] although PsA can precede skin disease or occur simultaneously in 15% of patients.[8] The joint disease occurs in five different subtypes, as defined by the Moll and Wright classification (Fig. 58.6), including oligoarticular, polyarticular, distal interphalangeal (DIP)-predominant, axial or spondyloarthritis, and arthritis mutilans.[8] Oligoarticular PsA tends to be asymmetric, affecting four or fewer joints. Polyarticular PsA may be more symmetric and affects five or more joints. The distal subtype usually occurs with other subtypes but can occur alone in 5% of patients. Axial subtype affects the spine and SI joints. Arthritis mutilans is the deforming subtype with telescoping, flail digits, and marked bone resorption.[8] Enthesitis is encountered in 30% to 50% of patients. Acute or chronic dactylitis is reported in 40% to 50% of patients and is often associated with severe articular disease.[8] Extraarticular manifestations with nail disease including pitting or onycholysis is common, whereas uveitis is less common,[8] occurring in less than 10% of patients with PsA.[8]

TABLE 58.5 Skin Involvement in Psoriasis

Clinical Pattern of Skin Involvement	Description
Plaque	Most common type. Also called psoriasis vulgaris. Scaly erythematous plaques on the trunk and extensor surfaces
Inverse	Slightly erosive plaques in intertriginous locations
Guttate	Usually affects children or adolescents. Often triggered by group-A streptococcal tonsil infection
Pustular	Localized or generalized coalescing sterile pustules. Typically affects palms and soles
Erythrodermic	Most severe type, covering about 90% of the total body surface

From: Rendon A, Schäkel K. Psoriasis pathogenesis and treatment. *Int J Mol Sci.* 2019;20(6).

Enteropathic Arthritis

The arthritis associated with inflammatory bowel disease (IBD), called enteropathic arthritis, is most commonly nondestructive and reversible. Two patterns of peripheral arthritis have been recognized. Type 1 is pauciarticular, involving the knees and ankles more than the upper extremities. It tends to resolve in less than 10 weeks, may precede the diagnosis of IBD, runs parallel to intestinal disease, and is associated with HLA-B27, -B35, and -*DRB1*01:03*.[41,42] The second type has a polyarticular presentation, is more likely to involve the metacarpophalangeal (MCP) and proximal interphalangeal (PIP) joints than the lower extremities, and is more likely to have a chronic course. It may run independent of intestinal disease and is associated with HLA-B44.[41,42] A third type has more recently been described, involving both peripheral and axial arthritis.[41] The symptoms of peripheral arthritis tend to coincide with activity of the bowel disease in ulcerative colitis (UC) but not Crohn disease (CD). Total colectomy is associated with remission of arthritis in half of patients. In contrast, axial involvement may precede the development of IBD, has no gender predilection, and resembles the development of AS. The axial symptoms do not parallel activity of IBD. In addition to spondylitis, an isolated sacroiliitis occurs that is often asymmetric and not associated with HLA-B27. Axial involvement is more common in patients with CD than those with UC.[41]

Mucocutaneous complications of IBD include erythema nodosum, which occurs in up to 15% of those with CD and up to 10% in those with UC, although other reports showed lower frequency[41]; pyoderma gangrenosum, more rare and severe, with a prevalence of less than 2% in IBD patients; and, rarely, erythema multiforme or Sweet syndrome.[41] Oral lesions, including painful aphthous ulcers, periodontitis, aphthous stomatitis, and pyostomatitis vegetans, can parallel underlying IBD.[42]

Episcleritis is the most common ocular manifestation associated with IBD and is more associated with intestinal disease activity than other ocular manifestations. Anterior uveitis in IBD patients is more often bilateral, insidious in onset, chronic, and does not parallel intestinal activity.[43]

Undifferentiated Spondyloarthritis

Patients who do not meet criteria or clinical features of the "classic" SpA are regarded as having undifferentiated SpA, accounting for about 40% of patients at presentation (Table 58.6).[44] Follow-up studies suggest that about one-third will go into remission over time and more than half will develop a "classic" SpA, usually AS.[44]

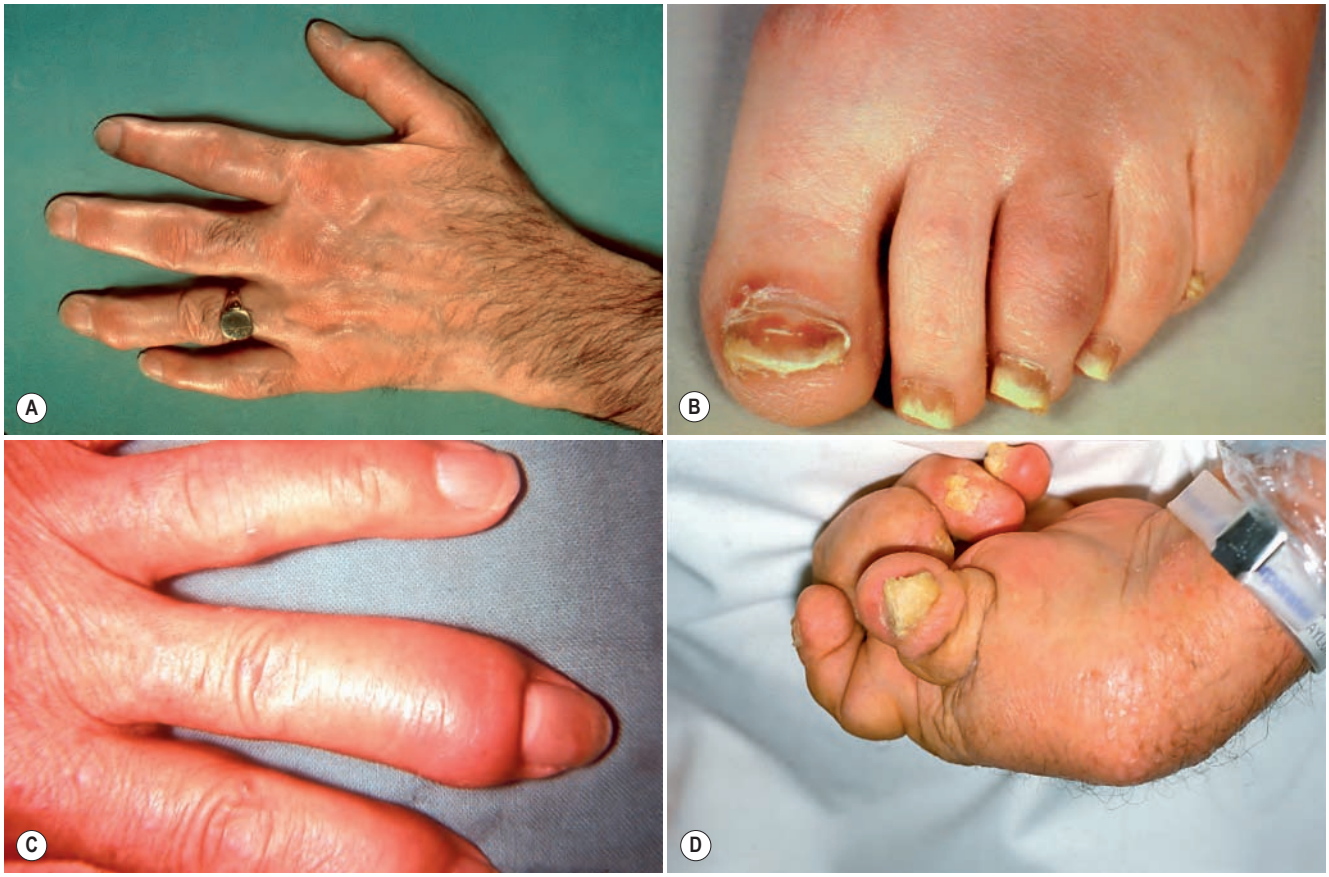


FIG. 58.6 Patterns of Psoriatic Arthritis. (A) Rheumatoid-like distribution; (B) sausage digits; (C) distal interphalangeal involvement; and (D) psoriatic arthritis mutilans.

CLINICAL RELEVANCE

Utility of Human Leukocyte Antigen (HLA)-B27 Testing in the Evaluation of Inflammatory Back Pain and Spondyloarthritis

- Not indicated where the diagnosis is unquestionable, as it has little value in prognosis.
- Although patients with SpA of African and Middle Eastern ancestry are often HLA-B27 negative, the finding of HLA-B27 in these patients has higher predictive value.
- Most useful in patients with either inflammatory back pain without radiographic changes or with other features of spondyloarthritis (unexplained lower extremity arthritis in a young adult, uveitis, etc.).

Measures of SpA Activity and Severity

In the past few years, outcome measures have been developed and validated to quantitate disease activity and severity; these are summarized in Table 58.7. These instruments are extensively validated and easy to administer in clinical practice and have been shown to perform well in clinical trials. There is no consensus on which are preferable to use.

Radiographic Imaging of Spondyloarthritis

Axial Spondyloarthritis

Radiographic AxSpA (AS) is the demonstration of sacroiliitis on plain radiographs (Fig. 58.7).[3] One problem with radiographic

TABLE 58.6 Frequencies of Different Symptoms and Signs in Patients With Undifferentiated Spondyloarthritis

Feature	Percent (%)
Demographic	
Males	62–88
Mean age at onset (years)	16–23
Clinical	
Low-back pain	52–80
Peripheral arthritis	60–100
Polyarthritis	40
Enthesopathy	56
Heel pain	20–28
Mucocutaneous involvement	16
Conjunctivitis	33
Genitourinary disease	28
Inflammatory bowel disease	4
Cardiac abnormalities	8
Laboratory	
Elevated erythrocyte sedimentation rate	19–30
Human leukocyte antigen (HLA)-B27 positive	80–84
Radiographic	
Sacroiliitis	16–30
Spinal radiographic changes	20

Adapted from Chen CH, Lin KC, Yu DT, et al. Serum matrix metalloproteinases and tissue inhibitors of metalloproteinases in ankylosing spondylitis: MMP-3 is a reproducibly sensitive and specific biomarker of disease activity. *Rheumatol (Oxf).* 2006;45:414–420.

TABLE 58.7 Measurements of Disease Outcome in Spondyloarthritis

Ankylosing Spondylitis

1) Disease activity
   (a) Bath Ankylosing Spondylitis Disease Activity Index (BASDAI)
   (b) Ankylosing Spondylitis Disease Activity Score (ASDAS)
   (c) Patient and Physician Global Assessments
2) Function
   (a) Bath Ankylosing Spondylitis Functional Index (BASFI)
   (b) Dougados Functional Index
3) Quality of Life
   (a) SF-36
   (b) Ankylosing Spondylitis Quality of Life Index (ASQOL)
   (c) ASAS Health Index
4) Metrometry
   (a) Schober test (lumbar flexion)
   (b) Chest expansion
   (c) Occiput-to-wall distance
   (d) Lateral bending
   (e) Bath Ankylosing Spondylitis Metrology Index (BASMI)
5) Imaging
   (a) Standard radiography, computed tomography, magnetic resonance imaging
   (b) Modified Stroke Ankylosing Spondylitis Spinal Score (mSASSS)
   (c) Bath Ankylosing Spondylitis Radiographic Index (BASRI)
6) Assessment in Ankylosing Spondylitis (ASAS) 20
   (a) An improvement of 20% and absolute improvement of 10 units on a 0–100 scale in three of the following four domains:
      i. Patient global assessment (by visual analogue scale [VAS] global assessment)
      ii. Pain assessment (the average of VAS total and nocturnal pain scores)
      iii. Function (represented by BASFI)
      iv. Inflammation (the average of the BASDAI's last two VAS concerning morning stiffness, intensity, and duration)
      v. Absence of deterioration in the potential remaining domain (deterioration is defined as 20% worsening)

Psoriatic Arthritis

1) Arthritis
   (a) ACR response criteria
   (b) Psoriatic Arthritis Response Criteria (PsARC)
2) Skin response
   (a) Psoriasis Area and Severity Index (PASI)
   (b) Target lesion score
   (c) Static global assessment
3) Quality of life (HAQ, SF-36, DLQI)
4) Radiographic
5) Composite measures
   (a) Psoriatic Arthritis Disease Activity Score
   (b) Composite Psoriatic Disease Activity Index
   (c) Group for Research and Assessment of Psoriasis and Psoriatic Arthritis (GRAPPA) Composite Exercise (GRACE) instrument

imaging is that up to 10 years may elapse from the onset of inflammatory back pain to the appearance of radiographic sacroiliitis.[3] The introduction of MRI of the spine and entheses has allowed not only correct anatomical description of spinal structures but also differentiation of AS-related and unrelated inflammatory spinal lesions earlier than is possible with standard radiography.[45] MRI of the sacroiliac joints and spine is currently the only imaging tool to localize and quantify spinal inflammation accurately (Fig. 58.8) and is being developed as a measure of disease activity and treatment response.

Two outcome instruments have been introduced in the assessment of disease damage and progression in AS: the Bath Ankylosing Spondylitis Radiographic Index (BASRI) and the modified Stroke Ankylosing Spondylitis Scoring System (mSASSS) (see Table 58.7). As a rule, these instruments have a low sensitivity to change (7.5% over 2 years).

Psoriatic Arthritis

PsA has characteristic radiographic manifestations, including asymmetrical involvement, involvement of the DIP joints, and the classic "pencil-in-cup" deformities. Also seen are periostitis, bony ankylosis, and eccentric erosions with new bone formation. Radiographic severity is quantitated using different scoring methods: the modified Steinbrocker global scoring, the modified (from RA) Sharp/van der Heijde method, and PsA Ratingen score (PARS).[47]

There are a number of disease impact quantification methods used in assessing PsA (see Table 58.7).

DISEASE COURSE AND PROGNOSIS

Ankylosing Spondylitis

AS significantly impacts the lives of those affected. Patients with AS are more likely to be work-disabled, not participate in the labor force, and more likely to have never married or to be divorced. Women with AS were less likely to have had children.[49]

There has been some debate about whether non-radiographic-AxSpA is a self-limited form of AxSpA with a more favorable course, an early stage within the disease spectrum, or even a different disease.[48] Only about 5% of patients with early AxSpA change from nr-AxSpA to radiographic-AxSpA.[48] A recent study showed the incidence of peripheral and extra-articular manifestations of nr-AxSpA and r-AxSpA to be similar and with equivalent disease burden over 5 years of follow-up.[48]

Although AS is a chronic condition that can frequently have an unpredictable course, some studies suggest that those with higher levels of disease activity early in the course of the disease are more likely to have active disease in the future (Fig. 58.9).[49] Hip involvement is a predictive factor for severe disease.[45] Other factors that may suggest a worse outcome include ESR greater than 30 mm/h; unresponsiveness to NSAIDs; limitation of motion of the lumbar spine; dactylitis; oligoarthritis; or onset at less than 16 years.[49]

A growing body of data have shown that patients with AS are at risk for early mortality as a result of cardiovascular disease.[50] However, the impact of newer agents—such as anti-TNF and anti–IL-17 drugs—on the natural history of this disease remains to be seen.

Reactive Arthritis

Most cases tend to remit within 6 months of onset. For those that last longer than 6 months, this is considered a sign of development of chronic ReA.[51] An American study with 5 years of follow-up showed that one-third of patients fully recovered and two-thirds continued to have subjective complaints, with half of the total patients developing chronic arthritis.[51] In a Finnish study with 11 years of follow-up, 16% of patients had chronic arthritis. The majority of these patients were also HLA-B27 positive and some developed sacroiliitis.[51]

Psoriatic Arthritis

Deformities and joint damage occur in many PsA patients. Within the first 2 years of diagnosis, 47% of patients have bony erosions despite use of disease-modifying antirheumatic drugs (DMARDs). PsA is also associated with increased risk

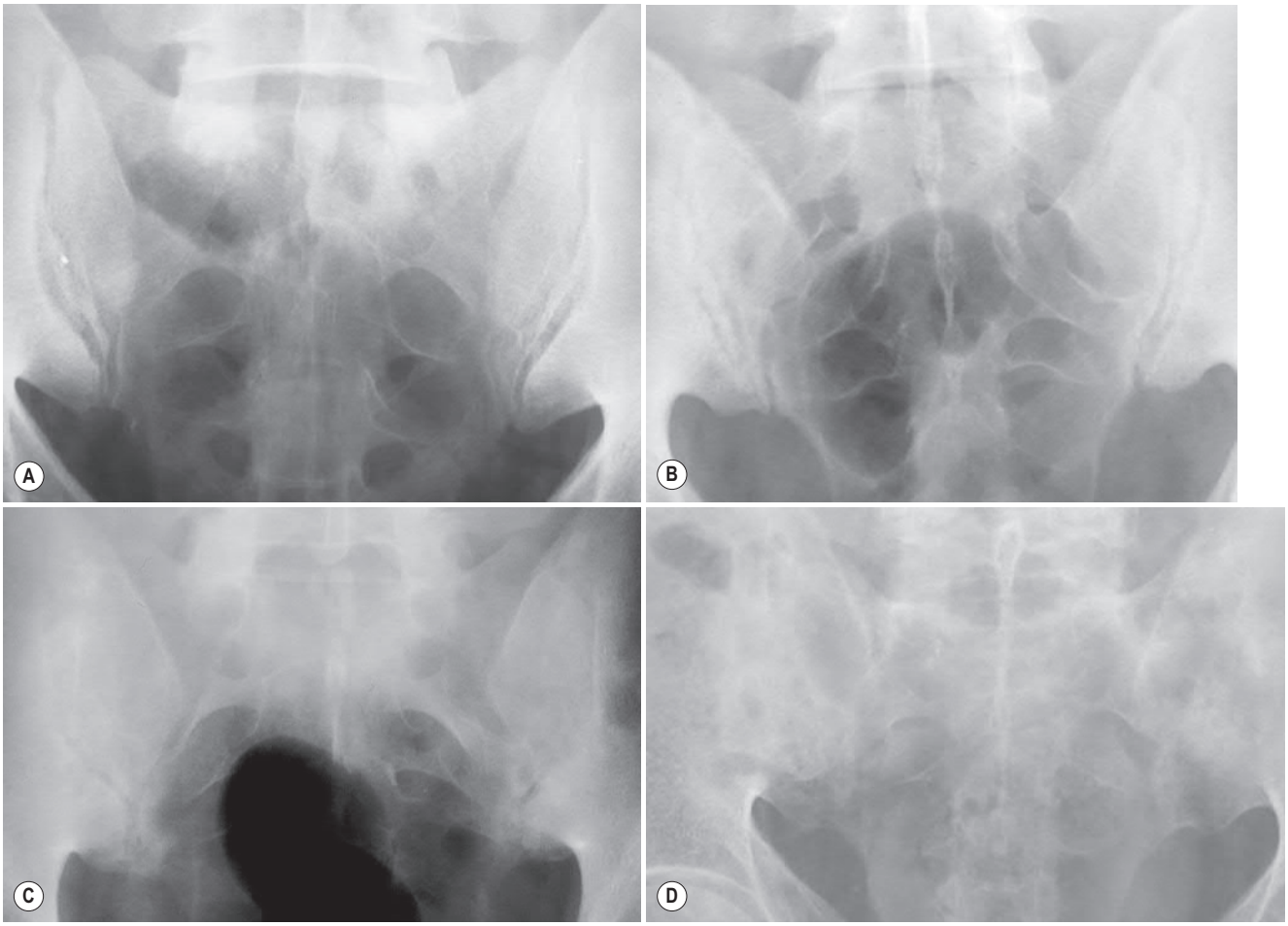


FIG. 58.7 Grading of Radiographic Sacroiliitis. (A) Grade 0–1 (normal); (B) grade 2–3, with sclerosis and small erosions; (C) advanced grade 3, with joint space narrowing and large erosions; and (D) grade 4 (total sacroiliac fusion).

of cardiovascular events, hypertension, diabetes, metabolic syndrome, and fatty liver. However, mortality rates have decreased and are now similar to the general population.[8]

Juvenile Spondyloarthritis

Although not extensively studied, the prognosis in JSpA is guarded.[40] Available data suggest that children with disease activity for greater than 5 years are more likely to be disabled. In fact, the probability of remission was only 17% after 5 years of disease. Nearly 60% of children with JSpA have moderate to severe limitation after 10 years of disease. What is not clear is the extent to which the outcome in juvenile-onset AS is different from that in adult-onset disease.

TREATMENT

Evidence-based recommendations have recently been updated for treatment of AS and nr-AxSpA by the American College of Rheumatology (Table 58.8).[46]

Patient Education and Physiotherapy

A great deal of educational information is available for patients (http://www.spondylitis.org and http://www.arthritis.org). Unsupervised recreational exercise improves pain and stiffness, and back exercise improves pain and function in patients with AS and other types of SpA, but these effects differ with disease duration. Health status is improved when patients perform recreational exercise at least 30 minutes/day and back exercises at least 5 days/week.

THERAPEUTIC PRINCIPLES

Treatment Principles for Medical Management of Spondyloarthritis

- Patient education, regular exercise, smoking cessation, and physiotherapy should be initiated early in the disease course.
- Nonsteroidal anti-inflammatory drugs (NSAIDs) remain the "first-line" treatment.
- Disease-modifying antirheumatic drugs (DMARDs: sulfasalazine) are used for peripheral arthritis.
- Intraarticular/intralesional corticosteroid injections are administered.
- Biological agents for axial disease refractory to NSAIDs, peripheral arthritis refractory to DMARDs, and entheseal lesions refractory to NSAIDs.
- It is important to remember to treat coexistent/complicating conditions (inflammatory bowel disease [IBD], psoriasis, osteoporosis, premature atherosclerosis).

Medical Treatment

Nonsteroidal Anti-Inflammatory Drugs

NSAIDs remain the starting point of treatment, and many patients will attain satisfactory symptom control with these

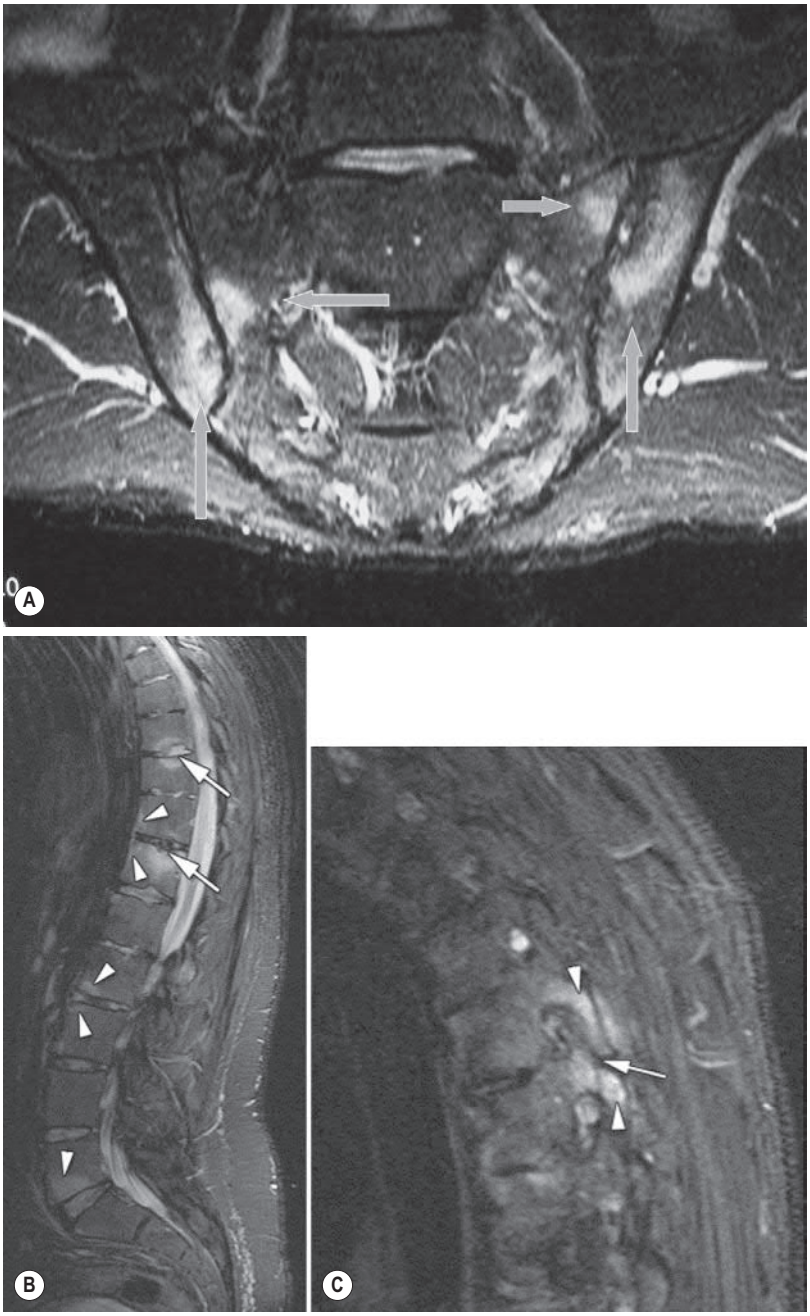


FIG. 58.8 (A) Magnetic resonance imaging (MRI) of the sacroiliac joints, showing areas of marrow edema (*indicated by arrows*) on STIR sequences. (B) Lateral spine, showing enhancement of the insertion of the annulus fibrosis on the disk (*arrowheads*) and subchondral bone (*arrows*). (C) Involvement of the subchondral bone of the apophyseal joints.

agents alone. There are no strong data to suggest the superiority of any specific NSAID in patients with SpA. There is inconsistent evidence that continuous NSAID use results in slower rates of radiographic progression over 2 years, compared to on-demand use of NSAIDs. As such, the new 2019 ACR/SAA/SPARTAN treatment guidelines conditionally recommend use of continuous treatment over on-demand treatment with NSAIDs for controlling disease activity in active AS.[46] COX-2 antagonists are recommended mainly for patients with proven peptic ulcer disease. Of concern is the association of the use of NSAIDs with flares of colitis, suggesting they should be used with care in this setting.

Disease-Modifying Anti-Inflammatory Drugs

DMARDs may be considered in patients whose peripheral arthritis is uncontrolled despite NSAIDs or when TNFi is not available.[46]

Sulfasalazine. The efficacy of sulfasalazine in the treatment of peripheral joint involvement in AS and other SpAs is well established. Coincident with improvement in peripheral arthritis is a fall in acute-phase reactants, such as the ESR and CRP. Most of the evidence to date shows there is little benefit for treatment of axial disease. Consideration may be given in patients with contraindications to or limited access to TNFi or biological agents or for those who decline treatment with biological agents.[46]

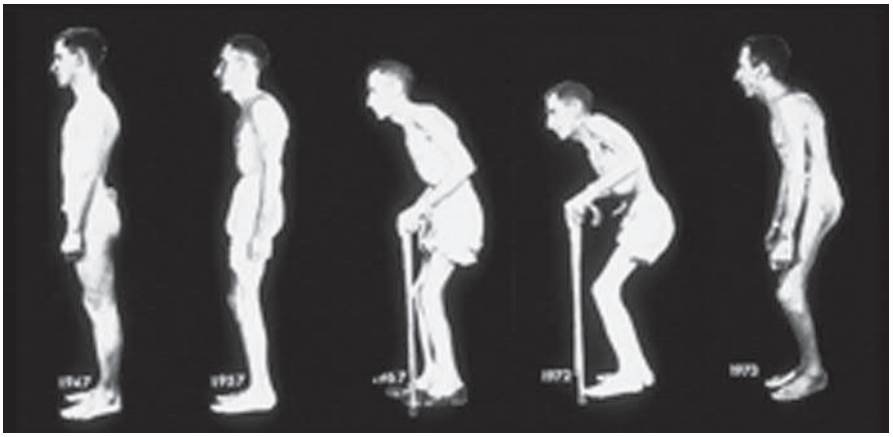


FIG. 58.9 The "Classic" Course of Ankylosing Spondylitis. Disease progression from shortly after disease onset in 1947 until just before the patient's death in 1973. The slight improvement between 1972 and 1973 was as a result of his having undergone total hip arthroplasties.

TABLE 58.8 Treatment of Spondyloarthritis

- Patient education
- Physiotherapy
- Medications
- Nonsteroidal anti-inflammatory drugs
- Disease-modifying antirheumatic drugs:
  - Sulfasalazine (especially for peripheral arthritis)
  - Methotrexate (especially for peripheral or psoriatic arthritis)
  - Apremilast (especially for psoriatic arthritis)
  - Abatacept (especially for psoriatic arthritis)
- Corticosteroids:
  - Systemic
  - Intraarticular, intralesional
- Biological agents:
  - Tumor necrosis factor blockers
  - Interleukin (IL)-17 blockers
  - Interleukin (IL)-12/23 and (IL)-23 blockers or (IL)-12/23 blockers
  - Janus kinase inhibitors
- Treatment of osteoporosis
- Surgery
- Hip replacement
- Corrective spinal surgery

Methotrexate. Although less well studied than sulfasalazine, methotrexate has been shown to be effective in some studies of peripheral arthritis and psoriasis in patients with AS and other SpA. Its efficacy in treating axial SpA has not been established.

Other DMARDs. For AS, leflunomide, apremilast, and thalidomide are not recommended.[46] Apremilast (an oral phosphodiesterase 4 inhibitor) and abatacept (selectively modulates T-cell activation) have been shown to be effective in psoriatic arthritis but not in AS.[52]

Corticosteroids. Although not well studied in patients with AS, many clinicians add low-dose glucocorticoids to the management of active SpA where NSAIDs or DMARDs fail to achieve a satisfactory response. On occasion, pulse steroids have also been utilized. Given the lack of controlled data as to their effectiveness, the side effects of long-term glucocorticoid therapy (including osteoporosis, a major cause of morbidity in AS patients, and possible worsening of psoriasis), and the emergence of more effective treatments, their use is not recommended unless more effective treatments are not available.

Intraarticular/Intralesional Corticosteroids

Intraarticular and peritendinous injections of depot steroid preparations are frequently employed by clinicians for symptomatic relief of local flare-ups, although they have not been extensively studied in controlled trials. Injecting around the Achilles or patellar tendons is generally not recommended because of the risk of tendon rupture.

Antibiotics

There is no convincing evidence that treatment of gastrointestinal infection alters the course of ReA, although antibiotic therapy may be warranted in severe bacterial GI infection. Benefits shown in one trial of antibiotics in *Chlamydia*-induced ReA have not yet been confirmed.[53] Overall, there is little evidence that antibiotics have a place in the management of ReA or other SpA.

Tumor Necrosis Factor-α Blockers

This category of medications has been shown to be effective in controlling inflammation and improving function in patients with AS (Table 58.9). Currently five TNFi agents have been approved for use by the Food and Drug Administration (FDA) for the treatment of AS in the United States: infliximab, an infusion of a chimeric mAb to TNF-α (infliximab) at 5 mg/kg of infliximab every 6 to 8 weeks; etanercept, given 50 mg subcutaneously weekly; adalimumab, which is used at a dose of 40 mg administered subcutaneously every other week; golimumab, at a dose of 50 mg administered subcutaneously monthly or 2 mg/kg intravenously every 8 weeks; and certolizumab, at either 200 mg administered subcutaneously every other week or 400 mg once a month. The onset of action is quite rapid, usually following the first infusion or injection. There is no evidence to support use of one TNFi over another for musculoskeletal disease, although adalimumab and infliximab are recommended over etanercept for the treatment of patients with AS with recurrent uveitis. Etanercept is also not approved for use in Crohn disease or ulcerative colitis, so the use of another TNFi would be a better choice for patients with AS and coexisting IBD.[46] These agents are effective also in PsA and nr-AxSpA.[46] Improvement was seen not only clinically but also radiographically, with improvement of lesions suggestive of disease activity on MRI, and, in studies extending greater

TABLE 58.9 2016 Update of the ASAS-EULAR Recommendations for Use of Biological Therapy in Axial Spondyloarthritis[a]

1. For the initiation of biological therapy:
   (a) A diagnosis of definitive AxSpA by a rheumatologist.
   (b) Elevated CRP, the presence of inflammation on MRI of the SI joints and/or spine, or the presence of radiographic sacroiliitis (according to modified New York criteria).
   (c) Presence of refractory disease defined by:
      (i) All patients: failure of at least 2 NSAIDS over 4 weeks in total.
      (ii) Patients with predominant peripheral manifestations: failure of one local steroid injection if appropriate or failure of therapeutic trial of sulfasalazine.
   (d) High disease activity with either ASDAS ≥ 2.1 or BASDAI ≥ 4.
   (e) Favorable opinion of the rheumatologist that benefit outweighs the risk, including considering potential contraindications to biological therapy.
2. For the monitoring of biological DMARD: both the BASDAI and the ASAS core set for clinical practice should be followed regularly
3. For the continuation of biological DMARD, consideration should be made after 12 weeks' treatment. Response is defined as improvement when:
   (a) ASDAS improvement ≥1.1 or BASDAI improvement ≥ 2 (0–10),[b] and
   (b) Rheumatologist opinion that treatment should be continued.

[a]van der Heijde D, Ramiro S, Landewe R, et al. 2016 update of the ASAS-EULAR management recommendations for axial spondyloarthritis. *Ann Rheum Dis.* 2017 Jun;76(6):978–991.
[b]Either ASDAS or BASDAI can be used but needs to be the same measure per patient.
ASDAS, Ankylosing Spondylitis Disease Activity Score; *AxSpA*, axial spondyloarthritis; *BASDAI*, Bath Ankylosing Spondylitis Disease Activity Index; *MRI*, magnetic resonance imaging; *SI*, sacroiliac; *NSAIDs*, nonsteroidal anti-inflammatory drugs.

than 4 years, slowing progression on standard radiographs, especially in those with shorter disease duration treated for an extended period.[54]

Interleukin-17 Blockers

Secukinumab and ixekizumab are anti–IL-17A monoclonal antibodies that have been shown to control the symptoms of active AS and PsA, and both have been approved by the FDA. The impact of these medications on other disease features, such as radiographic progression, remains to be determined. IL-17 blockers may be considered if TNF blockers are contraindicated such as in congestive heart failure (CHF) or demyelinating disease. Both medications have been associated with new onset or exacerbation of inflammatory bowel disease.[46]

Interleukin-12/23 Blockers

Ustekinumab is an IL-12/23 inhibitor that has been approved for treatment of psoriasis, PsA, and IBD. Tildrakizumab, risankizumab and guselkumab are IL-23 inhibitors which have shown benefit in PsA, although their efficacy in AxSpA has not been demonsrtated.[52]

Janus Kinase Inhibitors

Tofacitinib is an oral JAK inhibitor that may be another option in axial disease. Phase III study results are not available, but a phase II study showed benefit in clinical and radiographic changes in axial disease over 12 weeks.[46] Tofacitinib has been FDA approved for psoriatic arthritis and moderately to severely active ulcerative colitis. A number of other JAK inhibitors are being examined in the treatment of AxSpA.

Surgical Treatment of Ankylosing Spondylitis Complications

Because the hip is the joint most commonly involved in patients with AS, total hip arthroplasty is the most common surgical procedure.[55] Heterotopic new bone formation may be a potential problem.

Limited prevalence data suggest that patients with AS, even those with mild disease, are at increased risk for vertebral fracture, often resulting in neurological compromise.[56] Surgical options vary depending on acuity of fracture or dislocation and if there are neurological signs. In general, custom-fitted halo vest immobilization is recommended. Posterior fixation with or without laminectomy and anterior grafting if there is a severe bone gap may be necessary depending on the acuity and severity, but wedge osteotomies are typically considered first if possible. The fixed kyphotic deformities seen in patients with advanced AS are of considerable distress to patients and may result in substantial functional impairment. Indications for cervical extension osteotomy are the loss of horizontal gaze or the onset of breathing and eating difficulties due to stiff kyphosis.[57]

CONCLUSIONS AND RESEARCH OPPORTUNITIES

ON THE HORIZON

Research Opportunities in Spondyloarthritis

- Improved understanding of pathogenic mechanisms of spondyloarthritis (SpA)
- Elucidation of the roles of non–major histocompatibility complex (MHC) genes in SpA
- Definition of the link between gut microbiome and ankylosing spondylitis (AS)
- Improved measures of treatment outcomes
- Advances in biological therapies of SpA

Progress has been made in the classification and epidemiology of SpA, particularly in the elucidation of the factors involved in SpA pathogenesis. It has become clear that HLA-B27, followed by ERAP1 and IL-23R, are important genetic risk factors in SpA but are not sufficient alone to cause disease. GWAS and whole-genome sequencing continue to identify susceptibility genes for SpA, including those in the antigen processing and Th17 pathways, some of which are shared by some or all SpA and others that are specific for a given disease.

This continues to open doors for new treatment strategies, including IL-12/23 or JAK inhibition. However, why IL-12/23 or IL-23 inhibition is an ineffective target in AS is not understood and means further studies on the IL-23–IL-17 axis are needed. Further research may help to give a broader choice of therapies for SpA patients.

The link of gut inflammation to the triggering of AS is strongly suggested by data thus far, especially its possible link to the gut microbiome. How inflammation and the gut microbiome contribute to SpA pathogenesis is insufficiently understood.

Continued improvement in imaging techniques and specificity, early diagnosis of SpA, understanding pathological difference in nr-AxSpA and r-AxSpA, development of biomarkers, and understanding relationship between axial and peripheral disease are some of the many important unmet needs.[52] Understanding these in combination with the advances in treatment hold promise for a better future for patients with these diseases.

REFERENCES

1. Monnet D, Breban M, Hudry C, et al. Ophthalmic findings and frequency of extraocular manifestations in patients with HLA-B27 uveitis: a study of 175 cases. *Ophthalmology*. 2004;111:802–809.
2. Nurmohamed MT, van der Horst-Bruinsma I, Maksymowych WP. Cardiovascular and cerebrovascular diseases in ankylosing spondylitis: current insights. *Curr Rheumatol Rep*. 2012;14:415–421.
3. Bakker P, Moltó A, Etcheto A, et al. The performance of different classification criteria sets for spondyloarthritis in the worldwide ASAS-COMOSPA study. *Arthritis Res Ther*. 2017;19(1):96.
4. Stolwijk C, van Onna M, Boonen A, et al. The global prevalence of spondyloarthritis: a systematic review and meta-regression analysis. *Arthritis Care Res. (Hoboken)*. 2016;68:1320–1331.
5. Thom N, Ritchlin CT, Zhang X, et al. Prevalence of chronic axial pain, inflammatory back pain and spondyloarthritis in diagnosed psoriasis. *Arthritis Care Res. (Hoboken)*. 2015;67:829–835.
6. Loftus EV. Clinical epidemiology of inflammatory bowel disease: incidence, prevalence, and environmental influences. *Gastroenterology*. 2004;126:1504–1517.
7. Hugh JM, Weinberg JM. Update on the pathophysiology of psoriasis. *Cutis*. 2018;102(5S):6–12.
8. Ritchlin CT, Colbert RA, Gladman DD. Psoriatic arthritis. *N Engl J Med*. 2017;376(10):957–970.
9. Weiss PF, Colbert RA. Juvenile spondyloarthritis: a distinct form of juvenile arthritis. *Pediatr Clin North Am*. 2018;65(4):675–690.
10. Brown MA, Kennedy LG, MacGregor AJ, et al. Susceptibility to ankylosing spondylitis in twins: the role of genes, HLA, and the environment. *Arthritis Rheum*. 1997;40:1823–1828.
11. Robinson PC, Claushuis TA, Cortes A, et al. Genetic dissection of acute anterior uveitis reveals similarities and differences in associations observed with ankylosing spondylitis. *Arthritis Rheumatol*. 2015;67:140–151.
12. Reveille JD. The MHC and ankylosing spondylitis. *Clin Rheum*. 2014;33:749–757.
13. Hwang MC, Ridley L, Reveille JD. Ankylosing spondylitis risk factors: a systematic literature review. *Clin Rheumatol*. 2021 https://doi.org/10.1007/s10067-021-05679-7.
14. Reveille JD, Hirsch R, Dillon CF, et al. The prevalence of HLA-B27 in the United States: data from the U.S. National Health and Nutrition Examination Survey, 2009. *Arthritis Rheum*. 2012;64:1407–1411.
15. International Genetics of Ankylosing Spondylitis Consortium (IGAS), Cortes A, Hadler J, et al. Identification of multiple risk variants for ankylosing spondylitis through high-density genotyping of immune-related loci. *Nat Genet*. 2013;45:730–738.
16. Delay ML, Turner MJ, Klenk EI, et al. HLA-B27 misfolding and the unfolded protein response augment interleukin-23 production and are associated with Th17 activation in transgenic rats. *Arthritis Rheum*. 2009;60:2633–2643.
17. Reveille JD, Zhou X, Lee MJ, et al. HLA class I and II alleles in susceptibility to ankylosing spondylitis. *Ann Rheum Dis*. 2019;78(1):66–73.
18. Reveille JD. Inmunopatogenia. Predisposición Alélica de la Artritis Psoriásica. In Libro de Texto Artritis Psoriásica. Alba Feriz R, Muñoz-Louis R, Espinoza LR, Reveille JD Eds. 2019 1ERA Edición.
19. Li Z, Brown MA. Progress of genome-wide association studies of ankylosing spondylitis. *Clin Transl Immunology*. 2017;6(12):e163.
20. Genetic Analysis of Psoriasis Consortium and The Welcome Trust Case Control Consortium 2. A genomewide association study identifies new psoriasis susceptibility loci and an interaction between HLA-C and ERAP1. *Nat Genet*. 2010;42:985–990.
21. Ellinghaus D, Jostins L, Spain SL, et al. Analysis of five chronic inflammatory diseases identifies 27 new associations and highlights disease-specific patterns at shared loci. *Nat Genet*. 2016;48:510–518.
22. Cortes A, Maksymowych WP, Wordsworth BP, et al. Association study of genes related to bone formation and resorption and the extent of radiographic change in ankylosing spondylitis. *Ann Rheum Dis*. 2015;74:1387–1393.
23. Mielants H, De Keyser F, Baeten D, et al. Gut inflammation in the spondyloarthropathies. *Curr Rheumatol Rep*. 2005;7(3):188–194.
24. Costello ME, Robinson PC, Benham H, et al. The intestinal microbiome in human disease and how it relates to arthritis and spondyloarthritis. *Best Pract Res Clin Rheumatol*. 2015;29:202–212.
25. Melis L, Elewaut D. Progress in spondylarthritis. Immunopathogenesis of spondyloarthritis: which cells drive disease? *Arthritis Res Ther*. 2009;11:233.
26. van Kuijk AWR, Tak PP. Synovitis in psoriatic arthritis: immunohistochemistry, comparisons with rheumatoid arthritis, and effects of therapy. *Curr Rheumatol Rep*. 2011;13:353–359.
27. Rahman MA, Thomas R. The SKG model of spondyloarthritis. *Best Pract Res Clin Rheumatol*. 2017;31(6):895–909.
28. McGonagle D, Gibbon W, Emery P. Classification of inflammatory arthritis by enthesitis. *Lancet*. 1998;352:1137–1140.
29. Juanola X, Loza Santamaria E, Cordero-Coma M, et al. Description and prevalence of spondyloarthritis in patients with anterior uveitis: the SENTINEL Interdisciplinary Collaborative Project. *Ophthalmology*. 2016;123(8):1632–1636.
30. Danve A. Thoracic manifestations of ankylosing spondylitis, inflammatory bowel disease, and relapsing polychrondritis. *Clin Chest Med*. 2019;40(3):599–608.
31. Mercieca C, van der Horst-Bruinsma IE, Borg AA. Pulmonary, renal and neurological comorbidities in patients with ankylosing spondylitis; implications for clinical practice. *Curr Rheumatol Rep*. 2014;16(8):434.
32. Xiao M, Lv Q, Zhang Y, et al. Spondyloarthritis patients suffer increased risk of renal complications compared with general population: a retrospective observational study. *Front Pharmacol*. 2019;10:1073.
33. Ramirez J, Nieto-Gonzalez JC, Curbelo Rodriguez R, et al. Prevalence and risk factors for osteoporosis and fractures in axial spondyloarthritis: a systematic review and meta-analysis. *Semin Arthritis Rheum*. 2018;48(1):44–52.
34. Richards C, Hans D, Leslie WD. Trabecular Bone Score (TBS) predicts fracture in ankylosing spondylitis: the Manitoba BDM Registry. *J Clin Densitom*. 2020
35. Tang C, Moser FG, Reveille J, et al. Cauda equina syndrome in ankylosing spondylitis: challenges in diagnosis, management, and pathogenesis. *J Rheumatol*. 2019;46(12):1582–1588.
36. Nie A, Wang C, Song Y, et al. Prevalence and factors associated with disturbed sleep in outpatients with ankylosing spondylitis. *Clin Rheumatol*. 2018;37(8):2161–2168.
37. Hwang MC, Lee MJ, Gensler LS, et al. Longitudinal associations between depressive symptoms and clinical factors in ankylosing spondylitis patients: analysis from an observational cohort. *Rheumatol Int*. 2020;40(7):1053–1061.
38. Rusman T, van Vollenhoven RF, van der Horst-Bruinsma IE. Gender differences in axial spondyloarthritis: women are not so lucky. *Curr Rheumatol Rep*. 2018;20(6):35.
39. Lee W, Reveille JD, Davis JC, et al. Are there gender differences in the severity of ankylosing spondylitis: results from the PSOAS cohort. *Ann Rheum Dis*. 2007;66:633–638.
40. Tse SE, Laxer RM. New advances in juvenile spondyloarthritis. *Nat Rev Rheumatol*. 2012;8(5):269–279.
41. Peluso R, Di Minno MN, Iervolino S, et al. Enteropathic spondyloarthritis: from diagnosis to treatment. *Clin Dev Immunol*. 2013;2013:631408.
42. Vavricka SR, Schoepfer A, Scharl M, et al. Extraintestinal manifestations of inflammatory bowel disease. *Inflamm Bowel Dis*. 2015;21(8):1982–1992.
43. Troncoso LL, Biancardi AL, de Moraes Jr HV, et al. Ophthalmic manifestations in patients with inflammatory bowel disease: a review. *World J Gastroenterol*. 2017;23(32):5836–5848.
44. Zeidler H, Brandt J, Schnarr S. Undifferentiated spondyloarthritis. In: Weisman MH, Reveille JD, van der Heijde D, eds. *Ankylosing Spondylitis and the Spondyloarthropathies*. Philadelphia: Elsevier; 2006:75–93.
45. Maksymowych WP, Inman RD, Salonen D, et al. Spondyloarthritis Research Consortium of Canada magnetic resonance imaging index for assessment of sacroiliac joint inflammation in ankylosing spondylitis. *Arthritis Rheum*. 2005;53:703–709.
46. Ward MM, Deodhar A, Gensler LS, et al. 2019 Update of the American College of Rheumatology/Spondylitis Association of America/Spondyloarthritis Research and Treatment Network recommendations for the treatment of ankylosing spondylitis and nonradiographic axial spondyloarthritis. *Arthritis Rheumatol*. 2019;71(10):1599–1613.

47. Wassenberg S. Radiographic scoring methods in psoriatic arthritis. *Clin Exp Rheumatol*. 2015;33(5 suppl 93):S55–S59.
48. Lopez-Medina C, Molto A, Claudepierre P, et al. Clinical manifestations, disease activity and disease burden of radiographic versus non-radiographic axial spondyloarthritis over 5 years of follow-up in the DESIR cohort. *Ann Rheum Dis*. 2020;79(2):209–216.
49. Ward MM, Reveille JD, Learch TJ, et al. Impact of ankylosing spondylitis on work and family life: comparisons with the US population. *Arthritis Rheum*. 2008;59:497–503.
50. Szabo SM, Levy AR, Rao SR, et al. Increased risk of cardiovascular and cerebrovascular diseases in individuals with ankylosing spondylitis: a population-based study. *Arthritis Rheum*. 2011;63:3294–3304.
51. Hannu T. Reactive arthritis. *Best Pract Res Clin Rheumatol*. 2011;25(3):347–357.
52. Winthrop KL, Weinblatt ME, Bathon J, et al. Unmet need in rheumatology: reports from the Targeted Therapies Meeting 2019. *Ann Rheum Dis*. 2020;79(1):88–93.
53. Schmitt SK. Reactive arthritis. *Infect Dis Clin North Am*. 2017;31(2):265–277.
54. Haroon N, Inman RD, Learch TJ, et al. The impact of tumor necrosis factor α inhibitors on radiographic progression in ankylosing spondylitis. *Arthritis Rheum*. 2013;65:2645–2654.
55. Sweeney S, Gupta R, Taylor G, et al. Total hip arthroplasty in ankylosing spondylitis: outcome in 340 patients. *J Rheumatol*. 2001;28:1862–1866.
56. Carter S, Lories RJ. Osteoporosis: a paradox in ankylosing spondylitis. *Curr Osteoporos Rep*. 2011;9:112–115.
57. Lazennec JY, d-Astorg H, Rousseau MA. Cervical spine surgery in ankylosing spondylitis: Review and current concept. *Orthop Traumatol Surg Res*. 2015;101(4):507–513.

59

Small- and Medium-Vessel Primary Vasculitis

Raashid Ahmed Luqmani, Tamir Malley, Ana Águeda and Lorraine O'Neill

Small- and medium-vessel vasculitides are characterized by inflammation of the blood vessel wall, resulting in end organ failure or irreversible tissue damage and necrosis. In some cases, this is relatively trivial and may lead to minor inconvenience for the patient. However, in many forms of vasculitis, the consequences of rapid onset of ischemia and occlusion of blood vessels are devastating, leading to organ failure and death.

The distinction between small- and medium-vessel vasculitis entities is arbitrary, with three main patterns of vasculitis: small, medium, and large. Although there is merit in this classification, there are different patterns of involvement for patients with predominantly small-vessel involvement (characterized by capillaritis) compared with—but overlapping with—patients with medium-vessel involvement typified by small arteriolar inflammation. For example, in the kidney, small-vessel involvement leads to inflammation of glomeruli (glomerulonephritis); by contrast, medium-vessel inflammation of renal arterioles results in infarction of the kidney with tissue loss. Therefore, patterns of disease are well recognized, with typical dominance of kidney, lung, and upper airway involvement in patients with granulomatosis with polyangiitis (GPA), previously termed Wegener's granulomatosis. In microscopic polyangiitis (MPA), there is lack of upper airway involvement but significant renal and lung involvement. Finally, in eosinophilic glomerulonephritis with polyangiitis (EGPA), previously termed Churg-Strauss syndrome, the pattern of clinical involvement is dominated by upper and lower airway disease combined with neurological (peripheral nerve) features. These three entities are grouped together by their association with the presence of antineutrophil cytoplasmic antibody (ANCA) in most, but not all, cases.

In patients with polyarteritis nodosa (PAN), one of the main forms of medium-vessel vasculitis, the most characteristic findings are bowel ischemia or infarction and peripheral neuropathy. In a childhood onset form of medium-vessel vasculitis (Kawasaki disease [KD]), the clinical features are diverse and include mucocutaneous inflammation and systemic upset with fever, and in 2% to 4% of cases, there is coronary artery dilatation and/or aneurysm development, which can potentially rupture, leading to fatal consequences.[1]

The diversity of different forms of vasculitis with overlapping features suggests that the underlying mechanisms are varied, but some pathways are likely to be shared. This is reflected in treatment approaches, which—with some exceptions—are often very similar across diseases. Disease relapses can occur in up to 55% of patients within the first 3 years of achieving remission with on-going persistent risk of relapse. In addition, within the first 6 months of remission-induction therapy, many patients fail to achieve remission due to persistent or recurrent active ANCA-associated vasculitides (AAV).[2] This can lead to repeated high doses of immunosuppressive therapy, thus increasing the risk of adverse effects.[3]

EPIDEMIOLOGY

Despite improvement in our understanding of the epidemiology of the systemic vasculitides, patients are often not diagnosed as having vasculitis for extended periods. The discovery of ANCA and their association with small-vessel vasculitis has improved recognition.[4] Greater awareness of vasculitis may be a factor that explains an apparent increase in incidence from 1.5 to 6.1/million per year.[5] AAV have an incidence of 20 per million new cases per annum (GPA nine per million, MPA 9 per million, EGPA 2 per million), a prevalence of 200 per million and an average age of onset of 60 to 70 years.[6] In southern Europe, the number of cases of MPA is greater than those of GPA. MPA is also more common than GPA in Japan. ANCA directed against myeloperoxidase (MPO) is the predominant antibody detected in patients with AAV in Japan, whereas autoantibodies directed against proteinase 3 (PR3) are rarely seen in Japanese patients but are the most frequent ANCA antibody in patients in northern Europe.

The epidemiology of EPGA is less well understood compared with other forms of AAV. EGPA is characterized by an elevated eosinophil count in patients with late-onset asthma. Around 50% of cases have ANCA, usually directed against MPO. There is a potential overlap condition called *hypereosinophilic syndrome* (HES), and it is not clear whether some cases of HES are really cases of EGPA, or vice versa. Indeed, if patients are ANCA-negative, distinguishing the two conditions can be challenging.[7] Furthermore, because bronchospasm is a key feature of EGPA, it is possible that some cases of asthma, an extremely common condition, may, in fact, represent mild forms of EGPA. This has been brought more to light in cases of drug-induced EGPA, specifically in the setting of the use of montelukast, a leukotriene inhibitor, as a treatment for moderate to severe asthma. It has been suggested that these patients probably had underlying EGPA, which had previously been suppressed with systemic glucocorticoids, but when these were withdrawn, features of EGPA became more apparent.

AAV typically affects older individuals in their 60s or 70s, but can occur at any age. Most patients survive their initial illness as a result of effective immunotherapy, and therefore the prevalence of these diseases is growing. In southern Sweden, estimates of prevalence are 160 per million (95% confidence interval [CI] 114 to 206) for GPA, 94 (58 to 129) for MPA, 31 (11 to 52) for PAN, and 14 (0.3 to 27) for EGPA.[8] However, these are based on a relatively small population size and are higher than those reported from a Spanish series, where the prevalence of all forms of AAV was under 45 per million.

The two main forms of medium-vessel vasculitis are PAN and KD. PAN is extremely rare. There is very wide misconception among many physicians that PAN is the most common form of vasculitis, and this is partly encouraged by the older literature, which refers to all forms of vasculitis as PAN (initially called *periarteritis nodosa* and subsequently *polyarteritis nodosa*). In fact, most patients with so-called PAN probably did not have this disease but were more likely suffering from one of the forms of small-vessel vasculitis, particularly MPA and GPA. PAN has been associated with infection, especially hepatitis B and hepatitis C. The epidemiology of hepatitis B has been transformed by effective immunization; as a result of this, hepatitis B associated with PAN is now a rare disease. Recent figures suggest an incidence of PAN of 0.6 to 3.6 per million adults.[9]

KD is most common in children under the age of 5 years, but can occur in older children and young adults, in which case it is more difficult to diagnose because it is not suspected. In a recent Italian study of children under the age of 14 years, the incidence rate was 17.6/100,000 children under the age of 5 years, with a slight increase in reported cases during spring and winter compared with other times of the year. This is a slightly higher incidence rate than previously reported by other studies in Europe with a range of 3.6/100,000 to 15.2/100,000 children under the age of 5 years. Earlier figures from a large-scale study from the United States of over 6000 children admitted to the hospital because of KD suggested that the peak age of onset was 1 year, and children under the age of 2 years accounted for over a third of all cases. The under-5-year incidence was reported as 8.1/100,000 children in 1988, rising to 18.5/100,000 children in 1997, similar to the recent Italian experience. Males were more commonly affected than females (~60% males); there was no obvious seasonal variation. In a recent nationwide hospital survey in Japan, however, the incidence rates for KD during 2011 and 2012 were over 2400 cases per million children under the age of 5 years. The higher incidence of KD in patients of Japanese ethnicity appears to be independent of their geographical location. In fact, during 1996 to 2006, the average annual incidence of KD in Japanese American children in Hawaii was 210.5/100,000 children, in contrast to the rate for Caucasian children in Hawaii of 13.7/100,000, similar to the rate of 12.0/100,000 among Caucasian children in the continental United States.[10]

Leukocytoclastic skin vasculitis is one of the more common forms of small vessel vasculitis occurring with an annual incidence of about 45 per million (Table 59.1).[11] Less than a third is found in association with immunoglobulin A (IgA; Henoch-Schönlein purpura [HSP], or IgA vasculitis). IgA vasculitis is very common in children and is usually self-limiting; annual incidence is 100 to 200/million children ≤17 years of age or younger.[12] By contrast, this is a much less common disease in adults (around 13/million per year).

Cryoglobulinemic vasculitis is strongly associated with hepatitis C, and its epidemiology is likely to mirror the prevalence of the hepatitis C virus (HCV) infection. However, there are no published incidence and prevalence figures for cryoglobulinemic vasculitis itself.

TABLE 59.1 Diagnosis, Incidence and Prevalence in KD, PAN, AAV and Leukocytoclastic Vasculitis

Diagnosis	Incidence	Prevalence	References/ Reviews
KD	For children <5 years old: 2431/million (Japan) 1131/million (Korea) 690/million (Taiwan) 36–185/million (Italy)	Not applicable	Suka et al.[96]
PAN	3.6/million adults	2.6–14/million adults	Nesher et al.[5] Mohammad et al.[4]
AAV	9.5–16/million/year (Germany) 22.6/million (Japan) 21.8/million (UK)	149/million	Reinhold-Keller et al.[10] Herlyn et al.[11] Fujimoto et al.[94]
Leukocytoclastic vasculitis	45/million (equal male and female; increased incidence with age)	No data	Arora et al.[7]

AAV, ANCA-associated vasculitis; *ANCA*, antineutrophil cytoplasmic antibody; *KD*, Kawasaki disease; *PAN*, polyarteritis nodosa.

KEY CONCEPTS

Pathogenic Mechanisms in Antineutrophil Cytoplasmic Antibody-Associated Vasculitides

- The recognition of an association between some forms of vasculitis and the presence of ANCA has transformed understanding of the group of diseases considered to be associated with ANCA and in which ANCAs are suspected to have a pathogenic role.
  - GPA.
  - MPA.
  - EGPA.
- A genome-wide association study (GWAS) has demonstrated strong associations with specific alleles.
  - Anti-proteinase 3 ANCA with specific HLA-DP alleles and alleles encoded by genes for α_1-antitrypsin and proteinase 3.
  - Anti-myeloperoxidase ANCA with specific HLA-DQ alleles.
- Environmental factors.
  - Silica exposure.
  - Specific strains of *Staphylococcus aureus*.
- Immune dysregulation.
  - Defective regulation of T-cell immunity.
  - Neutrophil generation of extracellular traps (NETs) containing proteinase 3 and myeloperoxidase.
  - Activation of alternative complement pathway by ANCA-activated neutrophils.

PATHOGENESIS OF ANTINEUTROPHIL CYTOPLASMIC ANTIBODY-ASSOCIATED VASCULITIDES

The Pathogenic Role of ANCA in GPA and MPA

There is some evidence that ANCA play a role in the pathogenesis of GPA, MPA, and EGPA. Based on the immunofluorescence pattern, different forms of ANCA can be distinguished, but only

two are of direct clinical relevance: cytoplasmic c-ANCA (corresponding to antibodies directed against PR3 and perinuclear p-ANCA (predominantly corresponding to antibodies directed against MPO). p-ANCA can also be directed against other antigens, including bactericidal/permeability-increasing (BPI) protein, lactoferrin (LF), human neutrophil elastase (HNE), cathepsin G, and azurocidin, but their clinical significance is not well characterized. Although ANCAs are associated with vasculitis, titers are not reliable for monitoring disease status because there is no clear relationship with remission or relapse.

Transfer of MPO-ANCA in humans (maternal–fetal route) and animal models (necrotizing pauciimmune glomerulonephritis after passive transfer of purified antibody or splenocytes from MPO-deficient mice immunized with purified murine MPO) has resulted in features typical of MPA.[13] By contrast, the pathogenicity of antiPR3 antibodies is less well established. In one animal model of autoimmune-prone nonobese diabetic (NOD) mice, immunization with recombinant mouse PR3 (rmPR3) in complete Freund's adjuvant (CFA) had no clinical effect but resulted in high levels of circulating c-ANCA; transfer of splenocytes from these immunized animals into mice with severe combined immunodeficiency (SCID) resulted in vasculitis and severe segmental and necrotizing glomerulonephritis.[14] Transfer of splenocytes from the CFA-alone-immunized mice (controls) resulted in no disease, further suggesting that disease development depends on PR3-specific immune responses. In a second model of PR3-ANCA vasculitis, based on animals with a human–mouse chimeric immune system,[15] more than 70% of mice treated with IgG from patients with antiPR3 AAV (as compared with IgG from patients with non-vasculitic renal disease; or healthy controls) developed mild kidney disease with glomerular hypercellularity and focal pulmonary hemorrhage. Fifteen (83%) mice treated with antiPR3 IgG later showed mild kidney disease with glomerular hypercellularity, and 3 (17%) had severe glomerular injury. In the lungs, 13 (72%) showed areas of focal pulmonary hemorrhage, whereas lungs of the control group ($n = 8$) appeared normal ($P < .01$). There were no granulomatous lesions, but because granulomatous lesions are dependent on a robust T-cell–mediated response, the authors argued that a refinement of the model to include greater levels of chimerism and administration of interleukin-7 (IL-7)–Fc protein to augment T-cell development would be required to study this.[15]

Even though these studies strongly support a pathogenic role for ANCA, conventional serological assays fail to detect ANCA in some patients with classic clinical and pathological features of AAV, and titers do not correlate well with disease activity. Roth et al. examined MPO epitopes specificities,[16] reporting 25 different epitopes bound by anti-MPO antibodies; although some epitopes were associated with active disease, others were either not specific to active disease or not associated with disease at all. Igs purified from patients with ANCA-negative vasculitis could bind to a specific MPO epitope. Furthermore, the absence of ANCA in some patients could be explained by competitive binding to a fragment of ceruloplasmin (CP), the natural inhibitor of MPO. This CP fragment decreased anti-$MPO_{447–459}$ autoantibody reactivity by 30% to 50%, whereas full-length CP did not have any effect.[16] ANCAs are reported in small numbers of healthy individuals. How the pathogenic transformation of ANCA occurs is still unclear, but it is probably a multifactorial process requiring a complex interaction among genetics, the environment, and the immune system facilitating a break in immune tolerance.

GENETICS

There has been significant progress in the understanding of AAV genetics following the publication of two GWAS.[17] The strongest human leukocyte antigen (HLA) association is with the HLA-DPB1 haplotype, especially for the PR3-ANCA–positive subgroup, regardless of the clinical diagnosis.[18] Further analysis has revealed an association between MPO-ANCA and a single nucleotide polymorphism (SNP) in the HLA-DQ region, which had probably been masked previously as a result of the small number of MPO-ANCA–positive patients included in initial analyses. Other HLA associations are reported, such as HLA-DRB1*09:01 and HLA-DQB1*03:03 in Japanese patients with MPA. Less robust findings, not replicated in other studies, include a protective effect of HLA-DR13(6) and HLA-DR1, but an increased proportion of HLA-DR4 in Dutch patients with GPA compared with controls[18]; HLA-DRB1 in PR3-ANCA-positive (but not MPO-ANCA) patients[18]; HLA-B50, HLA-DR1, HLA-DR9, HLA-DQw7, and HLA-DR3 in GPA.[18] In EGPA, the most robust association is with HLA-DRB4. Overall, there is evidence for genetic susceptibility to AAV, related to specific SNPs in the HLA region. Other genetic associations with GPA include *PRTN3, SERPINA 1, PTPN22*, and *CTLA4*.

PR3 is either stored in neutrophil azurophilic granules or exposed on the cell membrane (where it can interact directly with ANCA). Although the proportion of neutrophils displaying membrane PR3 ($mPR3^+$) is stable over time, surface expression of PR3 may be enhanced. The percentage of $mPR3^+$ neutrophils is genetically determined. Schreiber et al. showed that among 125 healthy controls, 35 patients with GPA, 15 patients with other inflammatory diseases, and 27 pairs of monozygotic (MZ) and dizygotic (DZ) twins, the percentage of $mPR3^+$ neutrophils correlated significantly in MZ twins (but not in DZ twins) and the heritability percentage was estimated as 99%. Furthermore, the absolute number of PR3 molecules expressed on the cell membrane was correlated among MZ (but not DZ) twins, with a heritability estimate of 96.7%.[18] Following neutrophil activation and enzyme release, PR3 can mediate direct tissue damage. α_1-antitrypsin, encoded by the gene *SERPINA1* (found on chromosome 14), is the major inhibitor. Two α_1-antitrypsin alleles, Z and S, are associated with low enzymatic activity. A significant correlation with the Z allele (odds ratio [OR] 0.3; $P = 1.25 \times 10^{-5}$), but not the S allele, is reported in PR3-ANCA–positive patients with GPA.[18] *PRTN3* (the gene encoding PR3) is associated with GPA, especially among patients who are PR3-ANCA-positive.[18]

The 620WPTPN22 variant of *PTPN22* (encoding lymphoid tyrosine phosphatase) correlates with abnormal regulatory CD4 T regulatory cell (Treg) function, increased humoral activity, and enhanced neutrophil function in patients with GPA.[19] CTLA-4 (encoding cytotoxic T lymphocyte antigen-4) polymorphisms are associated with GPA.[19] CTLA-4 is a negative regulator of T-cell activation, which competes with the costimulatory molecule CD28 for binding of CD80 or CD86 on antigen-presenting cells (APCs). Abatacept, a monoclonal antibody (mAb), containing the binding domain of CTLA-4, reduces CD28–CD80/CD86 interaction and, therefore, T-cell stimulation—which could explain its potential benefit in a small trial of non–life-threatening GPA.[20] Haplotypes of IL-10 (a pleiotropic cytokine with complex and multiple effects in immune modulation) are associated with several immunological disorders, including systemic lupus erythematosus (SLE), rheumatoid arthritis (RA), and giant cell

arteritis. IL-10 production is largely (50% to 70%) determined by genetic factors, and elevated plasma levels are reported in EGPA but not in GPA.[21] Wieczorek et al. evaluated the impact of functionally relevant IL-10 polymorphisms in 403 patients with GPA, 103 with EGPA (compared with 507 controls).[21] There was a significant association with the 3575/1082/592 TAC haplotype—part of the extended ancient haplotype of IL-10.2—in ANCA-negative EGPA but not in GPA.[21]

EPIGENETICS

Genetic factors alone are not enough to explain the range of phenotypic presentations in AAV. Epigenetic dysregulation is increasingly recognized as a contributor to immune-mediated diseases. Epigenetics relate to heritable changes in gene function, without altering the DNA sequence. Most prominent epigenetic changes are DNA methylation, histone alterations, and microRNAs (miRNAs). Epigenetic modifications can be stable over time or can respond to developmental and environmental triggers, leading to phenotypic aberrances. Central to the pathogenesis of AAV is a dysregulated immune response resulting in ANCA production and aberrant expression of their target autoantigens, MPO and PR3. The expression of MPO and PR3 occurs primarily during early neutrophil development to produce intragranular constituents and is silenced in mature cells. However, in AAV, expression remains active. One epigenetic change that helps explain this is the reduced levels of the H3K27me3 histone modification at both the MPO and PR3 gene loci. H3K27me3 histone is associated with transcriptional silencing in patients with AAV compared with healthy individuals.[22]

ENVIRONMENTAL AND INFECTIOUS TRIGGERS

Genetic and epigenetic modifications may render a patient susceptible to developing disease in the presence of an appropriate trigger. Several triggers have been associated with AAV, including toxins, viral and bacterial infections, and some licit and illicit drugs.

Silica dust exposure has been associated with the development of AAV and other autoimmune diseases, such as SLE, RA, and scleroderma, as demonstrated in a recent large epidemiological survey of almost three million individuals.[23] Silica is an abundant earth material found in sand, grain, grass, and wool. Processing these materials may expose workers to respirable crystalline silica. Silica-induced ANCA-positive disease is often associated with p-ANCA, targeting MPO, and the clinical presentation is usually MPA rather than GPA.[23] The mechanism by which silica exposure triggers the development of AAV is not fully understood. *In vitro*, silica can activate monocytes and macrophages, releasing cytokines, such as IL-1 and tumor necrosis factor (TNF), as well as releasing oxygen radicals and lysosomal enzymes, including PR3 and myeloperoxidase. Furthermore, silica can inactivate α_1-antitrypsin. Asbestos, another silicon-containing mineral, has been suggested as a trigger for AAV in a small case-control study of 31 patients (22 GPA, 8 MPA, 1 EGPA) and 30 healthy controls; three patients had prior exposure to asbestos versus none of the controls.[24] There are few published studies investigating the potential pathogenic role of asbestos in the development of AAV.

It is estimated that 63% of patients with GPA are chronic nasal carriers of *Staphylococcus aureus*, resulting in an increased risk of relapse. Maintenance treatment with trimethoprim-sulfamethoxazole (cotrimoxazole) has been shown to reduce the incidence of relapse in a double-blind, placebo-controlled study in patients with GPA, where a maintenance dose of 960 mg two times daily of trimethoprim-sulfamethoxazole resulted in a 60% reduction in relapse. However, this has not been replicated in other studies. The mechanism for the pathogenic role of *S. aureus* is still unclear. Potential pathways include stimulation of B and/or T cells by *S. aureus* superantigens (SAgs); polyclonal activation of B cells by cell wall components of the bacterium, resulting in persistence of ANCA; and neutrophil priming leading to surface expression of PR3.[25]

Although parvovirus B19 has been proposed as a trigger for AAV in a few case reports, a case-control study failed to demonstrate any association because IgG antibodies to parvovirus B19 were detected equally in the sera of patients with AAV and control subjects, and all 13 patients with AAV and 39 controls were negative for IgM antibodies and viral DNA.[26] Hepatitis B virus (HBV) is implicated in the pathogenesis of PAN,[27] but there is no evidence to support a role for HBV or HCV in AAV.

Lysosomal-associated membrane protein 2 (LAMP-2) is a heavily glycosylated type 1 membrane protein, abundant on neutrophil and endothelial cell surfaces, which shuttles between lysosomes and the cell membrane. LAMP-2 cross-reacts with FimH, a gram-negative adhesin, which facilitates bacterial entry to host tissues. Preliminary studies have suggested a role for LAMP-2 in small-vessel vasculitis. The LAMP-2 epitope P41 to 49 has been reported to have a 100% homology with amino acids 72 to 80 of mature FimH.[28] Antibodies to human LAMP-2 have been shown to injure human microvascular endothelium *in vitro* and induce focal necrotizing glomerulonephritis (FNGN) in rats; immunization with FimH-induced pauciimmune FNGN associated with antibodies that bound human and rat LAMP-2; furthermore, patients with pauciimmune FNGN were found to have an increased likelihood of infections with FimH-expressing bacteria shortly before presentation of their FNGN.[28] LAMP-2 antibodies were found in patients with active disease or relapse but not in those in remission.[28] More recently, LAMP2-ANCA have been reported in 35% of children with small vessel vasculitis but not associated with disease severity.[29] Although this a promising new explanation for some cases of small-vessel vasculitis, the lack of replication of these findings makes them hard to substantiate.

CpG-oligodeoxynucleotides (ODN) is a short synthetic DNA containing unmethylated CpG motifs, highly prevalent in bacterial DNA, and recognized by Toll-like receptor 9 (TLR9), which is expressed by a variety of cells in the immune system. TLR9 triggering results in proinflammatory IL production. It has been reported that CpG motifs and IL-2 exposure can induce B lymphocytes to proliferate, increase antigen presentation, produce a range of cytokines, and differentiate into Ig-producing cells, ultimately leading to ANCA production.[30]

DRUG INDUCED ANTINEUTROPHIL CYTOPLASMIC ANTIBODY-ASSOCIATED VASCULITIDES

The most commonly implicated drugs that can induce an AAV-like syndrome are propylthiouracil (PTU), hydralazine, levamisole-adulterated cocaine, TNF inhibitors, sulfasalazine, D-penicillamine, and minocycline (Fig. 59.1).

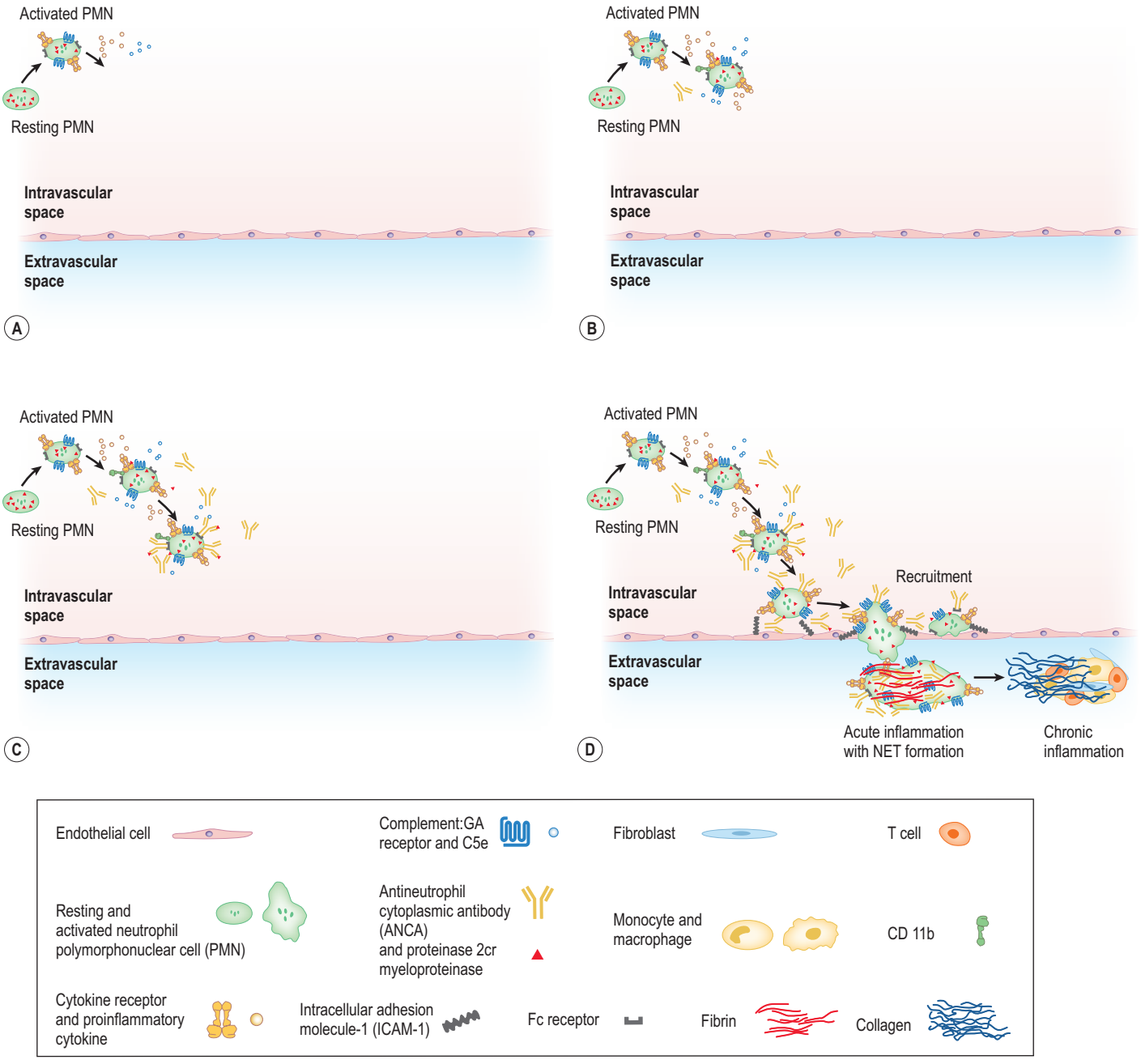


FIG. 59.1 Proposed pathogenic mechanisms in antineutrophil cytoplasmic antibody (ANCA) associated vasculitis. The interaction between activated neutrophils and ANCA is depicted in the figure, with resting neutrophils becoming activated (A), interacting with ANCA in the intravascular space (B), triggering a series of cascading events leading to vascular invasion (C) by inflammatory cells and the formation of neutrophil extracellular traps and eventually chronic inflammation in the extravascular space (D). *NET*, Neutrophil generation of extracellular trap.

Propylthiouracil

The most commonly described drug associated with the presence of ANCA is PTU. PTU is used to treat hyperthyroidism and was the first drug described to induce a condition mimicking AAV in the 1990s. Even though its use has declined over time, reports of PTU-induced AAV still emerge. Clinically, PTU-induced AAV may manifest as acute kidney injury caused by pauci-immune necrotizing and crescentic glomerulonephritis in addition to respiratory tract, joint, and skin diseases.

Reports of the proportion of patients who develop ANCA as a result of exposure to antithyroid medications varies widely from 4% to 46%; by contrast, the prevalence of antithyroid drug-induced AAV is much lower (0% to 1.4%). Although PTU is the most commonly reported antithyroid drug, methimazole and carbimazole have also been implicated. In a study by Noh et al,[31] the estimated annual incidence of antithyroid drug-induced vasculitis was 0.53 to 0.79 patients per 10,000 patients with Graves disease, especially for patients receiving PTU; the

ratio of the estimated incidence for methimazole and PTU was 1:39. Most patients with PTU-induced ANCA will not develop clinical features of vasculitis. Patients with PTU-induced AAV (usually MPO-positive) are younger, more commonly female compared with those with primary AAV, and usually have significantly higher anti-MPO ANCA titers compared with patients without vasculitis. Other ANCA subtypes may be found, including HNE-ANCA and LF-ANCA; however, when PTU-induced AAV develops, MPO-ANCA is usually present, regardless of the existence of other forms of ANCA. The pathogenesis of AAV related to antithyroid drugs is unclear. PTU is predominantly metabolized in the liver, but a proportion is modified by MPO in neutrophils.[32] Neutrophils are responsible for NET formation, and PTU can induce abnormal configured NETs *in vitro;* these abnormal NETs cannot easily be released into liquid phase and are not effectively digested by DNase I, therefore, remaining in the tissue.[33] Immunization of rats with abnormal NETs results in production of MPO-ANCA.[33]

Hydralazine

Hydralazine is widely used to treat hypertension, acting as a smooth muscle relaxant and causing arterial and arteriolar vasodilation. It has been implicated in the development of drug-induced SLE and AAV. High titers of MPO-ANCA, ANA, anti-double-stranded DNA (dsDNA), and anti-histone antibodies are found; hypocomplementemia is also frequent. The underlying mechanism of action for hydralazine-induced AAV is not clear. One potential pathway by which hydralazine acts is through reverse epigenetic silencing of tumor suppressors but also potentially of MPO and PR3.

Levamisole-Contaminated Cocaine

It is commonplace for illicit drugs, such as cocaine, to be combined with adulterants, such as levamisole, in an effort to increase profits. Levamisole is similar in physical appearance and has possible potentiating effects on levels of dopamine in the central nervous system. It is estimated that more than 75% of cocaine users in the United States are exposed to levamisole.

Clinical manifestations of AAV induced by levamisole-contaminated cocaine include constitutional features; arthralgia; retiform purpura involving the ears, face, and extremities; and, less commonly, renal and lung diseases. Laboratory abnormalities include leukopenia, neutropenia, and high-titer p-ANCA, directed against MPO-ANCA or against atypical p-ANCA–associated antigens, such as HNE, LF, and cathepsin G. PR3-ANCA, ANA, and antiphospholipid autoantibodies have also been described in these patients. This multiplicity of antibodies helps distinguish AAV induced by levamisole-contaminated cocaine from primary AAV, which usually targets just one antigen.

Levamisole has an estimated mean half-life of 5.6 hours; therefore, serum testing is likely to be negative if the most recent exposure occurred over 24 hours prior to sample collection. Urinary detection of levamisole is highly suggestive of drug-induced disease and is useful if exposure occurred less than 48 hours prior to testing.

The mechanism by which levamisole-contaminated cocaine induces AAV is unclear. Levamisole, like PTU, could serve as a substrate for MPO to form reactive metabolites that might induce autoimmunity. Levamisole can enhance NETosis through engagement of muscarinic (subtype 3) receptors; furthermore, NETs generated by levamisole can induce endothelial cell death and vascular dysfunction by disrupting normal endothelium-dependent vasorelaxation. Abusing cocaine without levamisole can also trigger a syndrome mimicking AAV. The clinical presentation can be identical to that of patients with GPA, although cerebral angiitis, urticarial vasculitis, and EGPA-like syndromes have also been described. In GPA-like vasculitis, patients typically present with cutaneous vasculitis, nasal septal destruction, and pauci-immune crescentic glomerulonephritis. The autoantibody profile usually includes c-ANCA with PR3 specificity, although there are reports of cases with negative ANCA or p-ANCA with PR3 specificity. In these cases, the target of p-ANCA may be atypical, such as HNE.[34]

Cocaine-induced midline destructive lesions (CIMDLs) in the upper respiratory tract mimic limited forms of GPA.[35] Patients with CIMDLs may present with or without ANCA, and when ANCA-positive, the pattern often varies. Some patients are positive for c-ANCA with PR3 specificity, but more often, patients with CIMDLs present p-ANCA with specificity for atypical ANCA, such as HNE. Wiesner et al. reported that 76% of the patients with CIMDLs were ANCA-positive (mostly p-ANCA): 57% had PR3-ANCA and 86% had HNE-ANCA; this compares with the absence of HNE-ANCA in GPA and MPA, suggesting that the presence of HNE-ANCAs may be helpful in distinguishing CIMDL from GPA.[36]

Table 59.2 provides a summary of the most significant associations between drugs and ANCA or AAV. The management of all forms of drug-induced AAV is withdrawal of the offending agent, supportive measures, and, in severe cases, immunosuppression, dialysis, and plasma exchange.

LOSS OF B- AND T-CELL TOLERANCE IN ANTINEUTROPHIL CYTOPLASMIC ANTIBODY-ASSOCIATED VASCULITIS

The imbalance of effector and Tregs underpins the autoimmune dysregulation seen in AAV with multiple alterations in the circulating T-cell population. Patients with AVV have a reduced number of circulating Tregs,[37] which are functionally impaired, and an expanded population of CD4 effector memory T cells with an increased number of activated T cells.[38] Persistence of CD4 T-cell activation in peripheral blood correlates with disease severity.[39]Aberrant T helper (Th) polarization, with an increase in proinflammatory Th17 responses, further contributes to vascular injury.[39] It has been suggested that these alterations to the peripheral T-cell compartment could be influenced by environmental factors, such as infection.

The discovery of B cells in the inflammatory lesions in AAV together with the success of B-cell depletion therapies suggests that B cells play a significant role in the pathogenesis of AAV; however, the exact mechanism of this involvement is still not known. Regulatory B cells are reduced in AAV, whereas B-lymphocyte stimulator (BLyS) levels are significantly increased, thereby promoting B-cell differentiation, proliferation, and survival.[40] B cells may play a number of roles in AAV: as precursors to antibody-producing plasma cells, as APCs, producing proinflammatory mediators, or in costimulation of T cells.

ROLE OF NEUTROPHILS

In addition to containing the antigens for ANCA, activated neutrophils release many mediators that modulate the inflammatory response and can directly contribute to tissue inflammation,

TABLE 59.2 Drug-Induced Vasculitis Associated With Antineutrophil Cytoplasmic Antibody Positivity—Implicated Drugs, Proposed Mechanisms of Action and Laboratory Findings

Drug/Class	Proposed Mechanism of Action	IF Pattern	ANCA Serotype	Other ANCA Autoantigens	Other Antibodies	Ref.
Allopurinol	Limited data	p-ANCA	MPO-ANCA	—	ANA	Jia et al.[50]
Anti–TNF (ADA, ETN, IFX)	TNF may induce the formation of immune complexes, activation of complement and mediate inflammation by switching from a cytokine response of T-helper type 1 to type 2, upregulating antibody production	p-ANCA c-ANCA	MPO-ANCA PR3-ANCA	—	ANA	Rowley et al.[51]
Benzylthiouracil	Limited data	p-ANCA	MPO-ANCA	HNE LF	—	Leung et al.[52]
Carbimazole	Limited data	p-ANCA c-ANCA	MPO-ANCA PR3-ANCA	—	—	
Cocaine	Enhanced formation of NETs enriched in neutrophil elastase and inflammatory mitochondrial DNA with enhanced release of B cell–activating factor belonging to the TNF family (BAFF)	c-ANCA p-ANCA	PR3-ANCA	HNE	—	Holden et al.[40] and Nakazawa et al.[41]
D-penicillamine	Limited data	p-ANCA	MPO-ANCA	HNE LF	ANA Anti-dsDNA abs	Schulte et al.[53]
Hydralazine	Hydralazine mediated reversal of epigenetic silencing of MPO and PR3, with subsequent increased expression of both autoantigens. Induction of cytotoxic products and neutrophil apoptosis mediated by hydralazine binding MPO	P-ANCA	MPO-ANCA PR3-ANCA	HNE LF	ANA Anti-dsDNA abs Antihistone abs Anticardiolipin abs	
Levamisole contaminated cocaine	Enhanced NETosis through engagement of muscarinic subtype 3 receptor Metabolism of levamisole by MPO with formation of reactive metabolites Genetic susceptibility to agranulocytosis in HLA B27+ patients	p-ANCA c-ANCA	MPO-ANCA PR3-ANCA (double + is very common)	HNE Cathepsin G LF	ANA Anticardiolipin abs	Holden [40] and Cid et al.[54]
Methimazole	Limited data	p-ANCA c-ANCA Atypical ANCA	MPO-ANCA PR3-ANCA	HNE	ANA	Yalcinkaya et al.[55] and Sansonno et al.[56]
Minocycline	Minocycline oxidation by MPO, with abnormal production of reactive metabolites and modification of enzymatic activity, triggering the induction of ANCA Cytotoxicity leading to premature apoptosis of neutrophils, with abnormal release of nucleosomes and drug-modified proteins (including myeloperoxidase, elastase, LL37, and HMGB1), which can be bound to NETs triggering lupus/vasculitis via type I IFN production	p-ANCA Atypical ANCA	MPO-ANCA PR3-ANCA	HNE Cathepsin G BPI	ANA Anti-dsDNA abs	Varricchi et al.[57]
PTU	Cross-reactivity between anti-thyroperoxidase (anti-TPO) antibody and MPO. Alteration of NET configuration PTU-induced structural change in MPO, which serves as a neoantigen Metabolism of PTU by MPO (in the presence of H_2O_2 and Cl^-) into strong toxic metabolites. Subsequent binding to neutrophils proteins and recognition by T cells with B cells activation, resulting in autoantibody production Competitive inhibition of MPO oxidation activity by PTU, in a dose dependent manner	p-ANCA atypical ANCA	MPO-ANCA PR3- ANCA	HNE LF BPI Azurocidin Cathepsin G	ANA Antihistone abs Anticardiolipin abs	Wang and Tsai[27], Yalcinkaya et al.[55], Craven et al.[58], and Walsh et al.[95]
Sulfasalazine	Sulfasalazine-induced neutrophil apoptosis, with subsequent membrane expression of PR3 and MPO	p-ANCA c-ANCA	MPO-ANCA PR3-ANCA	LF	ANA Anti-dsDNA abs	Jia et al.[50] and Luqmani and Ponte[59]

abs, Antibodies; *ADA*, Adalimumab; *ANA*, antinuclear antibodies; *ANCA*, antineutrophil cytoplasmic antibody; *BAFF*, B cell activating factor; *BPI*, bactericidal permeability increasing protein; *dsDNA*, double-stranded DNA; *ETN*, etanercept; *HLA*, human leukocyte antigen; *HNE*, human neutrophil elastase; *IF*, immunofluorescence; *IFN*, interferon; *IFX*, infliximab; *LF*, lactoferrin; *MPO*, myeloperoxidase; *NET*, neutrophil extracellular trap; *PTU*, **propylthiouracil;** *PR3*, proteinase 3; *Ref*, references; *TNF*, tumor necrosis factor.

vascular injury, and damage in AAV via phagocytosis, degranulation, and cytokine production. Neutrophils release B cell–activating factors (BAFFs) that enhance B-cell proliferation and retard apoptosis. Neutrophils from patients with AAV are more prone to spontaneous release of NETs.[41] In the normal immune system, NETs consist of chromatin fibers released from dying neutrophils and are designed to trap and kill extracellular pathogens. NETs not only contain proinflammatory proteins that directly cause endothelial cell damage and complement activation, but also form a link between innate and adaptive immunity via providing access to MPO and PR3.

ROLE OF COMPLEMENT

Despite the apparent paucity of immune complexes in AAV, complement (and in particular, activation of the alternative complement pathway) plays a crucial role in the pathogenesis of AAV. When primed neutrophils are activated by ANCA, they produce complement pathway (C5a), which, in addition to recruitment, primes additional neutrophils for further activation by ANCA.[42] C3a, C5a, and soluble C5b-9 levels are elevated in active disease; plasma levels of complement factor H, a regulator of the alternative complement pathway, are significantly lower in patients with active AAV.[43] Low serum levels of C3 at diagnosis are associated with a worse prognosis.[44] There is also a suggestion that C5a may play a role in the hypercoagulability associated with active AAV.

PATHOGENESIS OF KAWASAKI DISEASE

The incidence of KD is highest in East Asian countries, such as Japan, Korea, and Taiwan. The incidence in Japanese children is 10-fold higher compared with that in the United States. Japanese data show a seasonal variation in incidence, with an increase seen in the winter; furthermore, siblings of patients with KD are at increased risk of disease compared with the general population,[45] and the risk of sibling patients is increased among patients whose parents previously had the disease.[46] KD shows geographical clustering, with shared clinical features between clusters of disease. These findings support a role for an infectious environmental trigger, together with genetic susceptibility. Several disease susceptibility genes have been identified, including genes that affect the function of molecules involved in the calcineurin-nuclear factor of activated T cells (NFATs) pathway; transforming growth factor-β (TGF-β) lipopolysaccharide-induced endothelial cell inflammation pathways; and genes encoding Fcγ receptors, such as *ITPKC, FAM167-BLK, CD40, FCGR2A, HLA, CASP3, TGFB2, TGFBR2, SMAD3*, and *PLCB4/PLCB1*.[47] *CD40* and *FCGR2A* have been associated with disease susceptibility in Europeans.

SNPs identified within the *ITPKC* and *PLCB4/PLCB1* genes are associated with the risk of coronary artery aneurysms.[47] *ITPKC* and *CASP3* SNPs are associated with lack of response to intravenous immunoglobulin (IVIG) treatment.[47] ITPKC is a negative regulator of T-cell activation via downregulation of the Ca^{2+}/NFAT signaling pathway. The SNP located in the intron of *ITPKC* regulates splicing, decreasing negative regulation of T cells and leading to increased T-cell activation and IL-2 production.[47]

TNF, nuclear factor-κB (NF-κB), IL-17, TGF-β, interferon-γ (IFN-γ), granulocyte–colony-stimulating factor (G-CSF), IL-1β, IL-6, follistatin-like protein 1, and TLR2 are reported to be elevated in patients with KD.[48,49] In the acute phase of disease, circulating IL-17 levels and IL-17–induced cytokinesis are increased, suggesting an imbalance between Th17 and Tregs. Th17 levels are higher in children with coronary lesions.[50] There is evidence for marked activation of cytotoxic and Th lymphocytes and dendritic cells, with upregulation of type I IFN responses, supporting a possible viral etiology of KD—based on a transcriptomic study of patients who died or who were undergoing heart transplantation.[51]

KD has been suspected to be triggered by bacterial or viral infections, but no single causative agent has been identified. It has been suggested that SAgs are implicated. Several microorganisms can produce SAgs, including bacteria (*e.g.*, staphylococci, streptococci, mycobacteria, *Mycoplasma, Yersinia, Lactobacillus)* and viruses (*e.g.*, Epstein-Barr virus). SAg-producing bacteria have been isolated from children with acute KD, including toxic shock syndrome toxin 1 (TSST-1) producing *S. aureus* and pyrogenic exotoxin–producing *Streptococcus*.[52] In a mouse model involving injections of intraperitoneal *Lactobacillus casei*, disease could not be induced in recombinase activating gene-1–deficient mice, implicating T cells as critical drivers of the disease process.[53]

Pathogenesis of Polyarteritis Nodosa

The pathogenesis of idiopathic PAN is unknown; clinical responses to immunosuppressive therapy strongly support an underlying immunological mechanism. Elevated levels of IL-2, IL-8, IFN-γ, TNF, and IL-1β have been reported in PAN, as well as elevated levels of circulating soluble adhesion molecules, such as intercellular adhesion molecule-1, vascular cell adhesion molecule-1, and E-selectin, suggesting that endothelial cell activation could perpetuate and potentiate the inflammatory milieu. IFN-γ and TNF increase the expression of class I major histocompatibility complex (MHC) antigens and induce MHC class II expression, resulting in antigen presentation to T cells. In one study of 24 patients with idiopathic PAN, macrophages and CD4 T lymphocytes were the predominant cell types found in lesions. T lymphocytes were more abundant in nerve biopsy specimens compared with muscle biopsy specimens (52% vs 35%; $P < .001$), while macrophages were predominant in muscle specimens (45% vs 33%; $P = .005$). Histologically, T lymphocytes were distributed throughout the vessel wall, whereas macrophages predominated in the periphery.[54]

In HBV-induced PAN, there is direct injury to the vessel wall as a result of viral replication[27] and deposition or *in situ* formation of immune complexes, which activate complement, thereby recruiting and activating neutrophils. The immunological process usually occurs within 6 months of HBV infection. During the active phase of disease, complement levels are low, consistent with complement consumption resulting from immune-complex deposition.

Alterations in the *MEFV* gene (encoding pyrin) may represent an important susceptibility factor for PAN. Pyrin is critically important in regulating IL-1β production; mutations of pyrin may be associated with loss of regulation of inflammatory pathways via apoptosis and IL-1β release, resulting in augmented inflammation, similar to patients with familial Mediterranean fever, who can be carriers of *MEFV* mutations.[55]

PATHOGENESIS OF CRYOGLOBULINEMIC VASCULITIS

The pathological hallmark of cryoglobulinemic vasculitis is deposition of cold-precipitating Igs in small vessels, resulting in vas-

cular inflammation and damage. In most patients, chronic HCV infection is the trigger, but cryoglobulinemic vasculitis can also occur in the context of other chronic infections and in some autoimmune rheumatic diseases and lymphoproliferative disorders.

HCV plays several important roles in the pathogenesis of cryoglobulinemic vasculitis. HCV glycoproteins interact directly with B-cell surface receptors, lowering the activation threshold and resulting in polyclonal activation and expansion of B cells.[56] Circulating, clonally expanded B cells produce monoclonal IgM rheumatoid factor (RF). Mixed cryoglobulins contain a monoclonal IgM RF directed against the Fc segment of IgG and polyclonal IgG. When IgM RF is exposed to cold conditions, it undergoes conformational changes, resulting in cryoprecipitation. Cold-insoluble immune complexes are formed predominantly by IgM RF linked to IgG (which, in turn, is bound to HCV core protein). Cryoglobulinemic vasculitis should be considered an antigen-driven disease, because HCV persistence ensures sustained lymphoid proliferation and continued cryoglobulin production.

C1q protein levels and C1q binding are significantly increased in cryoglobulinemic vasculitis. C1q is required for binding of immune complexes to endothelial cells. Although mixed cryoglobulins can be detected in up to 60% of patients with chronic hepatitis C, cryoglobulinemic vasculitis occurs only in a minority, suggesting that host factors must be equally important in the pathogenesis of the disease.

PATHOGENESIS OF EOSINOPHILIC GLOMERULONEPHRITIS WITH POLYANGIITIS

Despite the established role of ANCA in GPA and MPA, the pathogenic role of ANCA in EGPA is less clear. In EGPA, only 30% to 40% of patients are ANCA-positive; the presence or absence of ANCA in EGPA defines distinct clinical phenotypes. The pathological hallmark of EGPA is prominent tissue and peripheral blood eosinophilia, suggesting a central pathogenic role for the eosinophil. EGPA is considered to be Th2 mediated, given the striking early allergic manifestations, together with the marked increase in circulating Th2 cytokines, including IL-5, which plays a crucial role in eosinophil differentiation in bone marrow, as well as recruitment and activation of eosinophils at sites of inflammation. IL-5 could delay apoptosis of eosinophils and modulate the function of mast cells and basophils.[57] Elevated levels of IFN-γ (a potent Th1 cytokine involved in granuloma formation) provide evidence for the involvement of other Th responses. Th17 responses are upregulated, and B lymphocytes and humoral responses are likely to be important, given that B-cell depletion can induce remission and reduce circulating IL-5 and eosinophils.

The precise mechanisms of eosinophil-mediated inflammation in EPGA remain unclear. Eosinophils are granulocytic cells capable of releasing multiple proinflammatory cytokines, chemokines, and reactive oxygen species, which have direct effects on the vessel and perivascular tissues—including tissue fibrosis, thrombosis, and allergic inflammation. Indirect effects include recruitment and activation of other inflammatory cells to perpetuate the inflammatory response.

CLASSIFICATION OF VASCULITIDES

Several sets of classification criteria have been published, the most recent being one by the American College of Rheumatology (ACR) in 1990. These are primarily for the purposes of epidemiological studies or for defining patients for inclusion in clinical studies, rather than being used in clinical practice. However, as is often the case with classification criteria, they are misapplied as diagnostic criteria in daily practice and may fail to provide a clear-cut distinction between vasculitis and other conditions.[58] An international effort is currently underway to develop improved classification criteria as well as diagnostic criteria for patients with vasculitis or suspected vasculitis, and it is likely to lead to a new set of evidence-based criteria to improve diagnosis and classification of vasculitis in the future.[59] There is considerable overlap between different forms of vasculitis. The group of patients with diseases clustered around the presence of ANCA share many features in common, particularly the presence of glomerulonephritis and pulmonary infiltrates. In the case of EGPA, patients also experience rhinitis, nasal polyps, asthma, and rashes and may develop mononeuritis multiplex or sensory peripheral neuropathy. In addition to renal and lung inflammation, patients with GPA typically suffer from upper respiratory problems, such as nasal crusting or discharge; subglottic stenosis; or hearing loss. Patients with MPA do not usually develop significant upper airway disease; their disease tends to be more limited to the kidneys and/or lungs. All of these conditions can also affect any other organs, such as skin. Skin is the most commonly affected site in most forms of vasculitis. This can vary from small infarcts around nail edges to purpura, ulcers, nodules, and even gangrene.

The development of consensus criteria for diagnosis by the Chapel Hill group has provided a more rational approach to define disease entities. They are not true classification or diagnostic criteria because they are not evidence based (they are consensus of opinion from several experts), but they have advantages over the 1990 ACR criteria. They were developed after recognition of ANCAs, which are an important part of the 2013 definitions. One of the major flaws of the ACR criteria is the failure to recognize MPA as an entity separate from PAN.[59] This failure to separate PAN and MPA has probably prevented proper diagnosis, and it is a historical legacy from an assumption that almost all forms of vasculitis are, in fact, variations of PAN.

Furthermore, current classification criteria offer limited information about the severity, prognostication or relapse risk of AAV. Identifying and predicting organ involvement guides effective treatment of AAV. For example, while isolated orbital disease may be considered as limited disease, it is organ-threatening and may also evolve to affect new organs—through a continuum of disease severity—thereby raising the question of whether all forms of AAV should be regarded as severe and be treated as such. Alternative classifications—involving MPO and PR3 ANCA serological subtypes—have been proposed to allow the opportunity to add new organ involvement if they occur, in turn recognizing the dynamic nature of AAV.[60] This may help individualize cases and provide information about the severity and localization of lesions in order to guide more effective therapeutic decisions.

DIAGNOSIS

AAV are a group of three main entities: GPA, MPA, and EGPA. These three conditions overlap with many shared clinical features. Figs. 59.2 and 59.3 describe clinical patterns in patients with MPA and GPA and show that there are many similarities, with a dominance of kidney and lung involvement in both conditions. EGPA is characterized by eosinophilia, late-onset asthma, and neuropathic involvement, but cardiac involvement

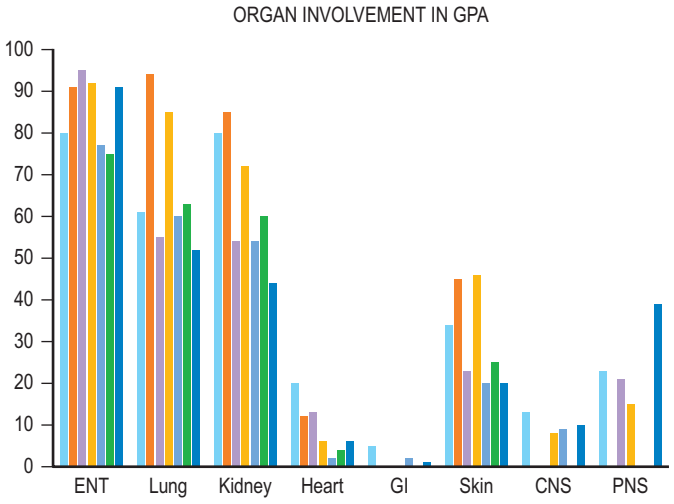


FIG. 59.2 Organ involvement in patients with granulomatosis with polyangiitis. *ENT*, Ear, nose, and throat; *GPA*, granulomatosis with polyangiitis. (Data compared across seven cohorts—reviewed by Luqmani R, Ponte C. Antineutrophil cytoplasmic antibody associated vasculitides and polyarteritis nodosa. In: Bijlsma JWJ, Hachulla E, eds. *EULAR Textbook on Rheumatic Diseases*. London, UK: BMJ Publishing Group Ltd; 2015:717-753, Chapter 27).[75]

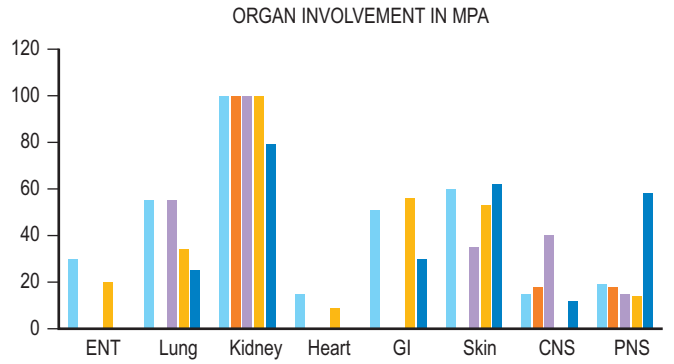


FIG. 59.3 Organ involvement in patients with microscopic polyangiitis. *ENT*, Ear, nose, and throat; *MPA*, microscopic polyangiitis. (Data compared across seven cohorts—reviewed by Luqmani R, Ponte C. ANCA associated vasculitides and polyarteritis nodosa. In: Bijlsma JWJ, Hachulla E, eds. *EULAR Textbook on Rheumatic Diseases*. London, UK: BMJ Publishing Group Ltd; 2015:717–753, Chapter 27.)[75]

may develop in a subset of patients with EGPA, particularly in association with hypereosinophilia. Systemic features, such as malaise, fever, weight loss, or myalgia, could be mistaken for a variety of other conditions, and it may delay recognition of disease. Patients with MPA can present just with isolated microscopic hematuria and hypertension, with nonspecific systemic features. In contrast, patients with GPA typically suffer from upper airway problems, with nasal crusting, blood-stained discharge, sinusitis, or hearing loss. In anti-glomerular basement membrane (anti-GBM) antibody disease, clinical features sometimes overlap with those of AAV, but the dominant features are pulmonary hemorrhage leading to hemoptysis and rapid deterioration in renal function. It is not entirely clear whether the anti-GBM disease should be regarded as vasculitis or vasculopathy, but it has been included in the current Chapel Hill Consensus definitions of a small-vessel vasculitis. Confusingly, some patients with anti-GBM antibodies also have ANCA.

The clinical course in MPA is typically acute, whereas GPA may remain undiagnosed for several months or even years before the upper airway and or lower airway symptoms are recognized as being related to vasculitis. For almost all forms of acute AAV, the differential diagnosis is infection, cancer, or drug toxicity. One of the drugs recently recognized to induce an AAV mimic is cocaine,[61] which can cause local tissue necrosis, especially in the palate, leading to palatal and nasal septal perforation. Interestingly, cocaine can induce ANCA production and therefore can be a true mimic of AAV. Typically, the ANCA is directed against elastase or is nonspecific. The use of cocaine continues to grow worldwide and is an increasing public health concern. Cocaine-induced vasculitis is an important differential diagnosis to consider. Two distinct vasculitic syndromes have been described due to cocaine: CIMDL and AAV. The former occurs as a result of direct vasoconstriction effect of cocaine causing ischemic necrosis of the septal cartilage and perforation of the nasal septum, mimicking GPA. The latter, which can present with high titers of p-ANCA and c-ANCA with MPO and/or PR3 positivity, is due to contamination of cocaine with levamisole; the main features being cutaneous involvement, arthralgia, otolaryngologic manifestations, agranulocytosis, and rarely renal involvement. Contamination of cocaine with levamisole is becoming a growing problem. Presumably, levamisole (a veterinary anthelmintic agent) is a cheap white powder that can be mixed readily with cocaine but can cause an acute necrotizing vasculitis of skin and the extremities. A high degree of suspicion and awareness is needed in order to properly diagnose and treat these patients. Careful history taking, sometimes with sensitive, confidential, repeated questioning, may be required to eventually identify this as a cause of a patient's condition. Cessation of cocaine and levamisole use is essential and may be the only step required to resolve this clinical condition.[62] However, once palatal or nasal perforation has developed, patients may require local surgical repair.

Cryoglobulinemic vasculitis typically presents with purpuric lesions on the legs, and the lesions may ulcerate. Patients often complain of joint pains and malaise. Neuropathies can occur with either mononeuritis multiplex or sensory neuropathy; membranous glomerulonephritis is a recognized complication. Most cases appear to be associated with hepatitis C (more than 90% in some series), and eradication of hepatitis C appears to be effective in reducing the manifestations of disease.

True PAN typically presents with neuropathies, systemic inflammation, and ischemic abdominal pain (difficulty eating because of medium-vessel supply to the gut and nerves). If it is associated with hepatitis B, the patient may or may not have obvious signs of liver disease.

Isolated cutaneous vasculitis typically presents as purpuric lesions, skin ulceration, or broken livedo, and biopsy of the skin will show typical changes of a necrotizing vasculitis.

KD usually presents in childhood as an acute illness, with relapsing high fevers, significant lymphadenopathy, and a systemic inflammatory syndrome. The child typically has mucosal inflammation with strawberry tongue; about 10 days after the onset, skin desquamation is a very typical feature. The most worrying feature of KD is the development of coronary artery aneurysms, which occur in around 2% to 4% of children and can be fatal, if untreated. This can be detected by using echocardiography.

Laboratory Investigations

The investigation of patients with suspected small- or medium-vessel vasculitis should follow a careful history and examination to determine the likely diagnosis and underlying illness. The differential diagnosis is very wide. It is important to be vigilant in looking for positive signs of vasculitis, but it is equally important not to forget to look for more common causes of the clinical presentation. Many of the studies performed can result in nonspecific findings, such as an elevated white blood cell count, platelet count, or the erythrocyte sedimentation rate. The C-reactive protein (CRP) level is typically raised. Patients may be anemic. Liver or, more importantly, renal function may be abnormal. The presence of hypereosinophilia can be suggestive of a diagnosis of EGPA, but there are other causes of hypereosinophilia, particularly drug reactions. It is important to test renal function in patients with suspected vasculitis in case there is nephritis, and it is always important to perform urinalysis for microscopic hematuria or proteinuria. Although these might be explained by the presence of a kidney infection or other causes, they raise a strong suspicion of glomerular inflammation. An abnormal urinary sediment in combination with hypertension should alert the physician to the possibility of kidney involvement by small-vessel vasculitis. Since the 1980s, the discovery of ANCA[4] has transformed the recognition of renal vasculitis, so that these patients can be managed more effectively.

Histology remains a critically important diagnostic test, not only to make a positive diagnosis but also to exclude other causes. Although histology from the airways can be nondiagnostic, this may still assist in ensuring that the patient does not have cancer, sarcoidosis, tuberculosis, or IgG4-related disease, all of which could present in a similar way with inflammation of the upper or lower airway. IgG4-related disease (IgG4-RD) is an important differential diagnosis due to its propensity to mimic AAV. For instance, nasal manifestations of IgG4-RD, causing chronic sinusitis and paranasal sinusitis, and orbital manifestations of IgG4-RD can be remarkably similar to limited GPA. The diagnosis of IgG4-RD is based on a combination of clinical presentation and histology. A histological diagnosis, the gold standard, requires demonstration of IgG4-positive plasma cells in tissue specimens and a ratio of IgG4-positive to total IgG-positive plasma cells greater than 40%. Accurate and swift diagnosis of IgG4-RD is important because potential irreversible tissue damage due to fibrosis may occur. Prolonged untreated inflammation due to IgG4-RD may also lead to amyloidosis, emphasizing the importance of timely recognition and treatment.[63]

Renal histology is still the gold standard to diagnose suspected glomerulonephritis and may be useful in predicting the prognosis.[64] Four categories of renal lesion have been proposed: focal, crescentic, mixed, and sclerotic. Follow-up of patients with different patterns has shown progressively worse renal outcome from focal (the best) through to sclerotic (the worst) over the subsequent 5 years.

ASSESSMENT

The outcome of small- and medium-vessel vasculitis has been completely transformed by immunosuppressive therapy. With treatment almost all patients with KD recover from their initial illness. In AAV, more than 70% survive at least 5 years from starting treatment. The majority of patients with PAN respond to initial therapy. However, the risk of recurrence of disease is high in small-vessel vasculitides. Relapse rates are likely to exceed 50% over time.[2] Patients experience comorbidities as a consequence of the disease or its treatment. A central issue is their immunocompromised state, rendering them susceptible to infections. In the long term, small-vessel vasculitides and their treatment can be associated with effects on the cardiovascular system with increased risk of hypertension, coronary heart disease, and stroke. Immunosuppressants given to patients can significantly improve disease control but may increase the risk of malignancy. The continued need for glucocorticoid therapy contributes to hypertension, diabetes, heart diseases, osteoporosis, and infection.

It is, therefore, obvious that careful evaluation of a patient's status is required throughout the course of disease. Constant vigilance is required to detect and manage any flare-up of disease because the consequences of untreated inflammation of vessels to vital organs can be severe. Although for most patients, initial therapy is very successful and improvements in disease status is obvious to the patient, the subsequent disease course can be much more complicated because of comorbidities, the evolution of disease-related damage, and morbidity caused by treatment as well as disease flare-ups. There are no suitable biomarkers that can be universally applied in small- and medium-vessel vasculitides to determine the patient's disease state to provide an evidence-based rationale for adjustment of therapy. ANCA titers fluctuate during the course of disease, and although ANCA remains a very useful diagnostic test, its value in managing variations in disease activity is very limited. Up to 40% of patients have an elevation of ANCA without any new deterioration in clinical state, but a recent study has shown that patients with renal involvement as part of their presentation are likely to have relapses associated with ANCA rises.[65] Among 166 patients with AAV, all were positive for ANCA and 104 had renal involvement (a mixture of PR3-ANCA and MPO-ANCA). The hazard ratio for ANCA rises predicting subsequent relapse was 11.09 (CI 5.01 to 24.55), suggesting that this test may be of value in detection of future relapse in this subset of patients. However, for most patients with systemic vasculitis, careful clinical evaluation remains a cornerstone of effective disease management as recommended by the EULAR guidelines on management of systemic vasculitis.[66]

In simple terms, at the onset of the condition, disease activity is the dominant problem, and treatment should be directed toward it; but over the course of time, with the development of consequences of vasculitis or its treatment, there is an increasing component of damage or scarring and also side effects of drug therapy. Similarly, different patients function at different levels with the same amount of disease, or damage, and therefore, the ability to perform normal tasks is an important component of their overall condition to consider (Table 59.3).

KEY CONCEPTS

Assessment of Disease Activity in Vasculitis

- For small- and medium-vessel vasculitis in adults, version 3 and a specific version of the Birmingham Vasculitis Activity Score (BVAS) for GPA are used.
- BVAS version 3 is the most generic and applicable across different forms of vasculitis; BVAS for GPA has been specifically designed for use in GPA and contains very specific items related to that condition.
- During sequential monitoring, a time frame of 3 months is recommended so that disease activity is considered to be of most relevance during this time. The time frame is based on pragmatic clinical experience that this is the usual time taken during which immunosuppressive therapy is likely to have a significant effect on active disease manifestations.

TABLE 59.3 Assessing Disease Activity in Vasculitis

System Assessed	BVAS	BVAS GPA	PVAS	Comments
General	Myalgia, arthralgia/arthritis, fever, weight loss	Arthritis, fever	Myalgia, arthralgia/arthritis, fever, weight loss	These are typical features in many forms of systemic vasculitis.
Skin	Infarct, purpura, skin ulcer, gangrene	Purpura, skin ulcer, gangrene	Polymorphous exanthema, livedo, panniculitis, purpura, skin nodules, infarct, ulcer, gangrene, other skin vasculitis	Skin is a common organ involved in most forms of small and medium vessel vasculitis. Skin manifestations in AAV tend to be less serious compared with manifestations in kidneys, lungs, and upper airways. For skin vasculitis in children, the manifestations are more diverse.
Mucous Membrane/eyes	Mouth ulcers, genital ulcers, adnexal inflammation, proptosis, red eye, (epi)scleritis, conjunctivitis, blepharitis, keratitis, blurred vision, visual loss, uveitis, retinal vasculitis	Mouth ulcer, conjunctivitis/episcleritis, retro-orbital mass, uveitis, scleritis, retinal exudates	Mouth ulcers, genital ulcers, adnexal inflammation, proptosis, red eye (epi)scleritis, conjunctivitis, blepharitis, keratitis, blurred vision, visual loss, uveitis, retinal vasculitis	Eye involvement is most common in GPA and less often seen in MPA or EGPA.
Ear, nose, and throat	Bloody nasal discharge, paranasal sinus involvement, subglottic stenosis, conductive hearing loss, sensorineural hearing loss	Blood nasal discharge, paranasal sinus involvement, swollen salivary glands, subglottic inflammation, conductive hearing loss, sensorineural hearing loss	Blood nasal discharge, paranasal sinus involvement, subglottic stenosis, conductive hearing loss, sensorineural hearing loss	ENT manifestations are most common in GPA and rarely seen in MPA or EGPA. However, EGPA is often characterized by the presence of nasal polyps, which are inflammatory in nature.
Chest	Wheezing, nodules or cavities, pleural effusion, infiltrates, endobronchial changes, massive hemoptysis, respiratory failure	Nodules or cavities, pleurisy, infiltrates, endobronchial changes, alveolar hemorrhage hemoptysis, respiratory failure	Wheezing, nodules or cavities, pleural effusion, infiltrates, endobronchial changes, massive hemoptysis, respiratory failure	The chest is commonly affected in all three forms of ANCA vasculitis. Wheezing is a common feature of EGPA as are infiltrates. In contrast, infiltrates, nodules, and endobronchial disease dominate in GPA. Massive hemoptysis can occur in patients with GPA or MPA and less frequently in patients with EGPA, but may also be a typical feature with GBM disease.
Cardiovascular system	Loss of pulses, ischemic cardiac pain, cardiomyopathy, congestive cardiac failure, valvular heart disease, pericarditis	Pericarditis	Loss of pulses, bruits, blood pressure discrepancy, claudication, ischemic cardiac pain, cardiomyopathy, congestive cardiac failure, valvular heart disease, pericarditis	Cardiovascular manifestations are most widely recognized in KD, which is not particularly well covered by PVAS. Although CVD manifestations do occur in small- and medium-vessel vasculitis, they are more typically seen in large vessel diseases, such as Takayasu arteritis.
Abdominal system	Ischemic abdominal pain, peritonitis, bloody diarrhea	Mesenteric ischemia	Abdominal pain, peritonitis, bloody diarrhea, bowel ischemia	Gut involvement is more typical in medium-vessel vasculitis, especially polyarteritis nodosa, but is well recognized in patients with GPA, especially with colitis giving rise to bloody diarrhea.
Renal system	Hypertension, proteinuria, hematuria, impaired renal function, deterioration in renal function	Hematuria, red cell casts, or glomerulonephritis, deterioration in renal function	Hypertension, proteinuria, hematuria, impaired renal function, deterioration in renal function	Renal involvement in small-vessel vasculitis is one of the major manifestations, leading to organ failure and death and should be carefully assessed. Renal involvement in medium-vessel vasculitis is much less common and takes the form of infarction of segments of the kidney, leading to hematuria and hypertension with consequent impairment of renal function.
Nervous system	Headache, meningitis, organic confusion, seizures, stroke, cord involvement, cranial nerve, lesion, sensory or motor neuropathies	Meningitis, stroke, cord involvement, cranial nerve lesion, sensory or motor neuropathies	Headache, meningitis, organic confusion, seizures, stroke, cord involvement, cranial nerve, lesion, sensory or motor neuropathies	Neurological involvement is a common feature in small- and medium-vessel vasculitis; often it does not lead to immediate loss of life. Strokes are less common, whereas peripheral neuropathies are more common and can cause long-term disability.

AAV, ANCA-associated vasculitides; *ANCA*, antibodies to neutrophil cytoplasm; *BVAS*, Birmingham vasculitis activity score; *CVD*, cardiovascular disease; *EGPA*, eosinophilic glomerulonephritis with polyangiitis; *GBM*, glomerular basement membrane; *GPA*, granulomatosis with polyangiitis; *KD*, Kawasaki disease; *MPA*, microscopic polyangiitis; *PVAS*, pediatric vasculitis activity score.

The primary tool used is the BVAS in adults and the Pediatric Vasculitis Activity Score (PVAS) in children (reviewed by Ponte et al.).[67] Training is recommended for assessment in the BVAS. BVAS provides a quantitative score based on individual items, providing an effective means to define the patient's status with regards to response to therapy. Many recent studies of different immunosuppressive agents have made use of the BVAS either to define improvement in terms of a fall in BVAS score or to define a cut-off representing active disease, or inactive disease, or flare-up, depending on the number of items present. Although it is subject to observer variability, it provides an effective means by which groups of patients can be compared against each other and allows individual patients to be followed up over the course of their condition. The score is weighted according to the organ system and the individual manifestation to reflect the severity of the disease. The range of scores for BVAS and PVAS are 0 to 63. For BVAS GPA, items are divided into major (scoring 3 points) and minor (scoring 1 point each); with 15 to 19 major items and 19 to 23 minor items, the range of sores is 0 to 76. The PVAS was developed as a pediatric version of the BVAS but validated in children with vasculitis and demonstrated to be effective at discriminating different disease states. It is increasingly used as a research tool in pediatric vasculitis.

Damage Assessment in Vasculitis

The concept of damage in patients with vasculitis is about the permanent consequences of having vasculitis. It is an attempt to measure disease burden regardless of cause. The Vasculitis Damage Index (VDI) is the most widely used and validated measure for assessment of damage in vasculitis (reviewed by Ponte et al.).[67] The VDI captures the long-term consequences of a diagnosis of vasculitis and its treatment and associated co-morbidities. Damage is defined as lasting at least 3 months or occurring at least 3 months ago for single time point events (*e.g.*, stroke or myocardial infarction) and should be recorded as a permanent change to the patient's damage status. A VDI of more than 5 points recorded within 6 months of disease carries a significant increased risk of subsequent mortality compared with a lower damage index at 6 months (reviewed by Ponte et al.).[67] VDI is a useful index of future harm. A further development of the damage index has been the combined damage assessment (CDA). In a comparative study with the VDI, CDA was shown to be inferior to it (reviewed by Ponte et al.).[67]

TREATMENT

Once a diagnosis of small- and medium-vessel vasculitis has been established, treatment should be focused on patients and their problems rather than the specific diagnosis. Treatment for different forms of vasculitis may look very different, but many aspects of different forms of vasculitis require the same therapeutic approach. Without a clear understanding of the underlying pathogenesis of disease, we are inevitably led by the need to suppress inflammation and reduce damage to prevent mortality and improve survival. However, as well as modifying immune dysregulation, it is important to consider other aspects of the patient's condition, such as comorbidities and the prevention of future comorbidities.

No Treatment/Symptom Relief

Small-vessel vasculitis, such as isolated cutaneous vasculitis resulting from infection or use of pharmaceutical agents, may respond to simple withdrawal of the offending agent or resolution of the infection without the need for specific treatment. However, for more recalcitrant disease, symptom relief may be required and occasionally systemic steroids. Symptom relief could be provided in the form of antipruritic agents or topical cream to reduce skin inflammation and/or topical steroids. NSAIDs can be helpful in relieving symptoms of joint pain or swelling. As monotherapy they are unlikely to resolve skin manifestations but could be tried in combination with other therapies. Colchicine has been used for skin vasculitis and occasionally can be effective, although the doses required should remain below 2 mg/day to avoid the predictable side effects of abdominal cramps and diarrhea.

Target-Directed Therapies

In diseases where there is a clearly defined provoking agent, as in hepatitis B–related PAN or hepatitis C-related cryoglobulinemic vasculitis, eradication of the virus is a key part of treatment of the disease. Effective antiviral agents play a vital role in the management of virus, associated with the need for immunosuppression. Hepatitis B-related PAN is treated with a combination of antiviral therapy plus plasma exchange to remove immune-complexes and other inflammatory mediators combined with a course of glucocorticoid therapy. For hepatitis C-related cryoglobulinemic vasculitis, recent reports of virus eradication may also be transforming the outcome of this disease. Unfortunately, the toxicity of these regimens can be considerable, with over 40% requiring erythropoietin, red blood cell transfusions, and/or G-CSF.

Specific Therapies

In KD, although the etiological factors have not yet been defined, it seems likely that this is related to some kind of infectious agent (see earlier section on pathogenesis). The most effective therapy is use of high doses of IVIG (0.4 g/kg per day for 5 days) combined with high doses of aspirin, which is usually curative. Whether or not this will prevent long-term harm to the cardiovascular system, particularly the coronary arteries, remains to be explored.

Glucocorticoids

Glucocorticoids (see Chapter 83) remain a cornerstone in the management of most forms of multisystem vasculitis. They are relatively contraindicated in KD because they may potentially worsen the development of coronary artery aneurysm, but they have been used in combination with IVIG and aspirin with beneficial outcomes. However, they are an integral part of almost all therapeutic regimens for management of vasculitis. In some instances, such as isolated skin vasculitis, they are the only treatment required, but more often they are insufficient on their own without causing significant morbidity from side effects. Typical doses of glucocorticoid therapy are 1 mg/kg per day over a period of 2 to 4 weeks, reducing to around 10 to 15 mg/day within 6 months, and then slowly withdrawing steroids in the next 6 to 12 months. The use of high-dose intravenous methylprednisolone is popular but lacks evidence. The only randomized trial of intravenous methylprednisolone compared it against plasma exchange in patients with severe AAV (reviewed by Ponte et al.).[67] This study demonstrated the superiority of plasmapheresis over IV prednisolone, both used as adjunct therapies (in combination with cyclophosphamide and high doses of oral prednisolone)

for treatment of severe AAV (mainly MPA plus some cases of GPA) with significant renal impairment (creatinine levels above 500 μmol/L [5.66 mg/dL]).

The role of glucocorticoid therapy has been increasingly challenged by more recent trials using smaller doses for shorter periods or even eliminating steroid use completely in some instances. While intensive immunosuppression using high doses of corticosteroids remains established as part of the initial management of AAV, close consideration needs to be given to the extensive side-effect profile associated with corticosteroids. There is significant morbidity associated with corticosteroid use in patients with AAV within just one year of treatment—8.2% developing new onset diabetes (50% of which occurred within 1.7 months), 29% gaining over 10 kg in weight, 2.6% developing peptic ulceration, 2.5% suffering insufficiency fractures, 2% developing cataracts and 0.4% developing avascular necrosis. These adverse effects become more pronounced with increasing length of corticosteroid exposure. After a median of 5 years, 41% of patients develop hypertension, 38% become osteoporotic, 28% develop diabetes mellitus and 25% develop cataracts.[68] The risk of osteoporosis is largely preventable with concurrent use of bisphosphonate therapy (unless there is significant renal dysfunction) with supplementary calcium and vitamin D replacement. The increased risk of cardiovascular disease (CVD) in AAV patients is not only associated with accelerated atherosclerosis due to inflammation, but also corticosteroid use causing multiple metabolic side effects including hypertension, hyperlipidemia, weight gain and diabetes, which are all significant cardiovascular risk factors.[68] Indeed, the challenge is balancing risk of AAV relapse with risk of corticosteroid-related morbidity. There are encouraging data from the PEXIVAS and CLEAR studies that indicate the possibility of effective treatment of AAV with reduced doses of corticosteroid.[2,69]

Other Immunosuppressive Therapies

Cyclophosphamide (see Chapter 84) has been available since the 1950s but was first used for the management of systemic vasculitis in the 1970s and remains the most effective agent we have for managing multiorgan systemic vasculitis. Initially it was used as a daily oral agent at 2 mg/day to 3 mg/day, and it transformed the outcome of patients with AAV from inevitable mortality to a high likelihood of survival. It is a cytotoxic agent and carries with it the risk associated with chemotherapy, including increased risk of malignancy, especially in the bladder because it is predominantly excreted through the kidneys and accumulates in the bladder. Initial protocols were associated with excessive risks of bladder cancer (~33-fold), but despite this, the daily oral cyclophosphamide dosing regimen was not effective in maintaining control of disease. Therefore, over the past 20 years or so, there have been a number of trials comparing reduced doses of cyclophosphamide, high-dose intermittent pulse therapy (reviewed by Yates et al.),[66] or with combination strategies of induction with short courses of cyclophosphamide followed by a switch to another drug for maintenance[66] or by replacing cyclophosphamide with another agent, such as methotrexate (MTX). All these studies[66] have demonstrated the equivalence of using shorter courses of daily oral cyclophosphamide and, more recently, of using of high-dose intermittent pulses of cyclophosphamide to reduce the total dose even further. The total cumulative dose from six cycles of cyclophosphamide over a period of 3 months would be 6 g (based on 15 mg/kg per treatment on six occasions). This compares with 9 g to 12 g of cyclophosphamide given as daily oral therapy over 4 to 6 months.[66]

Although the relapse rate for patients given high-dose intermittent cyclophosphamide was higher during the subsequent 5 years compared with that for patients who were not given daily oral cyclophosphamide, relapse was always effectively managed with reintroduction of therapy and never led to mortality.[66] The consequences of exposure to cyclophosphamide are likely to be a risk of cancer (more recent studies suggest that this risk is relatively modest now that the regimens include much lower total doses), infertility, hair loss, nausea, vomiting, diarrhea, cytopenia, and increased risk of infection. Less common complications of cyclophosphamide include hyponatremia. The introduction of rituximab for AAV has had a significant impact on the use of cyclophosphamide, with increasing numbers of patients being managed with rituximab in place of cyclophosphamide, especially if patients are of child-bearing years or if there is a potential contraindication to using cyclophosphamide, such as a previous history of bladder cancer.

Because of the nature of AAV, relapse is common. Therefore, a single course of therapy rarely achieves long-lasting remission. Repeat cycles of treatment are likely to be required, which accounts for the accumulation of higher doses of cyclophosphamide, especially in the pre-rituximab era. Therefore, although each individual course of therapy may only contain 6 g to 9 g, during a patient's lifetime, they may require treatment for several relapses, which would then start building up the total exposure to cyclophosphamide. Nevertheless, even though it is being slowly replaced, it presently remains an important aspect in the therapy of vasculitis.

Azathioprine is an immunomodulator with cytostatic properties. It inhibits cell division; it has been an effective immunosuppressant agent for decades. It was first used in combination with steroids and found to reduce mortality in systemic vasculitis in an open-label retrospective study of 64 patients.[66] The 5-year survival of patients given no therapy was 12%, and those given steroids alone had a 53% survival rate, whereas those treated with steroids plus another agent (mainly azathioprine but a few had cyclophosphamide) had a survival rate of 80%. It is an oral medication given at 2 to 2.5 mg/kg per day, and it has largely been superseded as an induction agent by cyclophosphamide. Azathioprine is usually now used as maintenance therapy once the disease has been controlled with another agent. In newly diagnosed AAV following induction with cyclophosphamide and glucocorticoids, rituximab was shown to be more effective in leading to sustained remission at 28 months when compared to azathioprine.[70] Following remission-induction with rituximab, rituximab was shown to be superior to azathioprine for preventing disease relapse in patients with AAV with a prior history of relapse.[71] Prolonged remission therapy with combination azathioprine and prednisolone for a duration of 48 months following initial diagnosis of AAV resulted in fewer relapses and improved renal survival when compared to withdrawal of treatment 24 months after initial diagnosis.[72] Azathioprine has been shown to be equivalent to MTX and superior to mycophenolate[66] as maintenance therapy. It is a relatively safe immunosuppressant and can be used safely throughout pregnancy.

MTX is popular among rheumatologists but less so among renal physicians who regard it with some suspicion because of its potential for nephrotoxicity. The latter is only the case for patients who have well-established renal disease (typically with a creatinine level greater than 300 μmol/L). MTX is an effective

immunosuppressant agent used very widely in the management of inflammatory arthritis and has found its place in the management of GPA, where studies have demonstrated its efficacy in comparison to oral cyclophosphamide.[66] However, it needs to be given continuously, rather than as an induction regimen over the short period of time. Although MTX is as effective as cyclophosphamide in inducing remission in GPA, stopping the drug inevitably leads to relapse. In terms of maintenance therapy, MTX has been shown to be equally as effective as cyclophosphamide in maintaining disease remission at 12 and 24 months, following remission-induction with pulsed cyclophosphamide, although levels of proteinuria were significantly lower at 24 months in the cyclophosphamide arm.[73] MTX is recommended for non-life-threating AAV, usually in combination with steroid treatment. MTX is available either as oral, intramuscular, or subcutaneous administration. The dose is 20 mg/week to 25 mg/week in most studies of vasculitis. It is contraindicated in pregnancy.

Cotrimoxazole, an antibiotic containing sulfonamide and trimethoprim, was fortuitously discovered to have beneficial effects in patients with GPA who had been treated coincidentally for infections. There has been a significant advance in our understanding of GPA, partly as a result of this historical experiment and partly based on suggestions that *S. aureus* plays a role in initiating disease by its effect on the nasal mucosa. Eradication of the organism has been suggested as one mechanism by which it works, although the drug is not particularly effective against this organism. It is more likely that the drug has an immunosuppressive effect in itself; it has been demonstrated to be effective in combination with low-dose steroids in a randomized trial of localized forms of GPA.[66] It is also commonly used as a prophylactic agent against *Pneumocystis jiroveci* infection; it is given three times per week for patients receiving other more potent immunosuppressant therapy, such as cyclophosphamide or even MTX (despite the potential for drug interactions leading to anemia).

Mycophenolate mofetil is a widely used transplantation drug that has been tested in AAV but is less effective than azathioprine as maintenance agent for patients who have achieved remission. Trial data comparing mycophenolate to cyclophosphamide revealed that cyclophosphamide was more effective than mycophenolate for remission-induction in patients with non-life-threatening relapses of AAV (azathioprine being used as maintenance therapy).[74] In patients with newly diagnosed AAV, although mycophenolate was shown to be non-inferior to cyclophosphamide in remission-induction at 6 months, higher rates of relapse were seen in the mycophenolate group (azathioprine being used as maintenance therapy).[75] Mycophenolate is typically given as 2 to 3 g/day as an oral dose, along with reducing courses steroids. It is contraindicated in pregnancy.

Cyclosporine is a well-established immunosuppressive drug that has been used for transplantation for many decades. Cyclosporine has been given to limited numbers of patients with systemic vasculitis, with one small well-conducted trial[76] in 32 patients with GPA. In combination with plasmapheresis, it was as effective as continuous oral cyclophosphamide as a maintenance agent. It is generally limited by its toxicity and is not routinely used.

Leflunomide, an antilymphocyte agent used extensively for the management of inflammatory arthritis, has been tested in patients with AAV in limited trials demonstrating its ability to maintain remission. Indeed, in a recent meta-analysis, it was reported to be superior to azathioprine, MTX, and mycophenolate as a maintenance agent for AAV, but more trial data are needed. It is an oral agent that is characterized by a very long half-life, and it is not suitable for use in pregnancy.

The role of hydroxychloroquine in mild forms of vasculitis is uncertain. There is anecdotal evidence that it is beneficial in patients with skin manifestations with small-vessel vasculitis. Its use probably stems from its known effects in the treatment of connective tissue diseases with skin and joint manifestations, such as SLE. Other similar agents, such as mepacrine and dapsone, have also been used for skin vasculitis with occasional reported positive outcomes. However, the potential toxicity for each of these agents should be considered alongside the relatively limited evidence for benefit.

Specific Immunotherapy

Better understanding of the pathogenesis of vasculitis (see section on pathogenesis) has led to the development of targeted immunotherapy for some of these diseases. Rituximab is an mAb against B cells and has been in widespread use for treatment of RA. Its role as an effective agent in AAV is well established with two randomized trials demonstrating efficacy comparable with cyclophosphamide (reviewed by Yates et al.).[66] Rituximab induction therapy is just as effective as cyclophosphamide for moderate and moderate-to-severe AAV. Indeed, there were no differences in remission rates or increases in eGFR at 18 months (regardless of ANCA type) when comparing rituximab plus glucocorticoids versus cyclophosphamide plus glucocorticoids, in both new and relapsing GPA and MPA. There were no differences in relapse rates at 6, 12, or 18 months; and there were no differences in adverse events.[77] At 24 months, rates of the composite outcome of death, end-stage renal disease and relapse did not differ between rituximab and cyclophosphamide, although B-cell return was associated with relapse in the rituximab arm.[78] Maintenance therapy with rituximab is a real possibility with the potential to provide long-term control and reduce relapse risk. The long-term consequences of repeat cycles of rituximab, however, are unexplored. The risks are hypogammaglobulinemia, which occurs in most cases, and the potential increase in incidence of infections. The most feared complication is the risk of reactivation of the John Cunningham (JC) virus leading to the complication of progressive multifocal leukoencephalopathy (PML), which has a very high mortality rate. The COVID-19 pandemic has reemphasized the need for the host to produce protective antibodies, raising the possibility that B-cell depleting agents need to be used with great caution.

Other Therapies

Belimumab continues to be investigated as a maintenance agent for AAV. It is a fully humanized IgG1γ mAb directed against soluble BLyS. Belimumab plus azathioprine and glucocorticoids for the maintenance of remission in AAV did not reduce the risk of relapse when compared against placebo.[79] Mepolizumab is a mAb directed against IL-5, which controls eosinophil production. Mepolizumab has been successfully used in the treatment of HES. In EGPA, Mepolizimab resulted in significantly higher levels of sustained remission in EGPA, compared to placebo, thus allowing for reduced glucocorticoid use.[80] IVIG has been available as a replacement therapy for patients with hypogammaglobulinemia for several years, and it is the standard of care

for KD, but its use in AAV has been limited. An initial study suggested relatively short-term benefit.[81] It has been suggested that IVIG can provide clinical benefit as a short-term adjunctive therapy in refractory or relapsing AAV, with rapid improvements in BVAS, ANCA levels, and CRP.[82,83] Plasmapheresis has been available for several decades. It is not entirely clear how it works; there are many theories suggesting the removal of circulating immune mediators is effective in reducing inflammation. Plasmapheresis has been widely used in the treatment of patients with severely impaired renal function and/or alveolar hemorrhage in AAV. It has been successful for rescue therapy in patients with very aggressive AAV or anti-GBM disease. The MEPEX trial demonstrated that it was able to reduce renal dysfunction in patients with severe AAV.[66] However, the long-term follow-up of the MEPEX trial demonstrated that the difference between patients treated with plasma exchange and those given methylprednisolone pulses (both as adjunct treatment alongside cyclophosphamide and steroid) did not last.[84] It has been suggested this is accounted for by the fact that many patients with severe renal disease had already developed irreversible changes to their kidneys and that renal dysfunction would have been secondary to damage rather than active disease. However, its use has recently been called into question as it failed to demonstrate a reduction in the incidence of death or end-stage renal failure in severe AAV. The same data did in fact also demonstrate non-inferiority using a reduced-dose glucocorticoid regimen compared to a standard-dose regimen with respect to death or end-stage renal failure.[2]

Plasmapheresis is effective in conjunction with antiviral therapy and steroids in the management of hepatitis B–related PAN.

Avacopan, an orally administered selective C5a receptor inhibitor, has shown promise in an early phase randomized placebo-controlled trial in effectively replacing high-dose glucocorticoids in the treatment of AAV.[69] A recent randomized, controlled trial comparing oral avacopan to oral prednisone on a tapering schedule in patients with AAV (where all patients received induction therapy with either cyclophosphamide or rituximab, followed by maintenance azathioprine). Avacopan was non-inferior, but not superior, to prednisone taper with respect to remission at week 26; and was superior to prednisone taper with respect to sustained remission at week 52. Encouragingly, the avacopan group required significantly less glucocorticoid treatment, and reduced glucocorticoid-associated toxicity, which points towards the future possibility of steroid-reduced, or perhaps even steroid-free, treatment of AAV.[85]

OUTCOMES

The majority of patients have a successful initial outcome. Either the condition is self-limiting in the case of more isolated forms of skin vasculitis, or the initial immunosuppressive therapy is successful. Over 94% of patients with generalized AAV would expect to survive the first 18 months,[66] whereas patients with more severe disease, especially with significant renal impairment, have a mortality of around 25% after 2 years.[66] This contrasts with over 80% likelihood of dying without adequate treatment. However, with current therapy, 5-year survival figures suggest around 25% to 30% mortality among AAV-affected patients.[66]

The bigger problem, however, is morbidity. The quality of survival for most patients with multiorgan disease is complicated by episodes of relapse in 50% to 70% of cases and low-grade grumbling disease, which never quite goes into full remission in up to a third of cases.[86] This is added to the comorbidity experienced by older patients, usually a combination of vasculitis-related damage, steroid-induced side effects, and the long-term consequences of immunosuppressive agents. In the first year of diagnosis, the most likely cause of mortality is active vasculitis or infection,[64] the latter being a surrogate measure of the severity in immunosuppressant therapy required to control the disease.

Long-term adverse outcomes in vasculitis can be measured using a structured VDI (see Assessment section above). One of the most important outcomes is the development of end-stage renal failure and the requirement for dialysis. It is likely that this is significantly reduced as a result of effective therapy given within the first 4 months of diagnosis. Transplantation is successful in patients with AAV, and these patients should be offered this treatment. Ten-year survival rates (32.5%) are similar to those reported for other patients without diabetes receiving a kidney transplant.[87] The immunosuppressive regimens used for maintaining the transplant (see Chapter 89) are often sufficient to keep the vasculitis in remission, but there is a need for ongoing review.

Infection is a significant concern, especially in the early course of disease when potent treatment is being commenced, especially high doses of steroids. The risk of serious infection requiring hospitalization is very high in the first year,[64] especially if the steroid doses remain high after 6 months. Interstitial lung disease (nonspecific interstitial pneumonia) is reported in around 20% of Japanese patients, especially those with MPA. This is a higher prevalence than that seen in other populations and may reflect genetic and environmental differences unique to Japan. Chronic neuropathy occurs in up to 65% of patients with AAV,[88] especially for patients with eosinophilic GPA and can be very distressing for patients. Upper airway disease generally continues to cause long-term problems in 65% of patients with GPA because of chronic mucosal damage causing symptoms of chronic nasal congestion, discharge, and discomfort.[89] Symptom relief is only partially successful in alleviating these problems. CVD among patients with small-vessel vasculitis is present in approximately 9% within 6 months of diagnosis[90] and four times greater in patients with AAV compared with the general population. The risk of cardiovascular events is about 14% within 5 years of diagnosis, especially in older patients who have baseline hypertension and MPO antibodies.[91] Cancer is associated with the presence of small-vessel vasculitis. It may predate as well as occur at the same time as diagnosis or develop subsequently, but it is recognized as a risk among patients treated with immunosuppressive and cytotoxic agents. Cancer of the bladder particularly has been an established risk arising from treatment with cyclophosphamide for many years; the original data from the 1970s suggested up to 33-fold increased risk of bladder cancer among patients treated with cyclophosphamide for vasculitis compared with background controls. However, this risk has been reduced with the use of more limited courses of cyclophosphamide (typically 3 to 6 months duration) and particularly with use of intermittent cyclophosphamide delivery. In a recent large series from the European Vasculitis Study Group (EUVAS), the only increased risk of cancer was for non-melanoma skin cancer, and this may also have reflected the use of azathioprine as well as use of cyclophosphamide.[92] Shang et al.[93] showed that in a meta-analysis of over 2500 patients,

the standardized incidence rate of late-occurring malignancies, particularly nonmelanoma skin cancer, leukemia, and bladder cancer, was 1.74 (95% CI = 1.37 to 2.21). We advise all of our patients to wear sun protection.

The ability to work is significantly affected by AAV; among 410 patients interviewed, 26% of those of working age were classified as work disabled.[94] The strongest influences on this outcome were fatigue, depression, high levels of damage (measured using the VDI), and being overweight. Patients' functional outcome can be variably affected by vasculitis and its treatment. Patients report impairment of function as measured by using generic tools, such as EQ-5D or the short form 36.[95] The impairment is similar to that found in other chronic diseases. Physical functions tend to be more affected than mental functions, especially in older patients with evidence of neurological involvement, usually peripheral neuropathy. Functional outcome is not directly correlated to disease activity, although in a Japanese cohort, 18 months after initiation of therapy, many aspects of function had started to improve.[96] One of the problems of determining long-term outcomes in patients with vasculitis is the compounding effect of the very intensive immunosuppressive therapy required to control disease. Over the last 3 decades, we have seen dramatic shift away from long-course cyclophosphamide toward short courses of intermittent dose therapy, but we are now witnessing an era when targeted biological therapies are able to take the place of cyclophosphamide. Therefore, eliminating the use of cyclophosphamide altogether in some patients may reset potential future outcomes. If this is coupled with a reduced use of glucocorticoid therapies and maintaining better disease control with less frequent relapses, the outcome may well be improved considerably for patients in the future.

ON THE HORIZON

- Improved understanding of the underlying pathogenic mechanisms in different forms of vasculitis is starting to lead to more logical, rational therapies, some of which have become established as standard of care or should be available in the near future.
- Rituximab is available and is being increasingly used in place of cyclophosphamide in the management of AAV. The role of rituximab in maintenance treatment, compared to azathioprine, is currently being explored, but reduced ability to produce protective antibodies may impose long-term risk for increased susceptibility to infection. Rituximab therapy is also effective in patients with cryoglobulinemic vasculitis, even when associated with the presence of hepatitis C. There is a concern that this might lead to reactivation of the virus and increase the risk of hepatoma or induce the development of the B-cell clone. However, better control of hepatitis C with effective antiviral agent is now possible.
- Multiple proinflammatory cytokines are being explored as therapeutic targets, although this will only allow for the suppression of common pathways in the end stage of inflammation.
- Mepolizumab has recently been shown to be effective in the treatment of EGPA in a large randomized controlled trial.
- Belimumab continues to remain under investigation, although when used in maintenance of remission in AAV (alongside azathioprine and glucocorticoids) it did not reduce the risk of relapse when compared against placebo.
- Blocking IL-6 may be of benefit for some patients.
- The discovery of the involvement of the C5 in AAV has led to trials demonstrating effective treatment of AAV with steroid-reduced or steroid-free induction-remission.
- The ultimate goal remains to identify upstream pathologies leading to vasculitis, which promises to fundamentally change the management of these devastating diseases, as it may permit the induction and maintenance of drug-free remission.

ACKNOWLEDGMENTS

The authors thank Jana Vaskova for administrative and secretarial support.

REFERENCES

1. Kawasaki T, Kosaki F, Okawa S, et al. A new infantile acute febrile mucocutaneous lymph node syndrome (MLNS) prevailing in Japan. *Pediatrics*. 1974;54(3):271–276.
2. Walsh M, Flossmann O, Berden A, et al. European Vasculitis Study Group. Risk factors for relapse of antineutrophil cytoplasmic antibody-associated vasculitis. *Arthritis Rheum*. 2012;64(2):542–548.
3. Walsh M, Merkel PA, Peh CA, et al. PEXIVAS Investigators. Plasma exchange and glucocorticoids in severe ANCA-associated vasculitis. *N Engl J Med*. 2020;382(7):622–631.
4. van der Woude FJ, Rasmussen N, Lobatto S, et al. Autoantibodies against neutrophils and monocytes: tool for diagnosis and marker of disease activity in Wegener's granulomatosis. *Lancet*. 1985;1(8426):425–429.
5. Andrews M, Edmunds M, Campbell A, et al. Systemic vasculitis in the 1980s-is there an increasing incidence of Wegener's granulomatosis and microscopic polyarteritis? *J R Coll Physicians Lond*. 1990;24(4):284–288.
6. Mohammad AJ. An update on the epidemiology of ANCA-associated vasculitis. *Rheumatol (Oxford)*. 2020;59:42–50.
7. Mahr A, Moosig F, Neumann T, et al. Eosinophilic granulomatosis with polyangiitis (Churg-Strauss): evolutions in classification, etiopathogenesis, assessment and management. *Curr Opin Rheumatol*. 2014;26(1):16–23.
8. Mohammad AJ, Jacobsson LTH, Mahr AD, et al. Prevalence of Wegener's granulomatosis, microscopic polyangiitis, polyarteritis nodosa and Churg-Strauss syndrome within a defined population in southern Sweden. *Rheumatology*. 2007;46(8):1329–1337.
9. Nesher G, Ben-Chetrit E, Mazal B, Breuer GS. The incidence of primary systemic vasculitis in Jerusalem: a 20-year hospital-based retrospective study. *J Rheumatol*. 2016;43(6):1072–1077.
10. Holman RC, Christensen KY, Belay ED, et al. Racial/ethnic differences in the incidence of Kawasaki syndrome among children in Hawaii. *Hawaii Med J*. 2010;69(8):194–197.
11. Arora A, Wetter DA, Gonzalez-Santiago TM, et al. Incidence of leukocytoclastic vasculitis, 1996 to 2010: a population-based study in Olmsted County, Minnesota. *Mayo Clin Proc*. 2014;89(11):1515–1524.
12. Watts RA, Scott DGI. Epidemiology of the vasculitides. *Semin Respir Crit Care Med*. 2004;25(5):455–464.
13. Xiao H, Heeringa P, Hu P, et al. Antineutrophil cytoplasmic autoantibodies specific for myeloperoxidase cause glomerulonephritis and vasculitis in mice. *J Clin Invest*. 2002;110(7):955–963.
14. Primo VC, Marusic S, Franklin CC, et al. Anti-PR3 immune responses induce segmental and necrotizing glomerulonephritis. *Clin Exp Immunol*. 2010;159(3):327–337.
15. Little MA, Al-Ani B, Ren S, et al. Anti-proteinase 3 anti-neutrophil cytoplasm autoantibodies recapitulate systemic vasculitis in mice with a humanized immune system. *PLoS One*. 2012;7(1):e28626.
16. Roth AJ, Ooi JD, Hess JJ, et al. Epitope specificity determines pathogenicity and detectability in ANCA-associated vasculitis. *J Clin Invest*. 2013;123(4):1773–1783.
17. Lyons PA, Rayner TF, Trivedi S, et al. Genetically distinct subsets within ANCA-associated vasculitis. *N Engl J Med*. 2012;367(3):214–223.
18. McKinney EF, Willcocks LC, Broecker V, Smith KG. The immunopathology of ANCA-associated vasculitis. *Semin Immunopathol*. 2014;36(4):461–478. https://doi.org/10.1007/s00281-014-0436-6.
19. Li W, Huang H, Cai M, Yuan T, Sheng Y. Antineutrophil cytoplasmic antibody-associated vasculitis update: genetic pathogenesis. *Front Immunol*. 2021;12:624848.
20. Langford CA, Monach PA, Specks U, et al. An open-label trial of abatacept (CTLA4-IG) in non-severe relapsing granulomatosis with polyangiitis (Wegener's). *Ann Rheum Dis*. 2014;73(7):1376–1379.
21. Wieczorek S, Hellmich B, Arning L, et al. Functionally relevant variations of the interleukin-10 gene associated with antineutrophil cytoplasmic

antibody-negative Churg-Strauss syndrome, but not with Wegener's granulomatosis. *Arthritis Rheum*. 2008;58(6):1839–1848.
22. Ciavatta DJ, Yang J, Preston GA, et al. Epigenetic basis for aberrant upregulation of autoantigen genes in humans with ANCA vasculitis. *J Clin Invest*. 2010;120(9):3209–3219.
23. Boudigaard SH, Schlünssen V, Vestergaard JM, et al. Occupational exposure to respirable crystalline silica and risk of autoimmune rheumatic diseases: a nationwide cohort study. *Int J Epidemiol*. 2021;50(4):1213–1226. https://doi.org/10.1093/ije/dyaa287.
24. Rihova Z, Maixnerova D, Jancova E, et al. Silica and asbestos exposure in ANCA-associated vasculitis with pulmonary involvement. *Ren Fail*. 2005;27(5):605–608.
25. Glasner C, de Goffau MC, van Timmeren MM, et al. Genetic loci of Staphylococcus aureus associated with anti-neutrophil cytoplasmic autoantibody (ANCA)-associated vasculitides. *Sci Rep*. 2017;7(1):12211.
26. Eden A, Mahr A, Servant A, et al. Lack of association between B19 or V9 erythrovirus infection and ANCA-positive vasculitides: a case-control study. *Rheumatol (Oxford)*. 2003;42(5):660–664.
27. Wang CR, Tsai HW. Human hepatitis viruses-associated cutaneous and systemic vasculitis. *World J Gastroenterol*. 2021;27(1):19–36.
28. Kain R, Exner M, Brandes R, et al. Molecular mimicry in pauci-immune focal necrotizing glomerulonephritis. *Nat Med*. 2008;14(10):1088–1096.
29. Gibson KM, Kain R, Luqmani RA, et al. Autoantibodies against lysosome associated membrane protein-2 (LAMP-2) in pediatric chronic primary systemic vasculitis. *Front Immunol*. 2021;11:624758.
30. Hurtado PR, Jeffs L, Nitschke J, et al. CpG oligodeoxynucleotide stimulates production of anti-neutrophil cytoplasmic antibodies in ANCA associated vasculitis. *BMC Immunol*. 2008;9:34.
31. Noh JY, Yasuda S, Sato S, et al. Clinical characteristics of myeloperoxidase antineutrophil cytoplasmic antibody-associated vasculitis caused by antithyroid drugs. *J Clin Endocrinol Metab*. 2009;94(8):2806–2811.
32. Waldhauser L, Uetrecht J. Oxidation of propylthiouracil to reactive metabolites by activated neutrophils. Implications for agranulocytosis. *Drug Metab Dispos*. 1991;19(2):354–359.
33. Nakazawa D, et al. Abnormal conformation and impaired degradation of propylthiouracil-induced neutrophil extracellular traps: implications of disordered neutrophil extracellular traps in a rat model of myeloperoxidase antineutrophil cytoplasmic antibody-associated vasculitis. *Arthritis Rheum*. 2012;64(11):3779–3787.
34. Cascio MJ, Jen KY. Cocaine/levamisole-associated autoimmune syndrome: a disease of neutrophil-mediated autoimmunity. *Curr Opin Hematol*. 2018;25(1):29–36.
35. Marquez J, Aguirre L, Muñoz C, et al. Cocaine-levamisole-induced vasculitis/vasculopathy syndrome. *Curr Rheumatol Rep*. 2017;19(6):36.
36. Wiesner O, Russell KA, Lee AS, et al. Antineutrophil cytoplasmic antibodies reacting with human neutrophil elastase as a diagnostic marker for cocaine-induced midline destructive lesions but not autoimmune vasculitis. *Arthritis Rheum*. 2004;50(9):2954–2965.
37. Tsurikisawa N, Saito H, Oshikata C, et al. Decreases in the numbers of peripheral blood regulatory T cells, and increases in the levels of memory and activated B cells, in patients with active eosinophilic granulomatosis and polyangiitis. *J Clin Immunol*. 2013;33(5):965–976.
38. Abdulahad WH, Boots AMH, Kallenberg CGM. FoxP3(+) CD4(+) T cells in systemic autoimmune diseases: the delicate balance between true regulatory T cells and effector Th-17 cells. *Rheumatology*. 2011;50(4):646–656.
39. Marinaki S, Kälsch A-I, Grimminger P, et al. Persistent T-cell activation and clinical correlations in patients with ANCA-associated systemic vasculitis. *Nephrol Dial Transpl*. 2006;21(7):1825–1832.
40. Holden NJ, Williams JM, Morgan MD, et al. ANCA-stimulated neutrophils release BLyS and promote B cell survival: a clinically relevant cellular process. *Ann Rheum Dis*. 2011;70(12):2229–2233.
41. Nakazawa D, Shida H, Tomaru U, et al. Enhanced formation and disordered regulation of NETs in myeloperoxidase-ANCA-associated microscopic polyangiitis. *J Am Soc Nephrol*. 2014;25(5):990–997.
42. Yuan J, Gou S-J, Huang J, et al. C5a and its receptors in human anti-neutrophil cytoplasmic antibody (ANCA)-associated vasculitis. *Arthritis Res Ther*. 2012;14(3):R140.
43. Chen SF, Wang F-M, Li Z-Y, et al. Plasma complement factor H is associated with disease activity of patients with ANCA-associated vasculitis. *Arthritis Res Ther*. 2015;17(1):129.
44. Fukui S, Iwamoto N, Umeda M, et al. Antineutrophilic cytoplasmic antibody-associated vasculitis with hypocomplementemia has a higher incidence of serious organ damage and a poor prognosis. *Med (Baltimore)*. 2016;95(37):e4871.
45. Takahashi K, Oharaseki T, Yokouchi Y. Update on etio and immunopathogenesis of Kawasaki disease. *Curr Opin Rheumatol*. 2014;26(1):31–36.
46. Lin YJ, Chang J-S, Liu X, et al. Genetic variants in PLCB4/PLCB1 as susceptibility loci for coronary artery aneurysm formation in Kawasaki disease in Han Chinese in Taiwan. *Sci Rep*. 2015;5:14762.
47. Onouchi Y, Suzuki Y, Suzuki H, et al. ITPKC and CASP3 polymorphisms and risks for IVIG unresponsiveness and coronary artery lesion formation in Kawasaki disease. *Pharmacogenomics J*. 2013;13(1):52–59.
48. Gorelik M, Wilson DC, Cloonan YK, et al. Plasma follistatin-like protein 1 is elevated in Kawasaki disease and may predict coronary artery aneurysm formation. *J Pediatr*. 2012;161(1):116–119.
49. Lin IC, Kuo H-C, Lin Y-J, et al. Augmented TLR2 expression on monocytes in both human Kawasaki disease and a mouse model of coronary arteritis. *PLoS One*. 2012;7(6):e38635.
50. Jia S, Li C, Wang G, et al. The T helper type 17/regulatory T cell imbalance in patients with acute Kawasaki disease. *Clin Exp Immunol*. 2010;162(1):131–137.
51. Rowley AH, Wylie KM, Kim KY, et al. The transcriptional profile of coronary arteritis in Kawasaki disease. *BMC Genomics*. 2015;16:1076.
52. Leung DY, Meissner HC, Shulman ST, et al. Prevalence of superantigen-secreting bacteria in patients with Kawasaki disease. *J Pediatr*. 2002;140(6):742–746.
53. Schulte DJ, Yilmaz A, Shimada K, et al. Involvement of innate and adaptive immunity in a murine model of coronary arteritis mimicking Kawasaki disease. *J Immunol*. 2009;183(8):5311–5318.
54. Cid MC, Grau JM, Casademont J, et al. Immunohistochemical characterization of inflammatory cells and immunologic activation markers in muscle and nerve biopsy specimens from patients with systemic polyarteritis nodosa. *Arthritis Rheum*. 1994;37(7):1055–1061.
55. Yalcinkaya F, Ozçakar ZB, Kasapçopur O, et al. Prevalence of the MEFV gene mutations in childhood polyarteritis nodosa. *J Pediatr*. 2007;151(6):675–678.
56. Sansonno D, Carbone A, Re VD, et al. Hepatitis C virus infection, cryoglobulinaemia, and beyond. *Rheumatology*. 2007;46(4):572–578.
57. Varricchi G, Bagnasco D, Borriello F, et al. Interleukin-5 pathway inhibition in the treatment of eosinophilic respiratory disorders: evidence and unmet needs. *Curr Opin Allergy Clin Immunol*. 2016;16(2):186–200.
58. Craven A, Robson J, Ponte C, et al. ACR/EULAR-endorsed study to develop diagnostic and classification criteria for vasculitis (DCVAS). *Clin Exp Nephrol*. 2013;17(5):619–621.
59. Luqmani R, Ponte C. ANCA associated vasculitides and polyarteritis nodosa. In: Hachulla E, Bijlsma JWJ, eds. *EULAR Textbook on Rheumatic Diseases*. London: BMJ Publishing Group Ltd; 2015:717–753.
60. Oliva-Damaso N, Bomback AS. Proposal for a more practical classification of antineutrophil cytoplasmic antibody-associated vasculitis. *Clin Kidney J*. 2020:1–8. https://doi.org/10.1093/ckj/sfaa255.
61. Stahelin L, Fialho SCM, Neves FS, et al. Cocaine-induced midline destruction lesions with positive ANCA test mimicking Wegener's granulomatosis. *Rev Bras Reumatol*. 2012;52(3):431–437.
62. Berman M, Paran D, Elkayam O. Cocaine-Induced Vasculitis. *Rambam Maimonides Med J*. 2016 Oct 31;7(4):e0036.
63. Karim AF, Verdijk RM, Nagtegaal AP, et al. To distinguish IgG4-related disease from seronegative granulomatosis with polyangiitis. *Rheumatology*. 2017;56(12):2245–2247.
64. Flossmann O, Berden A, de Groot K, et al. Long-term patient survival in ANCA-associated vasculitis. *Ann Rheum Dis*. 2011;70(3):488–494.
65. Kemna MJ, Damoiseaux J, Austen J, et al. ANCA as a predictor of relapse: useful in patients with renal involvement but not in patients with nonrenal disease. *J Am Soc Nephrol*. 2015;26(3):537–542.
66. Yates M, Watts RA, Bajema IM, et al. EULAR/ERA-EDTA recommendations for the management of ANCA-associated vasculitis. *Ann Rheum Dis*. 2016;75(9):1583–1594.

67. Ponte C, Sznajd J, O'Neill L, et al. Optimisation of vasculitis disease assessments in clinical trials, clinical care and long-term databases. *Clin Exp Rheumatol*. 2014;32(5 Suppl 85):S118–S125.
68. King C, Harper L. Avoidance of harm from treatment for ANCA-associated vasculitis. *Curr Treatm Opt Rheumatol*. 2017;3(4):230–243.
69. Jayne DRW, Bruchfeld AN, Harper L, et al. CLEAR Study Group. Randomized trial of C5a receptor inhibitor avacopan in ANCA-associated vasculitis. *J Am Soc Nephrol*. 2017;28(9):2756–2767.
70. Guillevin L, Pagnoux C, Karras A, et al. French Vasculitis Study Group. Rituximab versus azathioprine for maintenance in ANCA-associated vasculitis. *N Engl J Med*. 2014;371(19):1771–1780.
71. Smith R, Jayne D, Merkel PA,. on behalf of RITAZAREM Investigators. OP0026 A randomized, controlled trial of rituximab versus azathioprine after induction of remission with rituximab for patients with ANCA-associated vasculitis and relapsing disease. *Ann Rheum Dis*. 2020;79:19–20.
72. Karras A, Pagnoux C, Haubitz M, et al. European Vasculitis Society. Randomised controlled trial of prolonged treatment in the remission phase of ANCA-associated vasculitis. *Ann Rheum Dis*. 2017;76(10):1662–1668.
73. Maritati F, Alberici F, Oliva E, et al. Methotrexate versus cyclophosphamide for remission maintenance in ANCA-associated vasculitis: a randomised trial. *PLoS One*. 2017;12(10):e0185880.
74. Tuin J, Stassen PM, Bogdan DI, et al. Mycophenolate mofetil versus cyclophosphamide for the induction of remission in nonlife-threatening relapses of antineutrophil cytoplasmic antibody-associated vasculitis: randomized, controlled trial. *Clin J Am Soc Nephrol*. 2019;14(7):1021–1028.
75. Jones RB, Hiemstra TF, Ballarin J, et al. European Vasculitis Study Group (EUVAS). Mycophenolate mofetil versus cyclophosphamide for remission induction in ANCA-associated vasculitis: a randomised, non-inferiority trial. *Ann Rheum Dis*. 2019;78(3):399–405.
76. Szpirt WM, Heaf JG, Petersen J. Plasma exchange for induction and cyclosporine A for maintenance of remission in Wegener's granulomatosis--a clinical randomized controlled trial. *Nephrol Dial Transpl*. 2011;26(1):206–213.
77. Geetha D, Specks U, Stone JH, et al. Rituximab for ANCA-Associated Vasculitis Immune Tolerance Network Research Group. Rituximab versus cyclophosphamide for ANCA-associated vasculitis with renal involvement. *J Am Soc Nephrol*. 2015;26(4):976–985.
78. Jones RB, Furuta S, Tervaert JW, et al. European Vasculitis Society (EUVAS). Rituximab versus cyclophosphamide in ANCA-associated renal vasculitis: 2-year results of a randomised trial. *Ann Rheum Dis*. 2015;74(6):1178–1182.
79. Jayne D, Blockmans D, Luqmani R, et al. Efficacy and safety of belimumab and azathioprine for maintenance of remission in antineutrophil cytoplasmic antibody-associated vasculitis: a randomized controlled study. *Arthritis Rheumatol*. 2019;71(6):952–963.
80. Wechsler ME, Akuthota P, Jayne D, et al. EGPA Mepolizumab Study Team. Mepolizumab or Placebo for Eosinophilic Granulomatosis with Polyangiitis. *N Engl J Med*. 2017;376(20):1921–1932.
81. Jayne DR, Chapel H, Adu D, et al. Intravenous immunoglobulin for ANCA-associated systemic vasculitis with persistent disease activity. *QJM*. 2000;93(7):433–983.
82. Shimizu T, Morita T, Kumanogoh A. The therapeutic efficacy of intravenous immunoglobulin in anti-neutrophilic cytoplasmic antibody-associated vasculitis: a meta-analysis. *Rheumatol (Oxford)*. 2020;59(5):959–967.
83. Crickx E, Machelart I, Lazaro E, et al. French Vasculitis Study Group. Intravenous immunoglobulin as an immunomodulating agent in antineutrophil cytoplasmic antibody-associated vasculitides: a French nationwide study of ninety-two patients. *Arthritis Rheumatol*. 2016;68(3):702–712.
84. Walsh M, Casian A, Flossmann O, et al. Long-term follow-up of patients with severe ANCA-associated vasculitis comparing plasma exchange to intravenous methylprednisolone treatment is unclear. *Kidney Int*. 2013;84(2):397–402.
85. Jayne DRW, Merkel PA, Schall TJ, Bekker P. ADVOCATE Study Group. Avacopan for the treatment of ANCA-associated vasculitis. *N Engl J Med*. 2021;384(7):599–609.
86. Hoffman GS, Kerr G, Leavitt RY, et al. Wegener granulomatosis: an analysis of 158 patients. *Ann Intern Med*. 1992;116:488–498.
87. Hruskova Z, Stel V, Jayne D, et al. Characteristics and outcomes of granulomatosis with polyangiitis (Wegener) and microscopic polyangiitis requiring renal replacement therapy: results from the European Renal Association-European Dialysis and Transplant Association Registry. *Am J Kidney Dis*. 2015;66(4):613–620.
88. Bischof A, Jaeger VK, Hadden RDM, et al. Peripheral neuropathy in antineutrophil cytoplasmic antibody-associated vasculitides: insights from the DCVAS study. *Neurol Neuroimmunol Neuroinflamm*. 2019;6(6):e615 20.
89. Martinez Del Pero M, Walsh M, Luqmani R, et al. Long-term damage to the ENT system in Wegener's granulomatosis. *Eur Arch Otorhinolaryngol*. 2011;268(5):733–739.
90. Monti S, Robson J, Klersy C, et al. Early development of new cardiovascular risk factors in the systemic vasculitides. *Clin Exp Rheumatol*. 2020;38(Suppl 124(2)):126–134.
91. Suppiah R, Judge A, Batra R, et al. A model to predict cardiovascular events in patients with newly diagnosed Wegener's granulomatosis and microscopic polyangiitis. *Arthritis Care Res (Hoboken)*. 2011;63(4):588–596.
92. Heijl C, Harper L, Flossmann O, et al. European Vasculitis Study Group (EUVAS). Incidence of malignancy in patients treated for antineutrophil cytoplasm antibody-associated vasculitis: follow-up data from European Vasculitis Study Group clinical trials. *Ann Rheum Dis*. 2011;70(8):1415–1421.
93. Shang W, Ning Y, Xu X, et al. Incidence of cancer in ANCA-associated vasculitis: a meta-analysis of observational studies. *PLoS One*. 2015;10(5):e0126016.
94. Basu N, McClean A, Harper L, et al. Markers for work disability in antineutrophil cytoplasmic antibody-associated vasculitis. *Rheumatol (Oxf)*. 2014;53(5):953–956.
95. Walsh M, Mukhtyar C, Mahr A, et al. Health-related quality of life in patients with newly diagnosed antineutrophil cytoplasmic antibody-associated vasculitis. *Arthritis Care Res (Hoboken)*. 2011;63(7):1055–1061.
96. Suka M, Hayashi T, Kobayashi S, et al. Improvement in health-related quality of life in MPO-ANCA-associated vasculitis patients treated with cyclophosphamide plus prednisolone: an analysis of 18 months of follow-up data from the JMAAV study. *Mod Rheumatol*. 2012;22(6):877–884.

Large Vessel Vasculitides

Cornelia M. Weyand

Most tissues have compensatory mechanisms that allow them to sustain the damaging effects of acute and chronic inflammation, but medium and large arteries are organs without redundancy and limited regenerative capacity. Life is unsustainable unless the major arteries have uncompromised function. Accordingly, autoimmune and autoinflammatory damage to such arterial vessels leads to severe clinical consequences, immediately posing a threat for the loss of function of vital organs. When affected by inflammation, the aorta and its branches have two possible response patterns: (i) Damage to the vessel wall leads to dilatation, aneurysm formation, and rupture. Alternatively, the wall layers dissect. (ii) The inflammation initiates a maladaptive response to injury, resulting in vasoocclusion, disruption of blood supply, and organ ischemia.

In contrast to other vasculopathies, especially those related to atherosclerosis, vasculitides of the large muscular and elastic arteries are almost always associated with a syndrome of intense systemic inflammation.[1] Systemic inflammation is no longer considered a spillover of inflammatory mediators from the vasculitic lesions. Instead, systemic activation of the innate immune system appears to be a pinnacle event that initiates the processes leading to vessel wall inflammation. The coincidence of malaise, fever, wasting, and myalgias, with signs of ischemia caused by vascular failure, remains a critical clue for the physician when diagnosing and treating large vessel vasculitis (LVV).

The two major forms of LVVs are giant cell arteritis (GCA) and Takayasu arteritis (TA). In addition, aortitis can infrequently be seen in other diseases, such as infections, connective tissue diseases, sarcoidosis, and inflammatory bowel disease (IBD), and occasionally is diagnosed as an idiopathic syndrome. It is now recognized that aortitis can occur as a side effect of checkpoint inhibitor immunotherapy. Polymyalgia rheumatica (PMR) is a condition closely related to GCA; it occurs in the same patient population and often precedes or follows the clinical diagnosis of GCA.[2] Patients with PMR do not have typical vascular lesions; consequently, PMR is a vasculitic forme frust. Patients with PMR do have a systemic inflammatory syndrome indistinguishable from GCA, and approximately 10% of them eventually progress to full-blown vasculitis. Similarities in the vascular lesions of GCA and TA have been interpreted as revealing parallels in immunopathogenesis. More recent studies emphasize dissimilarities between both LVVs, including disease-specific autoantibodies in TA patients and a role of cytolytic CD8 T cells and natural killer (NK) cells in TA pathogenesis. Whether the systemic inflammatory reactions accompanying GCA, TA, and PMR have disease-specific elements remains unanswered, but this has opened the possibility of developing biomarkers that are urgently needed for clinical monitoring. Excellent progress has been made in unraveling the pathogenesis of GCA, and this will inevitably lead to improvements in diagnosis, long-term management, and broadening of the therapeutic armamentarium.

EPIDEMIOLOGY

GCA may be a very old disease, as suggested by historic evidence that more than 1000 years ago temporal artery removal was recommended by a physician in Baghdad. In 1932, Horton et al. at the Mayo Clinic in Minnesota recognized that GCA was a vasculitis based on dense temporal artery inflammation of two patients who were systemically ill and had severe headaches. The first reports of TA, or "pulseless disease," in young women surfaced in Japan in the 19th century. The syndrome was named after Dr. Takayasu, an ophthalmologist, who, in 1905, described peculiar optic fundus abnormalities caused by ischemia-driven collateral formation.

The strongest risk factor for GCA, TA, and PMR is age.[3] GCA and PMR are essentially absent in individuals younger than 50 years of age, and their incidence climbs continuously during the seventh and eighth decades of life. TA is almost exclusively diagnosed in individuals younger than 40 years of age, with peak incidence during the second and third decades of life. Women are affected more often by all three syndromes compared with men, with a 2:1 ratio in PMR and GCA and a 9:1 ratio in TA.[3,4]

Marked geographic variations in the incidence and prevalence of GCA, TA, and PMR have encouraged speculations that environmental exposures are key determinants in disease pathogenesis. GCA is the most frequent vasculitis in the Western world, with yearly incidence rates reaching 10 to 20 cases per 100,000 persons older than 50 years of age.[3] In general, PMR is diagnosed threefold to fourfold more frequently, with a prevalence of up to 1 case per 133 individuals older than 50 of age.[2] Iceland, Norway, Sweden, and Denmark are high-risk areas; higher incidence rates are also seen in Scandinavian immigrant populations in the United States. The risk is significantly lower in Hispanics and African Americans. Although TA can afflict all races, a predilection exists for individuals of Asian and Central and South American origins. Japan, Thailand, India, Turkey, and Central and South America are considered high-incidence regions. TA is a rare disease, with an annual incidence of 1 to 2 cases/million. The typical patient is a female in her 20 s to 30 s. In middle-aged men and women, it can be challenging to differentiate TA from rapidly progressing atherosclerotic disease, especially as both disease processes may coexist.

ETIOLOGY AND PATHOGENESIS

Abnormal innate and adaptive immune responses are essential pathogenic elements in medium-vessel vasculitides and LVVs. The resulting disease process separates GCA, PMR, and TA

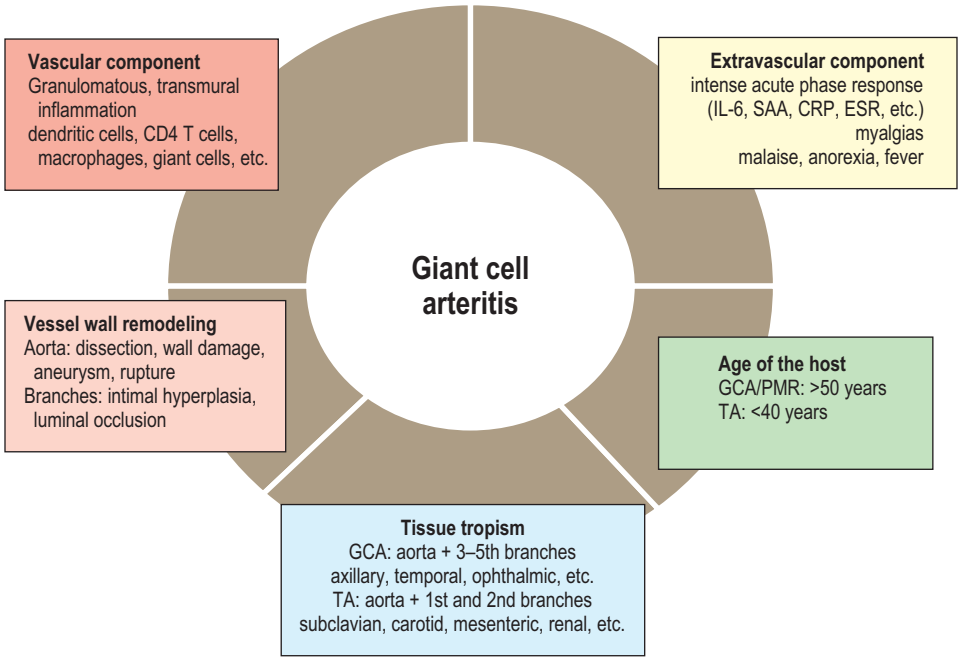


FIG. 60.1 Key Pathogenic Traits in Giant Cell Arteritis. Five hallmarks of disease represent fundamental pathogenic traits and distinguish giant cell arteritis *(GCA)* from other immune-mediated disorders. Vascular GCA (granulomatous vasculitis) is separated from extravascular GCA (systemic inflammation and intense hepatic acute phase response). Age remains the strongest risk factor and patients with GCA have a signature of immunosenescence. The stringent tissue tropism for selected vascular beds is suspected to reflect determinants of the tissue niche. Clinical complications are related to the patterning of vascular damage, spanning from wall destruction to luminal occlusion. *CRP*, C-reactive protein; *ESR*, erythrocyte sedimentation rate; *IL-6*, interleukin-6; *PMR*, polymyalgia rheumatica; *SAA*, serum amyloid A; *TA*, Takayasu arteritis.

from other vasculitides and from other autoinflammatory and autoimmune syndromes. Recent work has delineated disease-specific signatures, with the goal to mark the pinnacle abnormality triggering the disease and to discover intersection points that lend themselves for diagnostic and therapeutic purposes.

Pathogenic studies and careful clinical observation have prompted a substantial shift in the pathogenic model (Fig. 60.1): LVVs are now understood to be chronic conditions that target restricted vascular beds and have two major disease components: (i) granulomatous, intramural inflammation inducing vascular wall remodeling; and (ii) extravascular inflammation manifesting with an intense acute phase response, myalgias, and constitutional symptoms. Emerging data suggest at least partial autonomy in the pathogenic cascades of both disease components, predicting separate biomarkers, separate pathogenic mechanisms, and separate responses to immunosuppression. Similarities in tissue tropisms and histologic lesions of GCA and TA suggest some overlapping disease pathways, but abnormalities in cytolytic CD8 T cells and NK cells may be more relevant in TA. The etiopathogenesis of PMR is less well understood, but experimental evidence suggests that it represents a forme fruste of GCA, in which inflammatory attack to the vessel wall remains below a threshold, and standard histology describes noninflamed arteries.

Distinctive features—hallmarks of disease—provide clues to the fundamental pathogenic mechanisms (see Fig. 60.1), including a stringent age cutoff, a stringent tissue tropism for selected vascular territories (aorta and major branches), an extravascular disease module defined by an abrupt and intense acute phase response, a granulomatous transmural vasculitis, and two alternate patterns of vessel wall remodeling causing aortic wall destruction versus luminal occlusion in the branch vessels.

INNATE IMMUNE SYSTEM DEFECTS

Innate immune cells make critical contributions to the pathogenesis of LVVs, but where circulating monocytes and neutrophils encounter their activating stimuli remains unknown (Fig. 60.2). Circulating monocytes and macrophages are highly activated[5] and contribute to the array of proinflammatory cytokines measurably elevated in the serum of affected patients.[6] Altered neutrophil functionality has also been described.[7] Early in the disease process, circulating neutrophils resemble those induced by a lipopolysaccharide stimulus, displaying T-cell suppressive functions. Interleukin-8 (IL-8) and IL-6 are associated with neutrophil activation. IL-6, first described in the early 1990s, is highly elevated in GCA and PMR[8] and is explicitly steroid sensitive.[9] IL-6 acts as an inducer of the acute phase response in the liver, triggering production of C-reactive protein (CRP), serum amyloid A (SAA), and multiple other acute phase proteins. Accordingly, treatment with an IL-6 receptor–blocking antibody effectively reduces the laboratory abnormalities measured in patients with GCA.[10] Whether IL-6 has additional functions in the disease process, particularly in the inflamed vessel wall, is uncertain. Some acute phase proteins (*e.g.*, SAA) function as proinflammatory amplifiers. Although abnormal innate immunity dominates the extravascular component of LVVs, underlying mechanisms (*e.g.*, original triggers, site of activation, interface of systemic and vascular inflammation) are essentially unexplored.

Defects in innate immunity contributing to the early and late events in the vasculitic lesions are much better understood (Fig. 60.3).[11] A population of vessel wall–residing dendritic cells (vasDCs) have been identified in normal human arteries[12] localized at the adventitial–medial junction, close to the adventitial

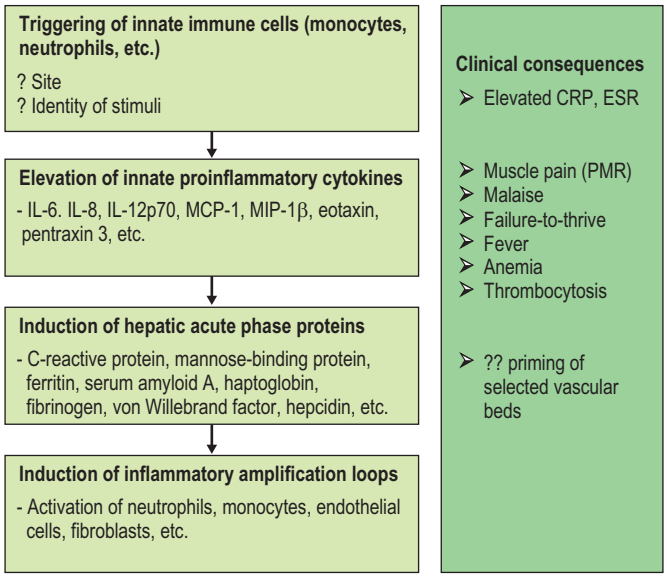


FIG. 60.2 Extravascular Giant Cell Arteritis. Circulating innate cells (monocytes, neutrophils, etc.) are highly activated and elicit a hepatic acute phase response. Acute phase proteins serve as biomarkers (C-reactive protein *[CRP]*, erythrocyte sedimentation rate *[ESR]*) of disease. Extravascular giant cell arteritis (GCA) is explicitly sensitive to corticosteroid therapy and to cytokine blockade. Whether the extravascular competent has a direct or indirect involvement in driving/sustaining granulomatous vasculitis remains unknown. Polymyalgia rheumatica *(PMR)* represents an isolated form of extravascular GCA. *IL*, Interleukin.

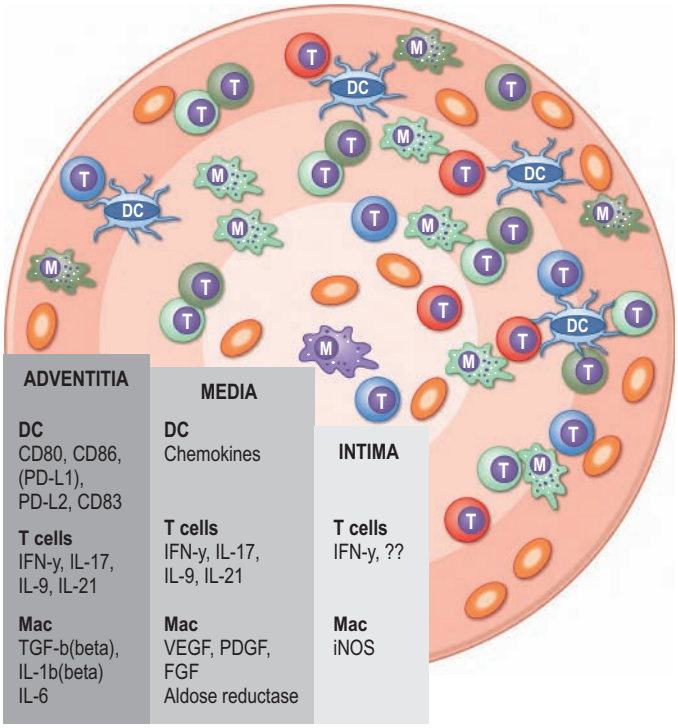


FIG. 60.3 Vascular Giant Cell Arteritis. T cells, macrophages, and dendritic cells *(DCs)* form granulomatous infiltrates in the vessel wall. DCs, present in healthy arteries at the adventitia-media border, serve as antigen-presenting cells and provide costimulatory signals but are deficient in the coinhibitory ligand programmed death ligand 1 *(PD-L1)*. Wall-infiltrating T cells are multifunctional, producing a broad spectrum of effector cytokines, including interferon *(IFN)-γ*, interleukin *(IL)*-17, IL-9, and IL-21. Macrophages, some of which transform into multinucleated giant cells, supply a multitude of cytokines, metalloproteinases, reactive oxygen species, growth factors, angiogenesis factors, and inducible nitric oxide synthase *(iNOS)*. Their functional commitment depends on their geographic location in the vessel wall. *FGF*, Fibroblast growth factor; *PDGF*, Platelet-derived growth factor*TGF-β*, Transforming growth factor-β; *VEGF*, Vascular endothelial growth factor.

microvascular network. Given their strategic positioning, they are believed to guard the artery's "backdoor." vasDCs serve a gatekeeper function in vasculitis, and stimulation of such vasDCs is a prerequisite to break the natural immunotolerance (immune privilege) of the arterial wall. Their activation presages vasculitis, as indicated by their transition from the resting state to the activated state in PMR arteries, where no vessel wall infiltrates are as yet detectable.[13] In frank vasculitis, vasDCs undergo expansion, penetrate deep into the wall, produce chemokines and cytokines, and express costimulatory ligands.[14] They may present local antigen to induce T-cell clonal expansion and, through their chemokine production, shape the composition of the intramural granulomatous infiltrates. Considering that vital arteries are protected by immune privilege, malfunction of endogenous vasDCs may, indeed, be the pinnacle defect-initiating vasculitis.

Other innate immune cells sequestered in the vasculitic lesions, specifically macrophages, are key inflammatory effector cells.[15] Multiple functional domains of macrophages have pathogenic relevance. The functional commitment of lesional macrophages is closely linked to their geographic location in the tissue site.[16] Intima-positioned macrophages produce inducible nitric oxide synthase, regulating vascular tone. Adventitial macrophages secrete cytokines (IL-1β, IL-6, transforming growth factor-β [TGF-β]), conditioning the local inflammatory environment. Granuloma formation and giant cells occur mainly in the media and at the media–intima border. Functional profiling of media-residing macrophages has connected them to tissue damage; they produce matrix metalloproteinases (MMP) and reactive oxygen species (ROS), provide antioxidant regulation, and supply growth factors for myofibroblasts and newly formed microvessels. Essentially, they drive the hyperplasia of the intimal layer that leads to luminal obstruction and tissue ischemia. Whether macrophage effector functions are fundamentally different in aortitis, compared with branch vasculitis, is unknown. Differences of tissue damage patterns (wall destruction versus luminal occlusion) suggest at least substantial variances (Fig. 60.4).

ADAPTIVE IMMUNE SYSTEM DEFECTS

The hallmark lesions in the vessel wall are granulomatous infiltrates, a mixture of highly activated macrophages, giant cells, and surrounding lymphocytes (see Fig. 60.3). The overwhelming majority of these lymphocytes are memory CD4 T cells. CD8 T cells are infrequent, and B cells are rare. Multinucleated giant cells are present in approximately 50% of patients, often localized at the intima–media border adjacent to the lamina elastica interna. Fragmentation of that elastic membrane is a hallmark of GCA.

Analysis of clonal CD4 populations has yielded identical T-cell receptors (TCRs) in independent temporal artery biopsies from the same patient, strongly supportive for antigen-dependent T-cell expansion.[17] Wall-embedded vasDCs serve

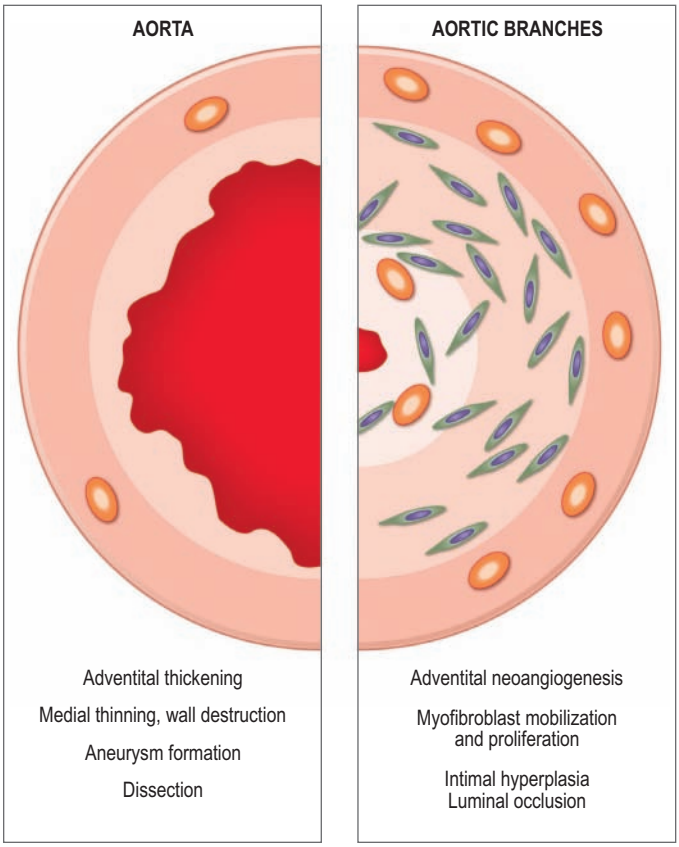


FIG. 60.4 Patterns of Vessel Wall Damage in Giant Cell Arteritis. Transmural granulomatous vasculitis results in two distinct damage patterns. Aortic inflammatory infiltrates destroy the wall, inducing dissection and aneurysm formation. As a result of neoangiogenesis and fibrosis, the adventitia expands. In the aortic branches, transmural granulomatous infiltrates predominantly elicit a maladaptive healing response, characterized by the mobilization, migration, and proliferation of myofibroblasts forming hyperplastic and lumen-obstructive neointima. Formation of neotissue is associated with microvascular neoangiogenesis in the intima, media, and adventitia. Intimal hyperplasia and neoangiogenesis are ultimately regulated through the intensity of the T-cell response in the vasculitic lesions.

as antigen-presenting cells (APCs); the nature of the antigen remains unknown, and reports about infectious agents have remained uncorroborated. Carefully designed experiments need to probe whether pathogens isolated from temporal artery tissue represent the vascular microbiota and whether they have a pathogenic role.

T helper 1 (Th1) cells are a key pathogenic element in GCA. Interferon-γ (IFN-γ), the Th1 marker cytokine, has multiple disease-relevant functions: activating endothelial cells and macrophages, regulating wall remodeling, inducing microvascular neoangiogenesis, and driving intimal-layer expansion.[11] IFN-γ is predominantly produced in the adventitia, where T cells are instructed by vasDCs. Th1 cells are present in untreated patients and persist despite prolonged corticosteroid therapy,[18] suggesting independence of vasculitogenic Th1 cells from the steroid-sensitive acute phase response and emphasizing their role as stabilizers of chronic-persistent lesions. During early vasculitis, Th1 cells are accompanied by Th17 cells, which are highly sensitive to steroid therapy and depleted from chronic lesions.[18] Comparative analysis of sequential temporal artery biopsies harvested from patients before and after treatment has demonstrated that adaptive immune responses evolve over the course of the disease and that different subpopulations of lesion-residing T cells are differentially responsive to immunosuppression.

Recent data have emphasized that the T-cell infiltrate in GCA-affected arteries is typically multifunctional and composed of a broad spectrum of functional lineages.[19] Besides Th1 and Th17 cells, Th9, Th22, and IL-21–producing T follicular helper (Tfh) cells are part of the infiltrate. Little is known about precise effector functions of these T cells, but lesional T-cell populations are functionally highly diverse, raising the possibility that a mixed set of antigens has vasculitogenic potential. Alternatively, the defect leading to granulomatous wall inflammation is the sequel of antigen-nonspecific defects.[19]

T cells have now been placed at the top of the artery's maladaptive response to injury, but precise effector pathways are not understood (see Fig. 60.4). Chronic aortitis leads to wall destruction and aneurysm formation. Dissection is increasingly recognized as a sequel of aortic wall inflammation. In rare cases, LVV results in the fatal complication of aortic rupture. In TA, direct cytotoxic function of CD8 T cells, NK cells, and γδ T cells has been implicated in local tissue injury.[20] Conversely, in the aortic branches, LVV typically causes luminal stenosis/occlusion by inducing lumen-obstructive neotissue. Proliferating myofibroblasts lay down matrix and build hyperplastic intima. Newly formed microvessels supply oxygen and nutrients to the thickening wall. How T cells instruct vascular cells to this maladaptive wound healing response and whether heterogeneity of such vascular cell populations imposes the disease's tissue tropism are the subjects of ongoing research. So far, the few wall-infiltrating B cells have not been assigned to a specific pathogenic mechanism.

KEY CONCEPTS

Large Vessel Vasculitis

- In humans, medium and large arteries have multiple wall layers and a wall structure substantial enough to be targeted by autoimmune disease.
- Vasculitis causes the rapid and concentric growth of hyperplastic intima, leading to luminal occlusion and ischemia of dependent tissues. Intramural inflammation of the aorta can result in wall damage followed by aneurysm formation and rupture.
- Because of the vital function and nonregenerative nature of large human arteries, the threshold for autoimmunity in the wall structures of such arteries must be explicitly high.
- Inflammatory infiltrates forming granulomatous vasculitis enter the vessel from the "back door," the adventitia, and not from the lumen.
- Besides their critical role as conduits for blood flow, medium and large arteries also possess immunoregulatory functions mediated by dendritic cells (DCs) indigenous to the vascular wall. DCs in each vascular territory express a distinctive pattern of Toll-like receptors (TLRs), giving each vessel its own immunologic identity.

Defective T Regulatory Cells and Insufficient Immune Checkpoints in Giant Cell Arteritis

Recent data have brought to the forefront that antigen-nonspecific immunoregulatory pathways have disease relevance in LVV, specifically in enabling unopposed and lasting adaptive immune responses to induce and sustain vessel wall inflammation.

Like in many chronic inflammatory lesions, GCA's arterial wall infiltrates lack sufficient antiinflammatory T regulatory cells (Tregs). This may be a consequence of the proinflammatory milieu, but recent studies have identified a novel mechanism of Treg-dependent immunosuppression that is nonfunctional in GCA. In healthy individuals, secondary lymphoid tissues are occupied by a population of CD8 Tregs that effectively control clonal expansion of CD4 T cells and thus the overall size of the CD4 T-cell compartment.[21] Such CD8 Tregs suppress activation and expansion of neighboring CD4 T cells by the directed transfer of nicotinamide adenine dinucleotide phosphate (NADPH) oxidase 2 (Nox2)-containing exosomes. The population of $Nox2^+$ $CD8^+$ Tregs is distinctly low in patients with GCA and remains low with antiinflammatory therapy. Implicating Nox2 released by CD8 T cells in the threshold setting of CD4 T-cell immunity opens an entirely new perspective in vasculitis research and redirects the focus away from antigenic triggers that induce protracted macrophage activity.

A second immunoregulatory defect weakening proper control of CD4 T-cell immunity has been localized to the programmed death 1 (PD-1) immunoinhibitory checkpoint.[22] Inhibitory checkpoints are designed to protect tissue from excessive immune reactivity. PD-1 is expressed on activated T cells, the ligand PD-L1 is expressed on APCs. Triggering of PD-1 sends negative signals into T cells and stops their proliferation and polarization. DCs in patients with GCA, both lesional vasDCs and circulating DCs, are $PD\text{-}L1^{low}$. GCA T cells thus lack negative signaling and fail to undergo clonal contraction. Indeed, in inflamed temporal arteries, $PD\text{-}1^+$ T cells are strongly enriched. In human artery–chimeric mice, blocking PD-1 with antibody treatment exacerbates vasculitis, and wall-infiltrating $PD\text{-}1^+$ T cells produce a spectrum of effector cytokines (IFN-γ, IL-17, IL-9, IL-21, IL-22, etc.). Most importantly, PD-1–PD-L1 interactions in the artery have direct impact on the density of newly formed microvessels and the thickness of the hyperplastic intimal layer.[22] Hyperactive checkpoints are currently targeted therapeutically to unleash antitumor T-cell immunity. In cancer patients treated with checkpoint inhibitors, aortitis and vasculitis have been reported as immune-related adverse events.

CLINICAL FEATURES IN GIANT CELL ARTERITIS

CLINICAL PEARLS

Clinical and Epidemiologic Clues in Giant Cell Arteritis

- Patient older than 50 years of age.
- Female.
- Northern European heritage.
- Laboratory findings of a highly activated acute phase response (e.g., elevated erythrocyte sedimentation rate [ESR] and C-reactive protein [CRP]).
- Insidious onset of nonspecific symptoms (weight loss, night sweats, malaise, fever).
- Ischemia of ocular structures, cranial muscles, scalp, or upper extremities.

Clinical manifestations of GCA reflect the combination of a systemic inflammatory syndrome with vascular insufficiency (Table 60.1).[2] Increased sensitivity of vascular imaging methods and longer survival of affected individuals have made it clear

TABLE 60.1 Clinical Features of Giant Cell Arteritis, Polymyalgia Rheumatica, and Takayasu Arteritis

Organ System	Clinical Features	FREQUENCIES GCA	PMR	TA
Vascular	Headaches	+++		*
	Limb claudication	+		***
	Scalp tenderness	**		
	Jaw claudication	**		
	Absent or asymmetrical pulses	*		***
	Asymmetric blood pressure readings	*		***
	Bruit	*		***
	Tongue claudication	*		
	Tissue gangrene	*		
	Abdominal angina			*
	Cough (dry, nonproductive)	*		*
Constitutional	Malaise	**	**	***
	Failure to thrive	*	**	*
	Weight loss	**	**	**
	Fever	*	*	*
Central nervous	Ocular symptoms	**		*
	Stroke/transient ischemic attack	*		*
Peripheral nervous	Peripheral neuropathy	*		
Cardiac	Aortic dilatation and regurgitation	*		*
	Myocardial infarction	*		*
	Congestive heart failure			*
Musculoskeletal	Proximal stiffness/muscle pain	**	***	
	Synovitis of peripheral joints		*	
Others	Intense acute phase response	***	***	***
	Normochromic or hypochromic anemia	**	*	**

*Low frequency (<20%); **moderate frequency (20%–70%); ***high frequency (>70%).
GCA, Giant cell arteritis; *PMR*, polymyalgia rheumatica; *TA*, Takayasu arteritis.

that most patients eventually have involvement of medium and large arteries. In some patients, the disease preferentially targets more peripheral branches of the aorta (*e.g.*, cranial arteries, such as the temporal artery). In others, the aorta and its proximal branches (*e.g.*, subclavian, axillary arteries) are involved to a major extent. In a subpopulation of patients, the clinical consequences of arterial inflammation are minimal, and these patients come to clinical attention because of failure-to-thrive or fever of unknown origin.

In cranial GCA, symptoms result from vascular stenosis of the neck and head arteries, most prominently the branches of the external carotid artery. Arteritis of the scalp arteries leads to the typical presentations of headaches and scalp tenderness. Patients report difficulties with wearing glasses or combing their hair. The headaches are often intense and unresponsive to standard analgesics. Headaches are a nonspecific clinical symptom, yet in an older individual with other findings of an inflammatory syndrome, physicians need to rule out GCA. Insufficient blood flow to the masseter muscles and the tongue causes jaw or tongue claudication, elicited by prolonged chewing and talking. Although this type of claudication is present in less than 30% of patients, it is clinically helpful because it rarely occurs outside of GCA. Similarly, painful dysphagia can be a useful clinical clue.

The orbits and the optic nerve are strictly dependent on blood supply from the external carotid system, particularly the

ophthalmic artery. GCA in ophthalmic artery branches, specifically the posterior ciliary arteries, leads to anterior ischemic optic neuropathy, presenting as sudden and painless vision loss. Typically, patients lose vision in the early-morning hours or wake up blind. Involvement of one eye may be followed by visual loss in the partner eye if the disease is not diagnosed and treated promptly. Besides anterior optic neuropathy, GCA can cause a number of ischemic complications in the orbits and along the visual axis, which may present as diplopia or partial vision loss.[15] If recognized and treated immediately, sight loss is preventable, which indicates that GCA should be considered an ophthalmologic emergency.

Chronic nonproductive cough can be related to arteritis in bronchial artery branches. If the vertebral and basilar arteries develop vasculitic stenosis, ischemia of the central nervous system manifests with transient ischemic attacks or frank stroke.

In patients with large vessel GCA, cranial symptoms may be minimal, and temporal artery biopsy can be negative.[23] Vascular insufficiency is focused on the upper-extremity vessels and the aorta. In rare cases, lower extremities are affected. Typically, patients have asymmetric blood pressure readings or experience total loss of upper-extremity blood pressure and pulse caused by occlusions in the distal subclavian and axillary arteries (Fig. 60.5). Patients with subclavian GCA are on average approximately 10 years younger at disease onset than those with dominant cranial manifestations. Diagnosis of large vessel GCA is often delayed because symptoms are nonspecific and systemic inflammation is less pronounced. Ischemic pain in the hands when using the arms can be combined with coolness and bluish discoloration. Gangrene of the fingertips is rare. Disability can be significant because patients have difficulties with activities of daily living. With stenotic lesions of the subclavian arteries, blood pressure readings are unreliable, requiring alternative strategies for blood pressure monitoring. Although carotid involvement is considered infrequent, it can be challenging to distinguish atherosclerotic disease and vasculitic disease. Patients with carotid GCA are at high risk for cerebral ischemic events. Aortic involvement preferentially targets the thoracic aorta and infrequently the abdominal aorta (Fig. 60.6). Dilation of the aortic root can lead to aortic insufficiency. Aortic aneurysms are often clinically silent. The diagnosis may first be made from tissue obtained surgically during aortic aneurysm repair. In extreme cases, the aortic wall ruptures.

The response pattern of arteries to inflammation may not include intimal hyperplasia, thus eluding luminal compromise. In such patients, the systemic inflammatory component dominates the clinical presentation. Fever, fatigue, malaise, weight loss, and depression are often intense enough to prompt a workup for a malignancy. GCA needs to be in the differential diagnosis in all cases of fever of unknown origin, particularly in older individuals. Patients with cranial GCA have abnormally thick and tender temporal arteries exhibiting nodularity and loss of pulses, whereas clinical findings in nonstenosing GCA can be unremarkable. Temporal artery biopsy needs to be pursued even if clinical examination does not suggest the diagnosis.

CLINICAL FEATURES IN POLYMYALGIA RHEUMATICA

PMR is diagnosed in patients experiencing pronounced stiffness and pain in the shoulder and pelvic girdle muscles (see Table 60.1).[2] Laboratory testing reveals a systemic inflammatory syndrome; arterial biopsy is negative for arteritis. Approximately 10% of patients with PMR without any signs of vascular inflammation will eventually develop full-blown vasculitis. Notably, PMR often occurs in patients with GCA, in approximately 40% of patients at disease onset. Tapering of immunosuppressive therapy in GCA is frequently associated with new or remittent PMR symptoms. Complaints are focused on muscle pain and stiffness, classically affecting the neck, shoulders, and pelvic girdle. The muscles of the torso may be involved. Peripheral arms and legs are spared. Muscle pain is most intense in the early morning and improves during the day. Inability to get out of bed, stand up from a chair, or get off the toilet seat should alert the physician to consider PMR. Some patients with PMR have synovitis or bursitis in their shoulder and hip joints[2] difficult to distinguish from seronegative polyarthritis. No diagnostic procedure that allows for the diagnosis of PMR is available; the syndrome remains an exclusion diagnosis in cases of myalgia combined with laboratory signs of systemic inflammation. On clinical examination, passive motion of shoulder and hip joints is maintained, but active motion is restricted because of pain. Muscle strength is often normal. Careful evaluation of the temporal arteries is warranted to avoid missing fully developed GCA.

CLINICAL FEATURES IN TAKAYASU ARTERITIS

The clinical manifestations of TA are diverse and depend on the affected vascular territory (see Tables 60.1 and 60.2). Initial symptoms are usually nonspecific and include fever, cough, malaise, weight loss, night sweats, myalgias, and arthralgias. Signs of vascular deficiency develop later in the disease course and generally are ischemic in nature. Geographic variations in disease pattern have been reported, likely reflecting the interplay between host risk genes and dysfunctional immunity. In North American, Japanese, and Korean patients, the aortic arch and its primary cervical and upper extremity branches are preferentially targeted, giving rise to aortic insufficiency, cerebral ischemia, face and neck pain, ocular ischemia, and the typical presentation of "pulseless disease" (Fig. 60.7). In patients in India, the abdominal aorta and renal arteries are more commonly affected, causing renovascular hypertension and the long-term risk of cardiac failure (Fig. 60.8).

Nonspecific complaints of headaches, syncope, and face and neck pain are often misinterpreted as stress-related problems, particularly in young women. Consequently, the diagnosis can be missed for months. Only a few patients come to clinical attention because of catastrophic neurologic symptoms related to brain ischemia. Helpful clues are differences in blood pressure, loss of pulses, and vascular bruits heard on clinical examination. Retinal neoangiogenesis, induced by hypoperfusion of the eye is now relatively rare, but fleeting visual abnormalities may indicate transient ischemic attacks. Signs of aortic insufficiency are unlikely to be encountered in early disease, but continuous monitoring for aortic dilation is an essential part of follow-up care. Coronary artery stenosis in a young patient must prompt the physician to rule out TA. In a subset of patients, the origins of mesenteric arteries are involved. Clinical consequences include weight loss, nausea, vomiting, diarrhea, and abdominal claudication, typically elicited by the increased intestinal blood demand following a meal.

Renal artery stenosis may be clinically silent and is often noticed in routine screening. Correct measurement of blood pressure can represent a pressing clinical problem if the upper-extremity arteries are affected. Involvement of the infrarenal aorta can lead to lower-extremity claudication. Musculoskeletal examinations are usually unrevealing, although joint and muscle pains are common.

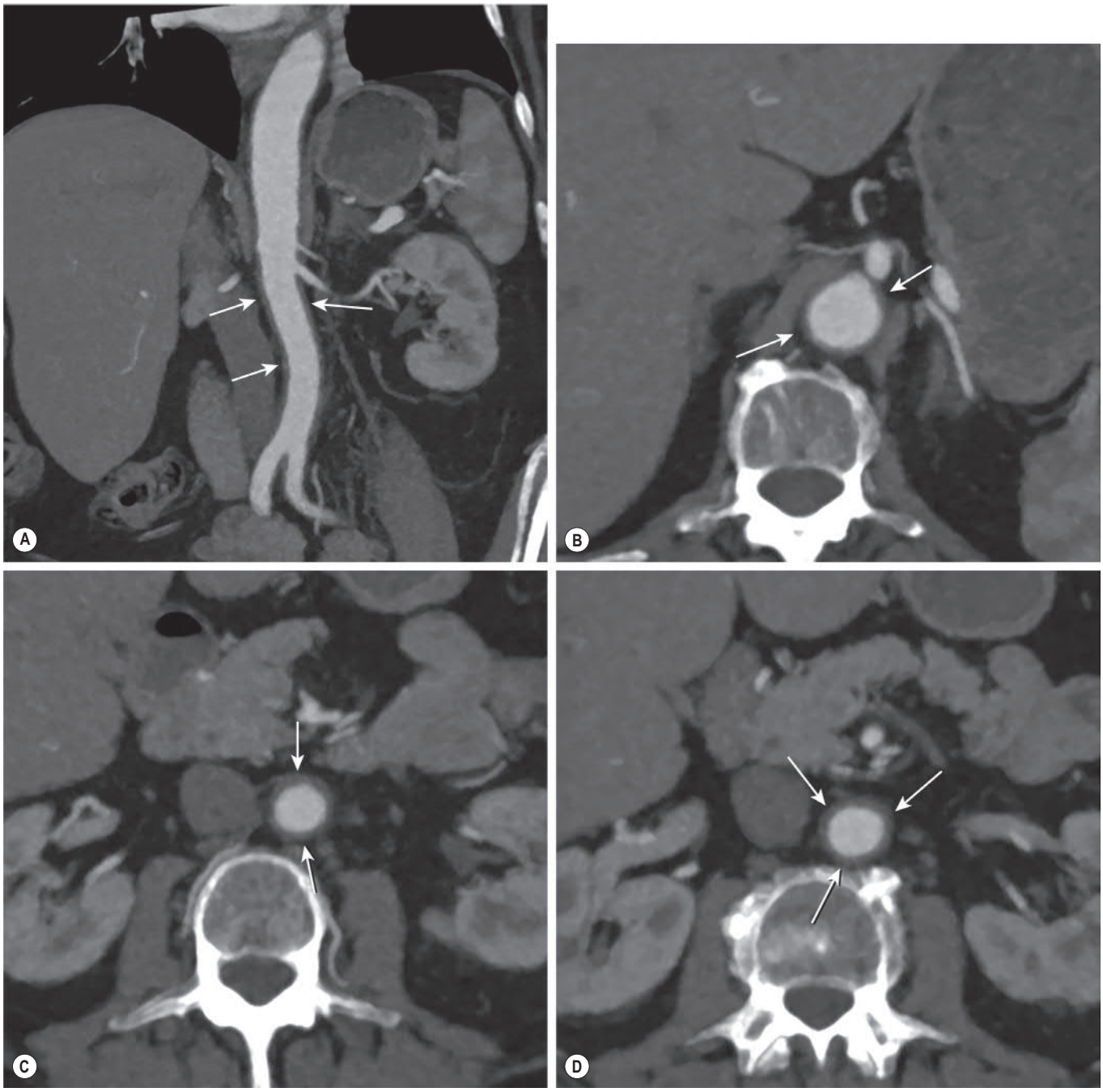


FIG. 60.5 Diagnostic Imaging in Giant Cell Arteritis: Computed Tomography Angiography. Contrast-enhanced computed tomography angiography (CTA) in a 74-year-old female with a temporal artery biopsy positive for giant cell arteritis. CTA imaging shows diffuse, circumferential mural thickening along the entire descending aorta and the abdominal aorta (Panel A). *Arrows mark* the thickened aortic wall. Axial images (Panels B–D) reveal the circumferential distribution *(arrows)* of the wall thickening from the distal arch to the intrarenal portion of the aorta. Thickness measurements can be used to monitor disease burden over time. The aortic diameter is within normal limits, indicating that the aortitis has not yet resulted in aneurysm formation. (Images were generated by Dr. D. Fleischmann, Department of Radiology, Stanford University.)

DIAGNOSIS

Classification criteria have been developed for GCA and TA to differentiate patients with LVV from those with other vasculitic entities (Tables 60.3 through 60.5).[24–26] Age at disease onset and the pattern of arteritis are clearly important for establishing the diagnosis and distinguishing between these two related vasculopathies. Diagnostic criteria for PMR remain a challenge (see Table 60.3) because they rely on nonspecific symptoms, such as muscle pain and stiffness and elevated erythrocyte sedimentation rate (ESR), all of which can occur in many other diseases.[2] No specific laboratory test is currently available to diagnose PMR. Therapeutic responsiveness of patients with PMR to low-dose corticosteroids is clinically helpful, emphasizing the need for objective diagnostic criteria.

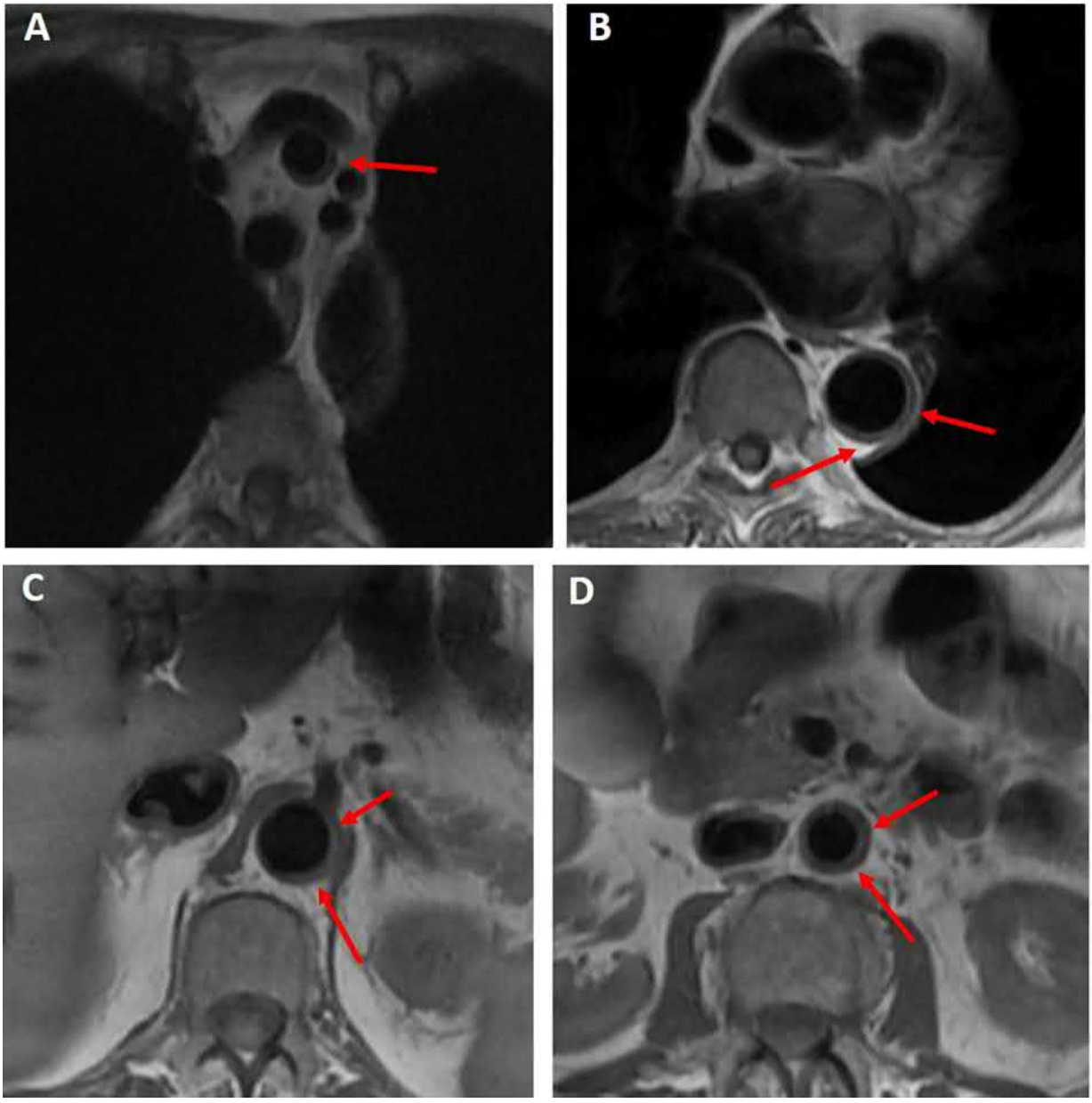


FIG. 60.6 Diagnostic Imaging in Giant Cell Arteritis: Magnetic Resonance Angiography. Contrast-enhanced magnetic resonance angiography of the chest and abdomen in the patient whose computed tomography angiography images are presented in Fig. 60.5. Double inversion recovery magnetic resonance images in the axial plane demonstrate diffuse, contiguous mural thickening of the great vessels (Panel A, brachiocephalic trunk) and the aorta (Panels B–D). *Arrows* are placed to mark the circumferential distribution of the mural thickening and compare luminal diameter and wall thickness at different levels of the vessels. (Images courtesy of Dr. D. Fleischmann, Department of Radiology, Stanford University.)

Laboratory Tests

In all three conditions—GCA, PMR, and TA—the laboratory tests indicate an intense acute phase response in a vast majority of patients.[2,27] In general, this is captured by measuring ESR or CRP. However, it is important to note that a subset of patients with GCA has normal ESR readings, even before initiation of immunosuppressive therapy. A normal ESR or CRP result is not sufficient to exclude the diagnosis, and further diagnostic work-up is required if the clinical suspicious is high. Other acute phase proteins, such as fibrinogen and SAA, are elevated as well. IL-6 is a potent inducer of hepatic acute phase proteins and is a sensitive marker of continuous systemic inflammation.[8,9] Other laboratory abnormalities, such as elevation of alkaline phosphatase, thrombocytosis, and anemia, are in line with a robust acute phase response.

Autoantibodies are not helpful beyond excluding differential diagnoses, such as rheumatoid arthritis (RA), systemic lupus erythematosus (SLE), or antineutrophil cytoplasmic antibody (ANCA)-related vasculitides. Recently, two endothelial autoantigens have been identified in Takayasu disease, but autoantibody measurements have not yet entered diagnostic testing.[28]

Tissue Biopsy

In patients with TA, tissue biopsies are rarely available unless the patient had to undergo vascular reconstructive surgery. In most patients, the diagnosis is made based on imaging procedures revealing luminal and wall abnormalities in affected blood vessels.

In contrast, arterial biopsy remains a critical diagnostic approach in patients with GCA. Temporal arteries are easily

TABLE 60.2 Takayasu Arteritis: Relationship Between Clinical Symptoms and Affected Vascular Territories

Vascular Bed Involvement	Approximate Frequency (%)	Predominant Clinical Symptoms
Subclavian	90	Arm claudication, pulselessness
Common carotid	60	Visual defects, stroke, transient ischemic attack, syncope
Abdominal aorta	45	Claudication, hypertension, abdominal angina
Renal	35	Hypertension
Aortic arch/root	35	Aortic insufficiency, congestive heart failure
Vertebral	35	Dizziness, visual impairment
Celiac axis	20	Abdominal angina
Superior mesenteric	20	Abdominal angina
Iliac	20	Claudication
Pulmonary	10	Dyspnea, chest pain
Coronary	10	Myocardial infarction, angina

accessible, and a segment can be removed in an outpatient setting. Recommendations include harvesting 2 to 3 cm of the temporal artery, starting at the most symptomatic side. Frozen tissue sections can lead to a quick diagnosis of granulomatous vasculitis. Whether the second side should be biopsied during the same surgical procedure remains a matter of debate. In cohorts that included several hundred patients, vasculitis was detected in 2% to 3% of tissue samples from the second side if the first side was negative. If the clinical suspicion is strong, biopsy confirmation can be sought from a second-side biopsy immediately after the first biopsy or after careful monitoring of the patient for several weeks. Negative findings on temporal artery biopsy do not exclude the diagnosis of GCA but make it unlikely. In a retrospective cohort study, approximately half of the patients with subclavian GCA had no vasculitis in the temporal arteries, consistent with preferential involvement of certain vascular territories. There has been a trend toward considering negative biopsies as "falsenegative" and classifying patients with a negative biopsy as having GCA. This may lead to unnecessary immunosuppressive therapy, and the patient may not obtain a proper diagnosis. Findings that prompt biopsies (headaches, elevated acute phase reactants) are notoriously nonspecific, and both physicians and patients are anxious to avoid treatable blindness, biasing toward overtreatment. The temporal artery biopsy remains a powerful diagnostic tool and the major tool that allows unequivocal classification of the disease process. A technically proper temporal artery biopsy will detect vasculitis in the vast majority of patients.

Corticosteroid therapy does not eradicate pathologic findings of vascular wall infiltrates, and biopsy can still be valuable in making the diagnosis in patients on steroids. Approximately half of the patients remain biopsy positive, even after 1 year of corticosteroid therapy.[29]

Histomorphologic reports describe mononuclear cell infiltrates penetrating through all layers of the vessel wall (Fig. 60.9).[1] Recent discussions have focused on the diagnostic relevance of isolated inflammatory cell clusters in the adventitia or perivascular lymphocytes limited to small blood vessels. These findings may not be sufficient to indicate arteritis. Multinucleated giant cells may or may not be found. They tend to lie along the internal elastic lamina, at the junction between the media and the intima. Media destruction is not unusual, but findings of fibrinoid necrosis should prompt a search for different vasculitic entities. The vessel lumen is compromised by hyperplastic intima formed from proliferating fibroblasts, smooth muscle cells, and deposition of acid mucopolysaccharides.

The histology of TA is similar to that of GCA, making it difficult to dissect both syndromes in aortic tissue samples. Lymphocytes and plasma cells accumulate around vasa vasorum and form transmural infiltrates. Marked wall thickening with inflammatory tissue extending into perivascular structures is typical for TA (Fig. 60.10). Destruction of elastic membranes is often extensive and combined with patchy areas of media necrosis. Weakening of the vessel wall can lead to aneurysm formation. Inflammatory lesions may be arranged in a "skipped" pattern, with normal vessel wall segments alternating with stretches of intense destructive inflammation.

Physicians may encounter morphologic findings of granulomatous aortitis in patients undergoing aortic aneurysm repair without any prior diagnosis of vasculitis. Detailed work-up of these patients is necessary to identify those with undiagnosed PMR, GCA, or TA. Rare causes of aortitis, including IBD, sarcoidosis, syphilis, relapsing polychondritis, and connective tissue disease, should be ruled out. Isolated granulomatous aortitis is diagnosed as idiopathic aortitis. The pathogenesis and prognosis of this condition are essentially unknown.

Diagnostic Imaging

Modern imaging modalities have fundamentally changed the diagnostic approach to LVV. Indeed, diagnosing TA mostly depends on identifying vascular lesions in typical distribution by imaging.[30]

Conventional angiography still has its place in preoperative planning and can be combined with intravascular interventions. It provides ideal visualization of the vascular lumen not only for large but also for medium-sized arteries, such as the axillary and brachial arteries (see Fig. 60.5). Ultrasound (US)-based methods are extremely useful for screening carotid arteries, but they have also emerged as the method of choice for initial assessment of the distal subclavian arteries, vertebral arteries, renal arteries, and femoral arteries. US examination is the optimal method for long-term monitoring of vessel bypasses after revascularization procedures. Magnetic resonance imaging (MRI), magnetic resonance angiography (MRA), and computed tomography (CT) are currently widely used for evaluating the vascular tree. These methods provide excellent information on abnormalities of the vascular lumen and wall, also capturing abnormalities in the more peripheral arterial branches. CT imaging is fast, well tolerated by patients with claustrophobia, and allows excellent assessment of the aorta and its wall (see Fig. 60.7). However, it has the disadvantages of contrast loading and radiation exposure. With its inherent multiplanar imaging capabilities, magnetic resonance is used to examine neck vessels, the aorta, and proximal aortic branches (see Figs. 60.6 and 60.8). Great hope was placed on its potential to measure wall edema and intramural vascularity, which would make magnetic resonance useful for estimating disease burden and responses to therapy. Comparisons of imaging results, laboratory parameters of inflammation, and surgical biopsy specimens have been disappointing, cautioning that edema-weighted magnetic resonance should not be used as a sole means of measuring disease activity and therapeutic responsiveness. Both CT angiography and MRA are currently used routinely to monitor progression/regression of vascular involvement and have an important place in managing the chronic phase of GCA and TA.

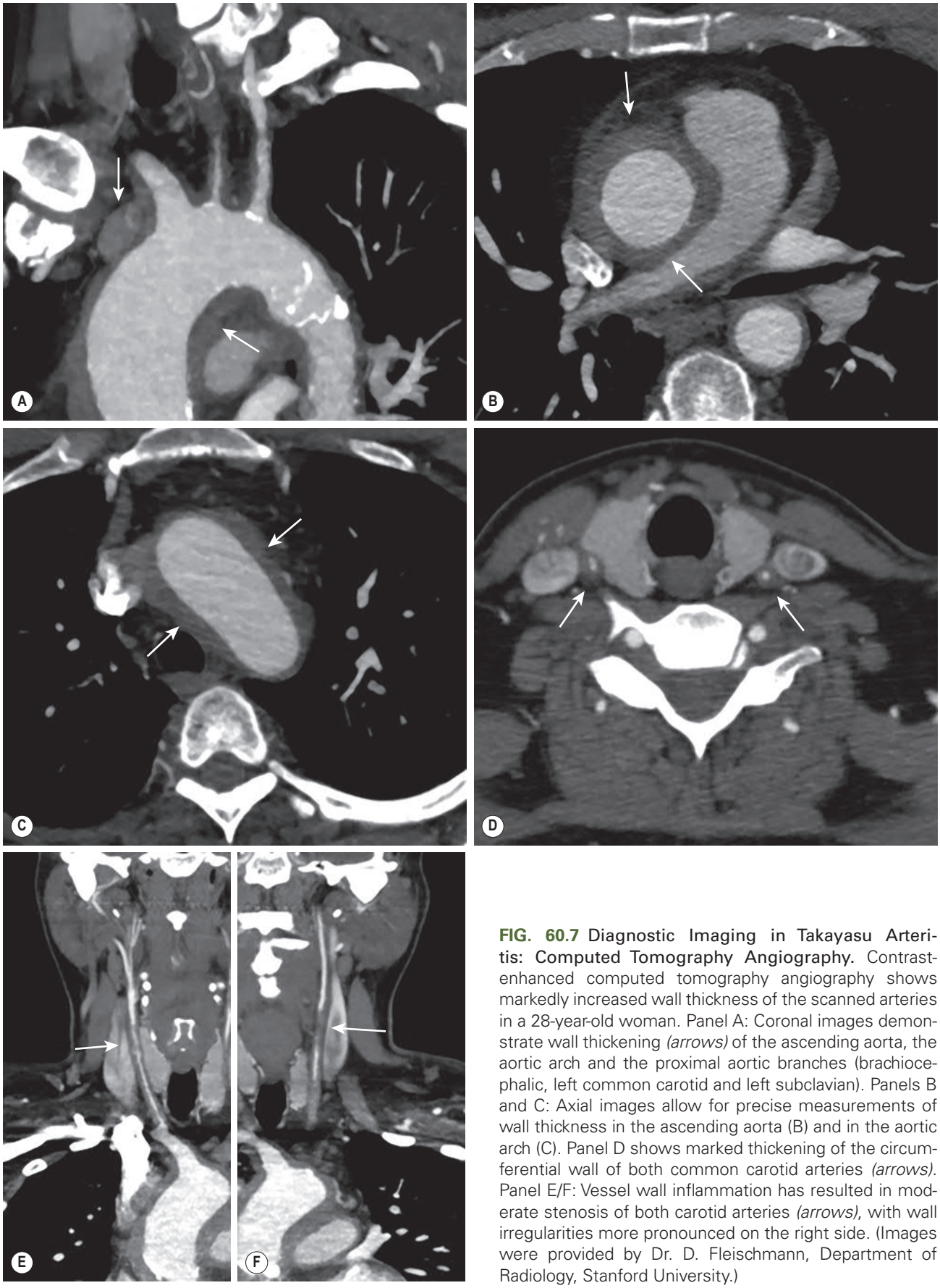


FIG. 60.7 Diagnostic Imaging in Takayasu Arteritis: Computed Tomography Angiography. Contrast-enhanced computed tomography angiography shows markedly increased wall thickness of the scanned arteries in a 28-year-old woman. Panel A: Coronal images demonstrate wall thickening *(arrows)* of the ascending aorta, the aortic arch and the proximal aortic branches (brachiocephalic, left common carotid and left subclavian). Panels B and C: Axial images allow for precise measurements of wall thickness in the ascending aorta (B) and in the aortic arch (C). Panel D shows marked thickening of the circumferential wall of both common carotid arteries *(arrows)*. Panel E/F: Vessel wall inflammation has resulted in moderate stenosis of both carotid arteries *(arrows)*, with wall irregularities more pronounced on the right side. (Images were provided by Dr. D. Fleischmann, Department of Radiology, Stanford University.)

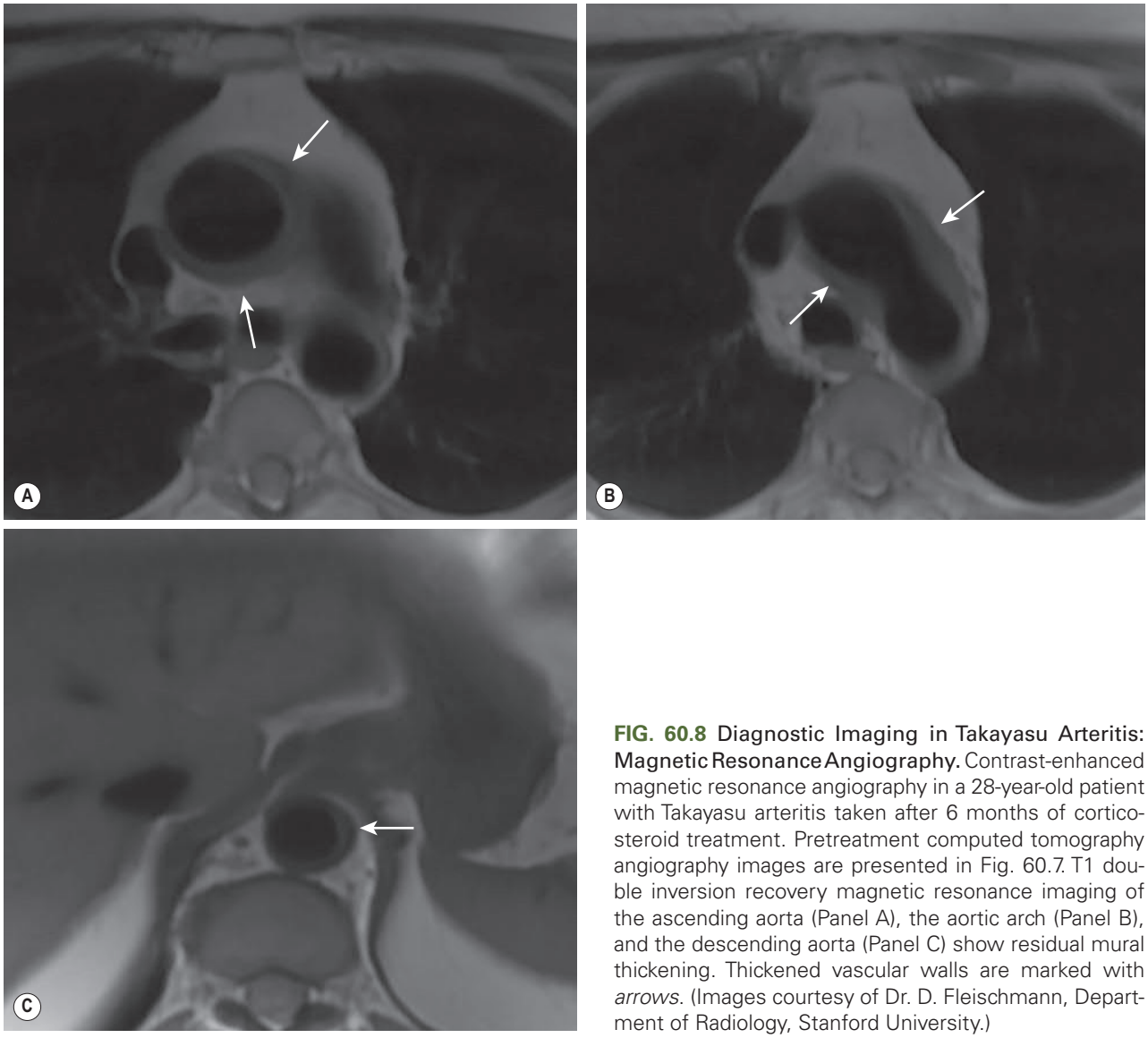


FIG. 60.8 Diagnostic Imaging in Takayasu Arteritis: Magnetic Resonance Angiography. Contrast-enhanced magnetic resonance angiography in a 28-year-old patient with Takayasu arteritis taken after 6 months of corticosteroid treatment. Pretreatment computed tomography angiography images are presented in Fig. 60.7. T1 double inversion recovery magnetic resonance imaging of the ascending aorta (Panel A), the aortic arch (Panel B), and the descending aorta (Panel C) show residual mural thickening. Thickened vascular walls are marked with *arrows*. (Images courtesy of Dr. D. Fleischmann, Department of Radiology, Stanford University.)

TABLE 60.3 American College of Rheumatology 1990 Classification Criteria for Giant Cell Arteritis[a] and Polymyalgia Rheumatica

Age at disease onset ≥50 years
New-onset or new type of headache
Temporal artery tenderness or decreased artery pulse
Elevated erythrocyte sedimentation rate (≥50 mm/h)
Histologic incidence of arteritis (characterized by a predominance of mononuclear cell infiltrates or a granulomatous process with multinucleated giant cells)

[a]A patient is classified as having giant cell arteritis if at least three of the five criteria are present.

Reprinted from Hunder GG, Bloch DA, Michel BA, et al. The American College of Rheumatology 1990 criteria for the classification of giant cell arteritis. . 1990;33:1122–1128, with permission of Wiley-Liss, Inc., a subsidiary of John Wiley & Sons, Inc. ©1990.*Arthritis Rheum*

TABLE 60.4 Provisional Classification Criteria for Polymyalgia Rheumatica

Age ≥50 years	Required
Bilateral shoulder ache	Required
Abnormal C-reactive protein (CRP) and/or erythrocyte sedimentation rate (ESR)	Required
Morning stiffness >45 min	2
Hip pain or limited range of motion	1
Absence of rheumatoid factor (RF) or anticitrullinated protein antibody (ACPA)	2
Absence of other joint involvement	1
A score of ≥4 is categorized as polymyalgia rheumatica.	

Reprinted from Dasgupta B, Cimmino MA, Kremers HA, et al. Provisional classification criteria for polymyalgia rheumatica: A European League Against Rheumatism/American College of Rheumatology collaborative initiative. *Arthritis Rheum.* 2012;64:943–954.

TABLE 60.5 American College of Rheumatology 1990 Criteria[a] for the Classification of Takayasu Arteritis

Disease onset at ≤40 years
Claudication of an extremity
Decreased brachial artery pulse
>10 mm Hg difference in systolic blood pressure between arms
Bruit over the subclavian arteries or the aorta
Arteriographic evidence of narrowing or occlusion of the entire aorta, its primary branches, or large arteries in the proximal upper or lower extremities

[a]For purposes of classification, a patient is classified as having Takayasu arteritis if more than three of the six criteria are fulfilled.
Reprinted from Arend WP, Michel BA, Bloch DA, et al. The American College of Rheumatology 1990 criteria for the classification of Takayasu arteritis. . 1990;33:1129–1134, with permission of Wiley-Liss, Inc., a subsidiary of John Wiley & Sons, Inc. ©1990. *Arthritis Rheum*

THERAPEUTIC MANAGEMENT

With increasing knowledge of the disease process and refinement of diagnosis and long-term treatment, the prognosis for patients with LVV has significantly improved. Life expectancy of patients with GCA is preserved. Follow-up studies of Japanese patients with TA have suggested good control of disease activity in about 75% of patients, with only 25% experiencing serious complications and cardiac manifestations determining long-term outcome. Whether vasculitis predisposes patients to accelerated atherosclerotic disease, given the combination of chronic inflammation and injury to vessel wall structures, is still being discussed. It is not known whether progression of atherosclerosis and its complications require a different management approach or whether standard vasoprotective measures (treating hypertension and hyperlipidemia, smoking cessation, etc.) are sufficient.

Pathogenic studies have pointed out that the traditional view of GCA as a self-limiting disease is incorrect.[9,19] To the contrary, granulomatous vasculitis has shown surprising resistance to immunosuppression, with vessel wall infiltrates persisting for greater than 12 months in almost 50% of patients despite appropriate immunosuppressive therapy. Based on examination of serial temporal artery biopsy specimens from patients before and after treatment, it is now clear that arteritis persists, albeit sustained by an immune network distinct from that in untreated patients.[18] It is unknown whether this persistent smoldering process requires treatment and what the risk/benefit ratio is for the older patient population affected by GCA. Unchanged life

THERAPEUTIC PRINCIPLES

Treatment of Large Vessel Vasculitides

- To prevent vision loss, patients with giant cell arteritis (GCA) require immediate treatment. Similarly, with the threat of catastrophic cerebral ischemia in Takayasu arteritis (TA), prompt initiation of therapy is imperative.
- Corticosteroids are the immunosuppressive drug of choice for large vessel vasculitides (LVVs). Often, the drugs must be given over a period of several years but may be clinically effective at very low doses.
- Clinical trials have failed to show convincing steroid-sparing effects for either methotrexate or tumor necrosis factor (TNF)-α blockade in GCA.
- A phase 3 clinical trial has demonstrated steroid-sparing effects of treatment with tocilizumab, an antibody against the interleukin-6 receptor
- Molecular studies of GCA vascular lesions have shown that early and untreated disease is characterized by two functional T-cell lineages; T helper 1 (Th1) and Th17 cells. Th17 cells respond rapidly to corticosteroids, whereas Th1 cells persist and promote chronic, smoldering vasculitis.
- A dual biopsy study in untreated and treated patients with GCA has demonstrated persistent vasculitis despite excellent control of peripheral inflammatory markers (erythrocyte sedimentation rate [ESR], C-reactive protein [CRP]), indicating separation of the vascular and extravascular disease component. Although extravascular disease (ESR, CRP, and myalgias) may be easy to control, vessel wall inflammation seems to be autonomous and difficult to suppress.
- Clinical experience (not evidence-based therapeutic trials) suggests that a combination of methotrexate, mycophenolate mofetil, or TNF-α–blocking agents with corticosteroids may be beneficial in controlling disease in some patients with TA.
- Close monitoring for diabetes, hypertension, and hyperlipidemia combined with bone-saving therapy should be part of the treatment regime in patients with LVVs on long-term corticosteroids.

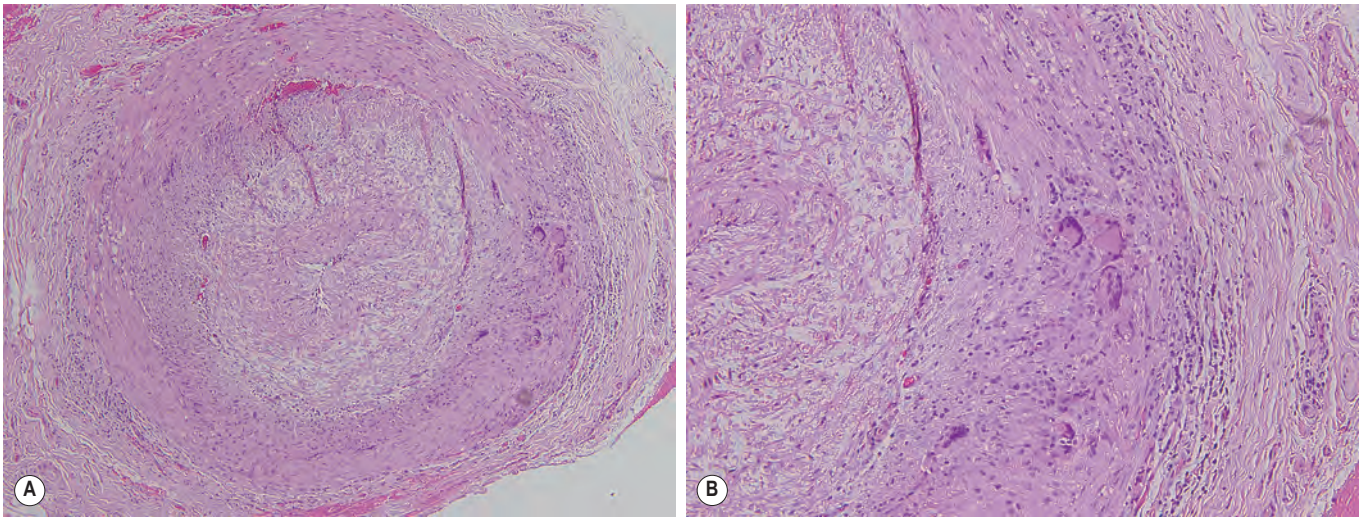


FIG. 60.9 Histomorphology of Giant Cell Arteritis. (A) Temporal artery cross-section with mononuclear infiltrates throughout all wall layers. The adventitia is infiltrated by round cells with cuffing of vasa vasorum by lymphocytes. The vessel lumen is occluded by intimal hyperplasia. (B) Higher magnification showing intense granulomatous inflammation with multinucleated giant cells in the proximal media and at the media–intima junction.

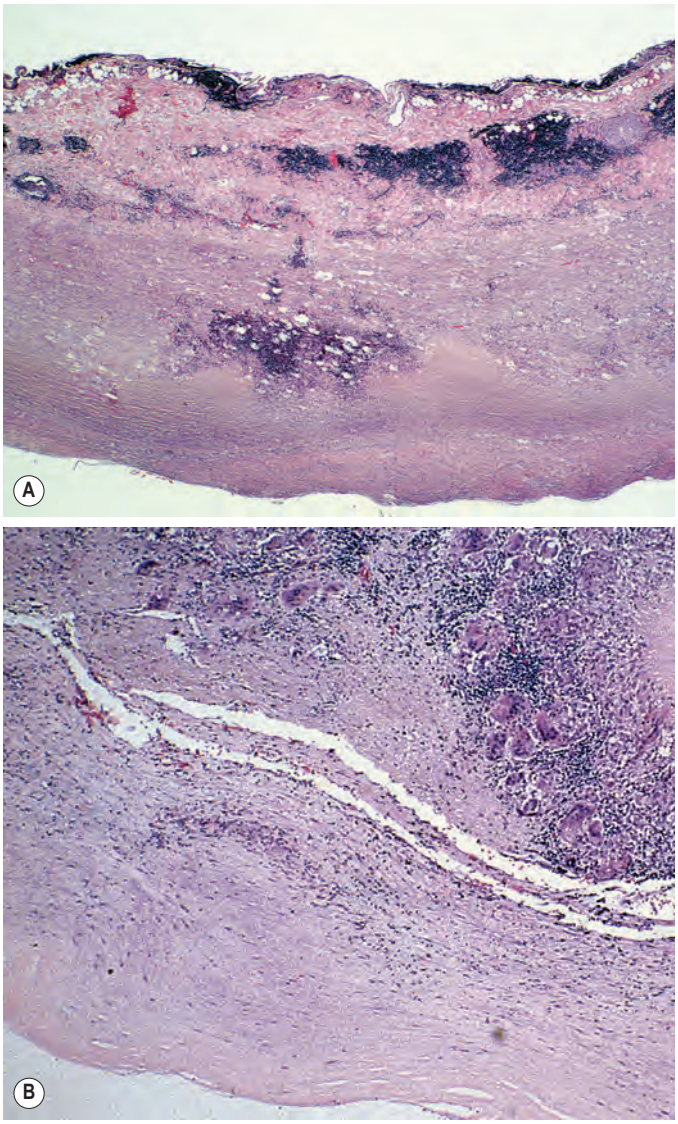


FIG. 60.10 Histopathology of Takayasu Arteritis. (A) Full-thickness section of the aortic wall shows dense mononuclear infiltrates in the adventitia and media. The intima is thickened and wavy; hematoxylin and eosin (H&E). (B) Florid granulomatous inflammation along the media–intima junction with numerous giant cells; H&E.

expectancy in GCA suggests adequacy of current management. Whether more intense immunosuppression or chronic maintenance therapy can prevent long-term complications, such as aortic aneurysm/dissection from GCA aortitis, and improve the overall prognosis is unknown. The ultimate decision depends on the cost/benefit analysis comparing the risk from smoldering disease with the risks imposed by long-standing immunosuppression. In that context, it is important to remember the profound impact of the immune aging process, which leaves older patients with an impaired immune system and amplifies the risk of immunosuppression.

Induction Therapy

In newly diagnosed patients with GCA, TA, and PMR, the immunosuppressants of choice are corticosteroids. Patients with GCA

ON THE HORIZON

- The vascular lesions of giant cell arteritis (GCA) are autonomous and self-maintained over prolonged periods. New therapeutic approaches are needed that can eradicate wall-residing immune cells.
- With extension of life span, patients may experience disease complications 20 years after diagnosis. Long-term immunosuppression may be required to protect the integrity of the large arteries.
- Several cellular activation pathways have disease relevance in GCA and Takayasu arteritis (TA). Interfering with cellular signaling through small-molecule reagents will provide new therapeutic opportunities.
- Patients with GCA have a defect in immunoprotective CD8 T regulatory (Treg) cells. Reconstituting Treg function emerges as a new therapeutic target.
- Patients with GCA have abnormalities in innate immune function (*e.g.*, hyperproduction of matrix metalloproteinase-9 (MMP-9) by monocytes). Monocyte reeducation represents a novel therapeutic strategy.
- Patients with GCA have a defective programmed death 1 (PD-1)/PD-ligand 1 immune checkpoint. Cancer patients treated with immune checkpoint inhibitors are at risk to develop aortitis/vasculitis.
- A subset of patients with TA build autoantibodies against endothelial surface receptors. The value of such autoantibodies to subset and monitor patients will need to be explored.

are started on a daily prednisone dose of 40 to 60 mg (approximately 1 mg/kg body weight). In patients with PMR, a daily dose of 20 mg prednisone is sufficient in almost all patients. The response is usually dramatic, with improvements within 24 to 48 hours. The promptness of clinical improvement is so exceptional that it has been suggested as a diagnostic criterion for PMR. The promptness of response may be limited to extravascular disease. Myalgias, fever, malaise, and headaches improve swiftly, in parallel with a fast reduction of acute phase reactants (CRP, IL-6, ESR). The vascular component is much more resistant to immunosuppression and may require a different therapeutic approach.

Once the condition is stabilized, steroid tapering is guided by close monitoring of the clinical presentation as well as laboratory markers of inflammation. In general, steroids should be reduced by 10% to 20% every 2 weeks. Monthly monitoring of ESR and CRP is mandatory to adjust therapy. Patients frequently return with signs or symptoms of recurrent disease as immunosuppression is lowered. Fortunately, disease exacerbations causing vision loss are infrequent. Disease flare-ups typically present with PMR symptoms or nonspecific manifestations of malaise and failure to thrive. In most patients, a transient small increase in the steroid dose reinstates disease control.

Much effort has been invested in identifying steroid-sparing agents. In a small study, treatment with pulse corticosteroids appeared to have long-term beneficial effects, reducing the overall steroid requirement and the rate of disease flare-ups.[31] Compared with the control arm, patients who received three initial steroid pulses (1000 mg methylprednisolone ×3 days, followed by oral steroids with a fast taper) had lower likelihoods of disease flare-ups. Particularly, once close to 10 mg/day of prednisone, these patients could tolerate steroid withdrawal significantly better, and most were taking 5 mg/day prednisone at 36 weeks.[31] The benefit from initial pulse therapy continued over subsequent months.

Several biologic agents have been explored or are currently undergoing testing in clinical trials.[32] Tumor necrosis factor-α (TNF-α) inhibitors may have a role in TA but had no steroid-sparing effect in GCA.[33] Preliminary study results suggest that

targeting T-cell costimulation with abatacept may prevent disease relapses in GCA. Ustekinumab, an antibody targeting IL-12 and IL-23, was reported to have potential efficiency in refractory GCA in a small open-label study. The IL-6 receptor blocker tocilizumab has been explicitly effective in reducing acute phase reactants (CRP, ESR) and in a phase 3 double-blind trial of tocilizumab given weekly or every other week demonstrated substantial steroid-sparing effect over a 1-year period.[10]

Maintenance Therapy

With a major shift in the pathogenic understanding of LVV, especially the realization that the disease process has two, partly independent components (extravascular, vascular) and that vasculitis persists chronically, the therapeutic needs for maintenance therapy have become the dominant issue for the treating physician. Patients with PMR are often managed successfully with low-dose corticosteroids (prednisone 5 mg daily) and typically are highly responsive to transient and very small dose increases (1 to 2 mg prednisone/day). Long-term management of patients with GCA and TA relies on low-dose corticosteroids as well unless there is objective evidence for progressive vascular wall inflammation. Unfortunately, no reliable biomarkers can separate the extravascular and vascular disease components, and no evidence has been presented that suppressing acute phase responses will ultimately restrict transmural vasculitis.

Methotrexate is considered to have mild-to-moderate steroid-sparing potential in GCA and PMR[34] but is more frequently used in TA. When given to human artery–severe combined immunodeficiency (SCID) chimeras, acetylsalicylic acid (aspirin) has marked antiinflammatory activities, with suppression of IFN-γ in vascular lesions. Clinical trials are needed to test whether this immunosuppressive action can translate into corticosteroid sparing. Because arteries are the primary targets of LVVs, the use of aspirin as an antiplatelet agent should be routinely recommended.

There is no evidence that immunosuppressants, such as azathioprine and cyclophosphamide, lower steroid needs, prevent vascular complications, or shorten the duration of steroid use. Whether any of the aforementioned biologic agents have a place to effectively suppress vessel wall inflammation and change the course of chronic disease is currently unknown.

An integral part of chronic immunosuppression with prednisone is regular monitoring for diabetes and hypertension. Patients should be encouraged to increase physical activity because steroid-induced myopathy occurs frequently. A major issue of chronic steroid treatment is the risk of excessive bone loss, possibly resulting from increased bone resorption and impaired bone formation. Several effective and safe therapies for osteopenia/osteoporosis are available. Calcium and vitamin D supplementation should be part of the therapeutic regimen.

In many, but not all, patients, immunosuppressive treatment can be discontinued 18 to 24 months after diagnosis. Markers of systemic inflammation may remain elevated, and continuous monitoring for aortic involvement and recurrence of cranial arteritis is recommended.

Most patients with PMR are sufficiently treated with an initial dose of 20 mg of prednisone per day. In some patients, 10 mg of prednisone can induce and sustain a clinical response. Steroids should be titrated to minimally needed doses to avoid side effects; tapering usually needs to be slow, over many months. In TA, long-term management should be tailored to individual patient conditions. It has been argued that patients should be maintained on a low dose of corticosteroids, such as 5 to 7 mg prednisone daily, even after successful control of active disease.

Given the age at disease onset in TA, preventive measures to counteract accelerated atherosclerosis and optimize blood pressure control are important aspects of management.

It has been suggested that up to 50% of patients with TA may require a second immunosuppressive agent. Steroid-sparing effects of methotrexate have been reported for some patients. Similarly, mycophenolate mofetil may have clinical efficiency, although published data are available only for a small patient cohort. Empirically, azathioprine may have a place in maintenance therapy of patients with TA. There may be a place for agents blocking TNF-α in patients with persistent disease activity. A randomized, double-blind, placebo-controlled, phase 3 trial of tocilizumab in Japanese patients with refractory TA did not meet the primary end point and disease progression in tocalizumab-treated TA patients has been reported, raising concerns that vascular inflammation in TA may be difficult to treat.[35,36]

Detecting and treating hypertension is an essential component of caring for patients with TA. Untreated hypertension leads to acceleration of atherosclerosis and cardiac insufficiency. In patients with upper-extremity involvement, obtaining accurate blood pressure measurements is a challenge and requires education of the patient and caregivers.

Revascularization Procedures

Besides pharmacologic therapy, revascularization procedures—including both surgical and endovascular interventions—have vastly broadened therapeutic options in patients with TA and large vessel GCA. To minimize the risk of complications, such as rapid reocclusion, vascular wall inflammation needs to be controlled before subjecting the patient to revascularization treatment. Conventional bypass grafts are still considered the method of choice in severely affected patients. Percutaneous transluminal angioplasty can be useful in managing renal artery stenosis or other short-segment lesions. Bypass surgery is needed in patients with cerebrovascular ischemia in whom catastrophic strokes may be prevented by bypassing critical stenosis of cervical vessels with grafts originating from the aortic arch. Reestablishing flow in the upper- and lower-extremity arteries can be complicated by multiple and long-segment stenosis, and arterial reconstructions with prosthetic graft materials or veins may be the only alternative to obtain long-term patency. Placing of conventional stents can be complicated by rapid restenosis, and it is unknown whether outcomes can be improved by drug-eluting stents. Occlusive disease of the coronary arteries usually represents a challenging clinical scenario, and most physicians opt for conventional bypass surgery. Surgical repair is the treatment of choice in patients with aortic regurgitation due to aortic wall and/or aortic valve inflammation.

REFERENCES

1. Weyand CM, Goronzy JJ. Medium- and large-vessel vasculitis. *N Engl J Med*. 2003;349(2):160–169.
2. Buttgereit F, Dejaco C, Matteson EL, et al. Polymyalgia rheumatica and giant cell arteritis: a systematic review. *JAMA*. 2016;315(22):2442–2458.
3. Nakagomi D, Jayne D. Outcome assessment in Takayasu arteritis. *Rheumatology (Oxford)*. 2016;55(22):1159–1171.

4. Isobe M. Takayasu arteritis revisited: current diagnosis and treatment. *Int J Cardiol*. 2013;168(1):3–10.
5. Wagner AD, Goronzy JJ, Weyand CM. Functional profile of tissue-infiltrating and circulating CD68+ cells in giant cell arteritis. Evidence for two components of the disease. *J Clin Invest*. 1994;94(3):1134–1140.
6. Baldini M, Maugeri N, Ramirez GA, et al. Selective up-regulation of the soluble pattern-recognition receptor pentraxin 3 and of vascular endothelial growth factor in giant cell arteritis: relevance for recent optic nerve ischemia. *Arthritis Rheum*. 2012;64(3):854–865.
7. Nadkarni S, Dalli J, Hollywood J, et al. Investigational analysis reveals a potential role for neutrophils in giant-cell arteritis disease progression. *Circ Res*. 2014;114(2):242–248.
8. Roche NE, Fulbright JW, Wagner AD, et al. Correlation of interleukin-6 production and disease activity in polymyalgia rheumatica and giant cell arteritis. *Arthritis Rheum*. 1993;36(9):1286–1294.
9. Weyand CM, Fulbright JW, Hunder GG, et al. Treatment of giant cell arteritis: interleukin-6 as a biologic marker of disease activity. *Arthritis Rheum*. 2000;43(5):1041–1048.
10. Stone JH, Tuckwell K, Dimonaco S, et al. Trial of tocilizumab in giant-cell arteritis. *N Engl J Med*. 2017;377:317–328.
11. Weyand CM, Goronzy JJ. Immune mechanisms in medium and large-vessel vasculitis. *Nat Rev Rheumatol*. 2013;9(12):731–740.
12. Pryshchep O, Ma-Krupa W, Younge BR, et al. Vessel-specific Toll-like receptor profiles in human medium and large arteries. *Circulation*. 2008;118(12):1276–1284.
13. Ma-Krupa W, Jeon MS, Spoerl S, et al. Activation of arterial wall dendritic cells and breakdown of self-tolerance in giant cell arteritis. *J Exp Med*. 2004;199(2):173–183.
14. Krupa WM, Dewan M, Jeon MS, et al. Trapping of misdirected dendritic cells in the granulomatous lesions of giant cell arteritis. *Am J Pathol*. 2002;161(5):1815–1823.
15. Weyand CM, Liao YJ, Goronzy JJ. The immunopathology of giant cell arteritis: diagnostic and therapeutic implications. *J Neuroophthalmol*. 2012;32(3):259–265.
16. Weyand CM, Wagner AD, Bjornsson J, et al. Correlation of the topographical arrangement and the functional pattern of tissue-infiltrating macrophages in giant cell arteritis. *J Clin Invest*. 1996;98(7):1642–1649.
17. Weyand CM, Schonberger J, Oppitz U, et al. Distinct vascular lesions in giant cell arteritis share identical T cell clonotypes. *J Exp Med*. 1994;179(3):951–960.
18. Deng J, Younge BR, Olshen RA, et al. Th17 and Th1 T-cell responses in giant cell arteritis. *Circulation*. 2010;121(7):906–915.
19. Weyand CM, Watanabe R, Zhang H, et al. Cytokines, growth factors and proteases in medium and large vessel vasculitis. *Clin Immunol*. 2019;206:33–41.
20. Seko Y, Minota S, Kawasaki A, et al. Perforin-secreting killer cell infiltration and expression of a 65-kD heat-shock protein in aortic tissue of patients with Takayasu's arteritis. *J Clin Invest*. 1994;93(2):750–758.
21. Wen Z, Shimojima Y, Shirai T, et al. NADPH oxidase deficiency underlies dysfunction of aged $CD8^+$ Tregs. *J Clin Invest*. 2016;126(5):1953–1967.
22. Zhang H, Watanabe R, Berry GJ, et al. Immunoinhibitory checkpoint deficiency in medium and large vessel vasculitis. *Proc Natl Acad Sci U S A*. 2017;114(6):E970–E979.
23. Brack A, Rittner HL, Younge BR, et al. Glucocorticoid-mediated repression of cytokine gene transcription in human arteritis-SCID chimeras. *J Clin Invest*. 1997;99(12):2842–2850.
24. Hunder GG, Bloch DA, Michel BA, et al. The American College of Rheumatology 1990 criteria for the classification of giant cell arteritis. *Arthritis Rheum*. 1990;33(8):1122–1128.
25. Dasgupta B, Cimmino MA, Kremers HM, et al. Provisional classification criteria for polymyalgia rheumatica: a European League Against Rheumatism/American College of Rheumatology collaborative initiative. *Arthritis Rheum*. 2012;64(4):943–954.
26. Arend WP, Michel BA, Bloch DA, et al. The American College of Rheumatology 1990 criteria for the classification of Takayasu arteritis. *Arthritis Rheum*. 1990;33(8):1129–1134.
27. Kerr GS, Hallahan CW, Giordano J, et al. Takayasu arteritis. *Ann Intern Med*. 1994;120(11):919–929.
28. Mutoh T, Shisrai T, Ishii T, et al. Identification of two major autoantigens negatively regulating endothelial activation in Takayasu arteritis. *Nat Commun*. 2020;11(1):1253. https://doi.org/10.1038/s41467-020-15088-0.
29. Maleszewski JJ, Younge BR, Fritzlen JT, et al. Clinical and pathological evolution of giant cell arteritis: a prospective study of follow-up temporal artery biopsies in 40 treated patients. *Mod Pathol*. 2017;30(6):M788–M796.
30. Uy CP, Tarkin JM, Gopalan D, et al. The impact of integrated noninvasive imaging in the management of Takayasu arteritis. *JACC Cardiovasc Imaging*. 2020;S1936–878X(20)30434–4. Online ahead of print. https://doi:10.1016/j.jcmg.2020.04.030.
31. Mazlumzadeh M, Hunder GG, Easley KA, et al. Treatment of giant cell arteritis using induction therapy with high-dose glucocorticoids: a double-blind, placebo-controlled, randomized prospective clinical trial. *Arthritis Rheum*. 2006;54(10):3310–3318.
32. Koster MJ, Matteson EL, Warrington KJ. Recent advances in the clinical management of giant cell arteritis and Takayasu arteritis. *Curr Opin Rheumatol*. 2016;28(3):211–217.
33. Hoffman GS, Cid MC, Rendt-Zagar KE, et al. Infliximab for maintenance of glucocorticosteroid-induced remission of giant cell arteritis: a randomized trial. *Ann Intern Med*. 2007;146(9):621–630.
34. Mahr AD, Jover JA, Spiera RF, et al. Adjunctive methotrexate for treatment of giant cell arteritis: an individual patient data meta-analysis. *Arthritis Rheum*. 2007;56(8):2789–2797.
35. Nakaoka Y, Isobe M, Takei S, et al. Efficacy and safety of tocilizumab in patients with refractory Takayasu araeritis: results from a randomized, double-blind, placebo-controlled, phase 3 trial in Japan (the TAKT study). *Ann Rheum Dis*. 2018;77(3):348–354.
36. Sanchez-Alvarez C, Koster M, Duarte-Garcia A, et al. Disease progression of Takayasu arteritis in two patients treated with tocilizumab. *Ann Rheum Dis*. 2020;79(2):e21.https://doi.org/10.1136/annrheumdis-2018-214642.

61

Antiphospholipid Syndrome

Molly M. Daughety, Doruk Erkan, Michael D. Lockshin, and Thomas L. Ortel

Diagnosis of antiphospholipid syndrome (APS) requires that a patient has *both* a clinical event (thrombosis and/or pregnancy loss) *and* persistent antiphospholipid antibody (aPL), documented by a solid phase serum assay (anticardiolipin [aCL] and/or anti-β_2-glycoprotein-I [β_2GPI] enzyme-linked immunosorbent assay) or an inhibitor of phospholipid-dependent clotting (lupus anticoagulant [LA] test), or both (Table 61.1).[1] Antiphospholipid syndrome occurs as an isolated diagnosis, referred to as primary APS, or is associated with other systemic rheumatic diseases such as systemic lupus erythematosus (SLE).

The primary antigen to which aPL binds is β_2GPI (apolipoprotein H), a phospholipid-binding plasma protein. β_2GPI is normally present at a concentration of 200 μg/mL and is a member of the complement control protein family. Structurally, it consists of 326 amino acids arranged in five short consensus repeat domains. *In vivo*, β_2GPI circulates in a circular form that is maintained by interactions between the first and fifth domains. This circular form unfolds into an open form upon binding to phosphatidylserine on membranes of activated or apoptotic cells, including those of trophoblast, platelets, and endothelial cells, exposing the immunogenic site in the first domain. This binding may initiate cell activation, clearance of apoptotic cells by macrophages, and/or coagulation.[2]

KEY CONCEPTS

- aPL exist as a family of autoantibodies directed against phospholipid-binding plasma proteins, most commonly β_2-glycoprotein-I.
- The origin of aPL is unknown but is hypothesized to be an incidental exposure to environmental agents inducing aPL in susceptible individuals.
- In humans, cross-sectional and prospective cohort studies demonstrate that aPL can predict future thrombosis. The pathogenic mechanism is unknown; more than one mechanism may be involved.
- Concomitant prothrombotic risk factors may promote clotting in an additive manner in aPL-positive patients

Autoimmune aPL binds β_2GPI (β_2GPI-dependent aPL), which in turn binds negatively charged phospholipids. Drugs (such as chlorpromazine, procainamide, quinidine, and phenytoin), malignancies (such as lymphoproliferative disorders), and infectious agents (such as syphilitic and non-syphilitic *Treponema, Borrelia burgdorferi*, human immunodeficiency virus, *Leptospira*, or parasites) induce β_2GPI-independent transient aPL. β_2GPI-independent aPL are usually made up of low-titer aCL, which may bind directly to phospholipids and are rarely associated with thrombosis.

EPIDEMIOLOGY

Low-titer aCL occur in less than 10% of normal blood donors, and moderate- to high-titer aCL and/or a positive LA test occurs in greater than 1%. The prevalence of positive aPL tests increases with age. Because the differential diagnosis of vascular occlusion is broader than it is in young adults, particular care is necessary when diagnosing APS in older patients. Thirty to 40% of SLE patients and approximately one-fifth of rheumatoid arthritis patients have positive tests for aPL.

The strength of association between aPL and clinical events varies among studies. Although a number of studies have tried to estimate the annual thrombosis risk in asymptomatic aPL-positive patients, most of these studies have predominantly included patients with SLE. The annual risk of first thrombosis is probably very low (<1%/year) in aPL-positive individuals with no other systemic autoimmune diseases or risk factors for thrombosis, though well-controlled studies are lacking. However, aPL-positive patients with other systemic autoimmune diseases, such as SLE, are at increased annual risk for first thrombosis (<4%/year).

In patients without underlying SLE, it is estimated that approximately 10% of first-stroke victims have aPL, especially those who are young, as do up to 20% of women who have suffered three or more consecutive fetal losses.

ETIOPATHOGENESIS

Antiphospholipid antibodies exist as a family of autoantibodies directed against phospholipid-binding plasma proteins, most commonly β_2GPI. Such proteins bind negatively charged phospholipids (cardiolipin, phosphatidylglycerol, phosphatidylserine, phosphatidylinositol) but not zwitterionic or neutral phospholipids (phosphatidylethanolamine, phosphatidylcholine). Other phospholipid-binding plasma proteins are prothrombin, thrombomodulin, protein C, protein S, and annexins I and V. *In vivo*, the likely relevant phospholipid to which these proteins bind is phosphatidylserine, which is normally sequestered on the inner cell membrane but is exteriorized during cell activation and apoptosis. Annexins bind to exposed phosphatidylserine and create a "remove me" signal on cells.

The commonly accepted hypothesis regarding the origin of aPL states that incidental exposure to environmental agents, the

TABLE 61.1 Revised Sapporo Classification Criteria for the Antiphospholipid Syndrome[1]

Clinical Criteria

1. Vascular thrombosis[a]
   (a) One or more clinical episodes[b] of arterial, venous, or small-vessel thrombosis,[c] in any tissue or organ.
2. Pregnancy morbidity:
   (a) One or more unexplained deaths of a morphologically normal fetus at or beyond the 10th week of gestation, *or*
   (b) One or more premature births of a morphologically normal neonate before the 34th week of gestation because of eclampsia, severe preeclampsia, or recognized features of placental insufficiency[d] *or*
   (c) Three or more unexplained consecutive spontaneous abortions before the 10th week of gestation, with maternal anatomical or hormonal abnormalities and paternal and maternal chromosomal causes excluded.

Laboratory Criteria[e]

1. Lupus anticoagulant present in plasma, on two or more occasions at least 12 weeks apart, detected according to the guidelines of the International Society on Thrombosis and Haemostasis.
2. Anticardiolipin antibody of IgG and/or IgM isotype in serum or plasma, present in medium or high titer (*i.e.*, >40 GPL or MPL, or > the 99th percentile), on two or more occasions, at least 12 weeks apart, measured by a standardized enzyme-linked immunosorbent assay (ELISA).
3. Anti-β_2 glycoprotein-I antibody of IgG and/or IgM isotype in serum or plasma, (in titer >the 99th percentile) present on two or more occasions, at least 12 weeks apart, measured by a standardized ELISA.

Definite antiphospholipid syndrome (APS) is present if at least one of the clinical criteria and one of the laboratory criteria are met. Classification of APS should be avoided if <12 weeks or >5 years separates the positive antiphospholipid antibody (aPL) test and the clinical manifestation. In studies of populations of patients who have more than one type of pregnancy morbidity, investigators are strongly encouraged to stratify groups of subjects according to a, b, or c above.

[a]Coexisting inherited or acquired factors for thrombosis are not a reason for excluding patients from APS trials. However, two subgroups of APS patients should be recognized, according to (a) the presence and (b) the absence of additional risk factors for thrombosis. Indicative (but not exhaustive), such cases include age (>55 in men and >65 in women) and the presence of any of the established risk factors for cardiovascular disease (hypertension, diabetes mellitus, elevated LDL or low HDL cholesterol, cigarette smoking, family history of premature cardiovascular disease, body mass index ≥30 kg/m^2, microalbuminuria, estimated glomerular filtration rate (GFR) <60 mL/min), inherited thrombophilias, oral contraceptives, nephritic syndrome, malignancy, immobilization, and surgery. Thus, patients who fulfill the criteria should be stratified according to contributing causes of thrombosis.
[b]A thrombotic episode in the past could be considered as a clinical criterion, provided that thrombosis is proved by appropriate diagnostic means and that no alternative diagnosis or cause of thrombosis is found.
[c]Superficial venous thrombosis is not included in the clinical criteria.
[d]Generally accepted features of placental insufficiency include (1) abnormal or nonreassuring fetal surveillance test(s), for example, a nonreactive nonstress test, suggestive of fetal hypoxemia; (2) abnormal Doppler flow velocimetry waveform analysis suggestive of fetal hypoxemia, for example, absent end-diastolic flow in the umbilical artery; (3) oligohydramnios, for example, an amniotic fluid index of 5 cm or less; or (4) a postnatal birth weight less than the 10*th* percentile for the gestational age.
[e]Investigators are strongly advised to classify APS patients in studies into one of the following categories: I, more than one laboratory criteria present (any combination); IIa, LA present alone; IIb, aCL antibody present alone; IIc, anti-$\beta_2$glycoprotein-I antibody present alone.

TABLE 61.2 Possible Mechanisms of Antiphospholipid Antibody–Induced Thrombosis

Endothelial Cells–Antiphospholipid Antibody (aPL) Interaction

Endothelial cell damage or activation (via increased expression of adhesion molecules)
Coexisting antiendothelial antibodies
aPL-induced monocyte adhesion to endothelial cells Increased tissue factor expression

Platelet–aPL Interaction

Platelet activation
Stimulation of thromboxane production

Coagulation System–aPL Interaction

Inhibition of activation of protein C by the thrombomodulin–thrombin complex
Inhibition of activation of protein C via its cofactor protein S
Interaction between aPL and substrates of activated protein C such as factors Va and VIIIa Interaction between aPL and an annexin V anticoagulant shield

Complement Activation

Complement activation by aPL provokes thrombosis
Mutations in complement regulatory genes predispose to uncontrolled complement activation

antigens of which contain β_2GPI-like peptides, induces aPL in susceptible individuals via molecular mimicry. In experimental animal models, passive or active immunization with viral peptides, bacterial peptides, and heterologous β_2GPI induces polyclonal aPL and clinical events associated with APS. β_2GPI polymorphisms influence the generation of aPL in individuals but have only a weak relationship to the occurrence of aPL-related clinical events. Persons congenitally lacking β_2GPI, as well as β_2GPI knockout mice, appear normal.

In humans, although cross-sectional and prospective cohort studies demonstrate that aPL can predict future thrombosis, the pathogenic mechanism remains unknown; more than one mechanism may be involved (Table 61.2). Because high-titer antibodies can persist for years in asymptomatic persons and because positive aPL tests can precede symptoms for years, it is likely that vascular injury and/or endothelial cell activation will immediately precede thrombosis in persons bearing the antibody. Platelet activation followed by binding of aPL to platelet membrane phospholipid-bound annexins may initiate platelet adhesion and thrombosis. Antiphospholipid antibodies can inhibit phospholipid-dependent reactions in the coagulation cascade, such as protein C and protein S activation. Interaction between aPL and an annexin A5 anticoagulant shield is another potential mechanism.[2] Also, aPLs induce cellular activation through receptors such as annexin A2 and apoER2 and induce the release of microparticles from endothelial cells.[3]

More recently, the role of complement in thrombotic APS has been examined.[4] In a prospective study, complement activation in APS patients was investigated using surface deposition of C5b-9 and complement-dependent cell killing (modified Ham assay [mHam]). Persistent complement activation was observed in patients with triple aPL positivity, recurrent thromboses, and catastrophic antiphospholipid syndrome (CAPS), suggesting a strong association between complement activation and thrombotic events in APS (see Table 61.2). Furthermore, rare germline mutations in complement regulatory genes were detected in a majority of patients with CAPS, suggesting a predisposition to uncontrolled complement activation leading to a more severe thrombotic phenotype.[4]

In experimental animal models, aPL cause fetal resorption (a proxy for recurrent fetal loss) and increase size and duration of trauma-induced venous and arterial thrombi. Inhibiting

complement activation prevents experimental aPL-induced fetal death, and C5 knockout mice carry pregnancies normally despite the presence of aPL, implying that a complement-mediated effector mechanism is a requirement for fetal death to occur.

Asymptomatic aPL-positive individuals may require a second trigger event (such as oral contraceptives or surgical procedures) to develop a thrombotic event, referred to as the "second-hit hypothesis." Acquired, as well as inherited, risk factors for thrombosis may increase the risk of thromboembolic events in aPL patients.

A proposed pathogenesis for aPL-mediated thrombosis and placental injury begins with activation or apoptosis (by unknown triggers, possibly infectious or traumatic) of platelets, endothelial cells, or trophoblasts. Negatively charged phosphatidylserine migrates from the inner to the outer cell membrane. Circulating β_2GPI binds to phosphatidylserine, which is followed by aPL binding to a β_2GPI dimer, activating a complement and, through C5a, initiating a signaling cascade that induces cell surface tissue factor (TF) expression and adhesion molecules (e.g., ICAM-1). Other components of the innate immune system may also be activated, leading to a milieu that promotes platelet activation and thrombosis. In addition, aPL adversely affects the formation of a trophoblast syncytium, placental apoptosis, and trophoblast invasion, all processes required for the normal establishment of placental function. *In vitro*, pathogenic aPL induce adhesion molecules and enhance adherence of leukocytes to cultured endothelial cells.

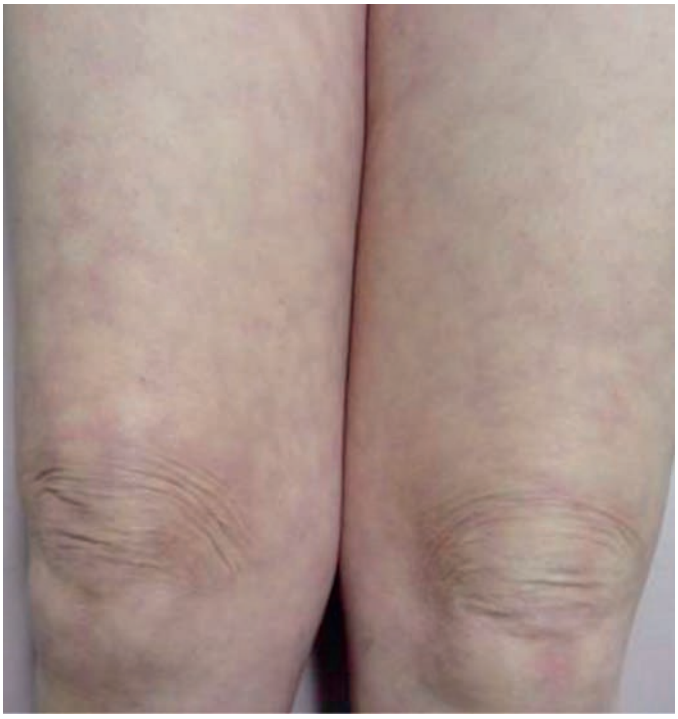


FIG. 61.1 Livedo Reticularis in a Patient With Primary Antiphospholipid Syndrome.

DIAGNOSIS

Clinical Manifestations

The clinical manifestations associated with aPL represent a spectrum from completely asymptomatic to catastrophic APS (Table 61.3). The principal manifestations are venous or arterial thromboses and pregnancy losses. Except for their severity, the young age of affected patients, and atypical anatomical locations (e.g., Budd-Chiari syndrome, sagittal sinus), thromboses in APS do not clinically differ from other thromboses. Stroke and transient ischemic attack are the most common presentation of arterial thrombosis, whereas deep vein thrombosis and pulmonary embolism are the most common venous manifestations of APS. Glomerular capillary endothelial cell injury or thrombosis of renal vessels (thrombotic microangiopathy) causes proteinuria without celluria or hypocomplementemia and may lead to severe hypertension and/or renal failure.

Many patients have livedo reticularis (a lattice-like pattern of superficial skin veins) (Fig. 61.1), cardiac valve disease (vegetations, valve thickening, and dysfunction), or other nonthrombotic manifestations described in several studies, but not included in the revised Sapporo criteria[1] due to nonspecificity or rarity (Table 61.4). These manifestations by themselves do not classify a patient as having APS for clinical studies, but they add information to the diagnosis of individual patients. The pathogenesis of cardiac valve disease in APS is unknown, but it can be severe enough to require valve replacement. A putative association of aPL and increased risk of atherosclerosis exists, but this is controversial.[5] Some patients develop nonfocal neurological symptoms such as lack of concentration, forgetfulness, and dizzy spells. Small hyperintense lesions can be seen on magnetic resonance imaging (MRI), primarily in the periventricular white matter, but do not correlate well with clinical symptoms. Hemorrhagic complications are uncommon in patients with APS but may be seen in patients with severe thrombocytopenia or those who develop hypoprothrombinemia due to the development of antibodies of prothrombin that deplete the protein.

Pregnancy losses in patients with aPL typically occur after 10 weeks gestation, but earlier losses can occur. Patients with

TABLE 61.3 The Clinical Spectrum of Antiphospholipid Antibodies

Asymptomatic[a] antiphospholipid antibody (aPL)–positivity
Antiphospholipid syndrome with vascular events
Antiphospholipid syndrome with only pregnancy morbidity
aPL-positivity with noncriteria clinical manifestations[b]
Catastrophic antiphospholipid syndrome[c]

[a]No history of thrombosis or pregnancy morbidity as per the Sapporo Criteria.[1]
[b]Defined in Table 61.4.
[c]Defined in Table 61.5.

TABLE 61.4 Noncriteria Features of the Antiphospholipid Syndrome

Type	Features
Clinical	Livedo racemose
	Autoimmune hemolytic anemia
	Thrombocytopenia (usually 50,000–100,000/mm^3)
	Multiple sclerosis-like syndromes
Laboratory	Immunoglobulin (Ig)A anticardiolipin and anti-β_2-glycoprotein-I antibodies
	Antiphosphatidylserine, phosphatidylinositol, phosphatidylglycerol, and/or phosphatidylethanolamine antibodies
	Antiprothrombin antibodies
	Antiphosphatidylserine-prothrombin antibodies

TABLE 61.5 Preliminary Criteria for the Classification of Catastrophic Antiphospholipid Syndrome

Criteria

1. Evidence of involvement of three or more organs, systems, and/or tissues[a]
2. Development of manifestations simultaneously or in <1 week
3. Confirmation by histopathology of small-vessel occlusion in at least one organ or tissue[b]
4. Laboratory confirmation of the presence of antiphospholipid antibody (aPL)[c]

Definite Catastrophic Antiphospholipid Syndrome (APS):

- All four criteria

Probable Catastrophic APS:

- Criteria 2–4 and two organs, systems, and/or tissues involved;
- Criteria 1–3, except no aPL confirmation 6 weeks apart due to the early death of a patient not tested before catastrophic episode;
- Criteria 1, 2, 4; or
  - Criteria 1, 3, 4, and development of a third event >1 week but less than 1 month after first despite anticoagulation.

[a]Usually, clinical evidence of vessel occlusions, confirmed by imaging techniques when appropriate. Renal involvement is defined by a 50% rise in serum creatinine, severe systemic hypertension, and/or proteinuria.
[b]For histopathological confirmation, significant evidence of thrombosis must be present, although vasculitis may coexist occasionally.
[c]If the patient had not been previously diagnosed as having APS, laboratory confirmation requires that the presence of aPL must be detected on two or more occasions at least 12 weeks apart (not necessarily at the time of the event), according to the proposed preliminary criteria for the classification of APS.

antiphospholipid syndrome may develop severe early preeclampsia and HELLP (*h*emolysis, *e*levated *l*iver enzymes, *l*ow *p*latelets) syndrome. Although placental infarction may be a cause of fetal growth restriction or death, non-thrombotic mechanisms of placental dysfunction are likely more important.

Catastrophic APS (CAPS) is a rare, abrupt-onset, life-threatening presentation. It is defined by multifocal thromboses associated with multi-organ failure occurring over a period of days in patients who meet laboratory criteria for APS. Proposed diagnostic criteria for CAPS are shown in Table 61.5. Early diagnosis can be a challenge, especially in patients with no history of APS or aPL-positivity, but diagnostic algorithms are available to facilitate diagnosis. Acute adrenal failure may be the initial clinical event, heralded by unexplained back pain and vascular collapse. Patients with CAPS often have moderate thrombocytopenia; fragmented erythrocytes can be seen, but they are less frequent than observed in hemolytic–uremic syndrome or thrombotic thrombocytopenic purpura. Renal failure and pulmonary hemorrhage may occur. Tissue biopsies show noninflammatory vascular occlusions involving both small- and medium-sized vessels.

Laboratory Tests

In the presence of characteristic clinical events, APS is diagnosed when patients have persistent aPLs, including moderate- to high-titer IgG, which may or may not be accompanied by IgM aCL or IgM aβ_2GP, and/or positive lupus anticoagulant (LA) test.[1] Approximately 80% of patients with positive LA tests have aCL, but only 20% of patients with positive aCL have positive LA tests. Some patients are only aβ_2GPI positive. Patients with positive LA tests are at higher risk for thrombosis than patients with aCL and/or aβ_2GPI alone. Several studies have shown that patients with all three tests positive have the highest risk for thrombotic complications, but this has not been consistently observed. LAs are detected in coagulation assays that detect the ability of aPL to interfere with phospholipid-dependent coagulation reactions. Guidelines from the International Society on Thrombosis and Haemostasis for diagnosis of LA[6] include the following:

1. Demonstration of a prolonged phospholipid-dependent coagulation screening test such as activated partial thromboplastin time (aPTT) or dilute Russell viper venom time (dRVVT).
2. Mixing patient plasma with normal plasma failed to correct the prolonged screening test, demonstrating the presence of an inhibitor.
3. Addition of excess phospholipid corrects or shortens the prolonged screening test, demonstrating phospholipid dependence.
4. Exclusion of other coagulopathies.

A positive screening coagulation test without confirmatory steps is not a positive LA test. Patients on anticoagulation may have false-positive or false-negative LA test results; thus, the LA test should be ordered when the patient is not receiving anticoagulation therapy, if possible.[6]

Interpretation of positive tests should take into account the following observations: moderate- to high-titer (>40 U) aCL or aβ_2GPI is more strongly associated with clinical events than is low-titer; LA is a more specific but less sensitive predictor of thromboses than other aPL tests;[7] multiple positive aPL tests impart a worse prognosis than does any single type of test[8]; and positive aPL tests require a repeat test after 12 weeks to exclude transient aPL.[1]

For most patients, laboratory results are stable over time. However, laboratory variability in the performance of these assays can be problematic. We showed that aPL results remain stable for at least three-quarters of subsequent tests during a mean follow-up of 2.4 years for the LA test, 3.5 years for the aCL test, and 1 year for the anti-β_2GPI test.[9] Based on same-day specimens, the consistency of aCL results among different commercial laboratories range from 64% to 88%, with moderate agreement for IgG and IgM, but marginal agreement for aCL IgA.[9]

Antiphospholipid antibody tests developed based on other phospholipids (*e.g.*, phosphatidylserine, phosphatidylinositol, or phosphatidylethanolamine) or prothrombin are not well standardized or widely accepted; their clinical significance is unknown. IgA aCL can rarely occur as the only aPL in patients with APS. When positive, an IgA aCL may justify a diagnosis of APS in LA- and aCL-IgG/IgM test-negative patients with clinically typical disease. A false-positive test for syphilis is not diagnostic for APS.

Antinuclear and anti-DNA antibodies occur in approximately 45% of patients with APS who are not diagnosed as having SLE. Thrombocytopenia occurs in APS and is usually modest (>50, 000/mm^3); proteinuria and renal insufficiency occur in patients with thrombotic microangiopathy. Erythrocyte sedimentation rate and hemoglobin and leukocyte count are usually normal in patients with uncomplicated APS, except during acute thrombosis. Complement levels are usually normal or only modestly low.

Imaging Studies

MRI studies show vascular occlusion and infarction consistent with clinical symptoms, without special characteristics, but multiple, otherwise unexplained, cerebral infarctions in a young person are more suggestive of the syndrome. Multiple, small, hyperintense, white matter lesions are common and do not

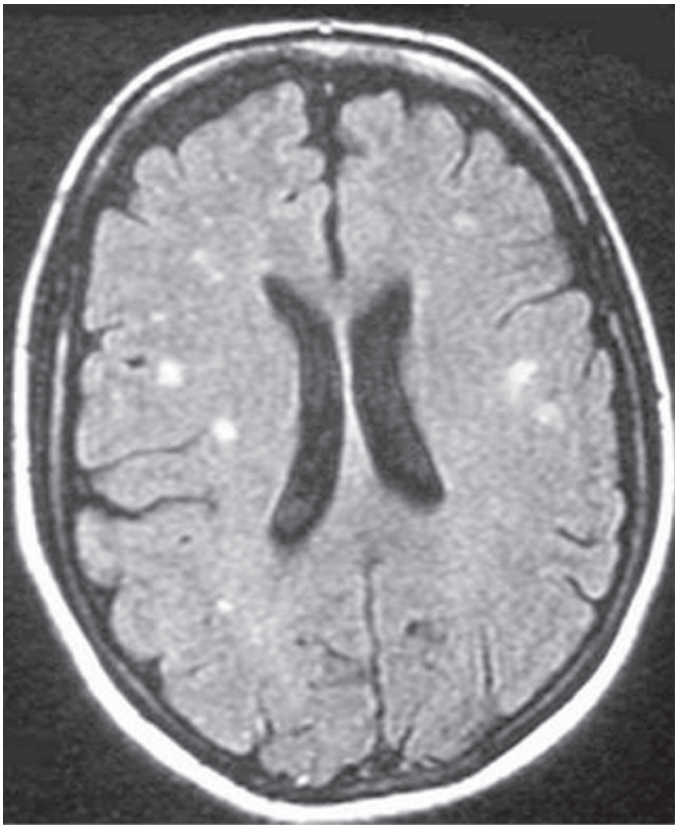


FIG. 61.2 Magnetic Resonance Imaging Demonstrating Multiple Periventricular White Matter Hyperintense Lesions.

unequivocally imply brain infarction (Fig. 61.2). Occlusions frequently occur in vessels below the resolution limits of angiography; hence angiography or magnetic resonance angiography is not indicated unless clinical findings suggest medium- or large-vessel disease. Echocardiography or cardiac MRI may show severe Libman-Sacks endocarditis and intracardiac thrombi.

Pathological Studies

Skin, renal, and other tissues show noninflammatory occlusion of all caliber arteries and veins, acute and chronic endothelial injury and its sequelae, and recanalization in late lesions. The finding of inflammatory necrotizing vasculitis suggests concomitant SLE or another connective tissue disease. There are no other diagnostic immunofluorescence or electron microscopic findings.

CLINICAL PEARLS

- The clinical manifestations of aPL represent a spectrum (from asymptomatic to catastrophic APS).
- Stroke and transient ischemic attack are the most common presentation of arterial thrombosis; deep vein thrombosis, often accompanied by pulmonary embolism, is the most common venous manifestation of APS.
- Pregnancy losses in patients with APS typically occur after 10 weeks gestation, but earlier losses also occur.
- Catastrophic APS is a rare, abrupt, life-threatening complication of APS, which consists of multiple thromboses with multi-organ failure occurring over a period of days.
- APS diagnosis should be made in the presence of characteristic clinical manifestations and *persistently* (at least 12 weeks apart) positive aPL.
- Warfarin, a vitamin K antagonist, is the preferred anticoagulant for patients with APS and a thrombotic event who are triple positive for aPL (*i.e.*, have lupus anticoagulant, anticardiolipin, and anti-β_2-glycoprotein I antibodies)

TREATMENT

Treatment recommendations for persistently aPL-positive patients are determined by the specific clinical indication (Table 61.6).

Asymptomatic Individuals

The ideal strategy for primary thrombosis prevention in asymptomatic, persistently aPL- positive individuals requires a risk-stratified approach based on aPL profile, age, systemic autoimmune diseases, traditional cardiovascular disease, or risk factors for venous thrombosis. Elimination of reversible risk factors for thrombosis (smoking, oral contraceptives) and prophylaxis during high-risk periods (surgical interventions or prolonged immobilization) is crucial for primary thrombosis prophylaxis in persistently aPL-positive individuals.

The role of aspirin for primary prevention of thrombosis in the asymptomatic aPL-positive individual is controversial. In one randomized, double-blind, placebo-controlled trial, low-dose aspirin (LDA) appeared to be no better than placebo in preventing first thrombotic episodes in asymptomatic persistently aPL-positive patients.[10] In contrast, a subsequent meta-analysis of seven retrospective observational studies of asymptomatic aPL carriers found that LDA reduced the risk of first thrombosis by half compared to those who did not use LDA.[11] The European League Against Rheumatism (EULAR) recently recommended the use of LDA in asymptomatic aPL-positive individuals with a high-risk aPL profile.[12] Risk prediction tools and prevention guidelines for cardiovascular disease in the general population, including an assessment of bleeding risk,

TABLE 61.6 Treatment Recommendations in Persistently Antiphospholipid Antibody–Positive Patients

Clinical Circumstances	Recommendation
Asymptomatic	No treatment[a]
Venous or arterial thrombosis	Warfarin international normalized ratio (INR) 2.0–3.0 indefinitely
Recurrent thrombosis	Warfarin INR 3.0–4.0 ± low-dose aspirin; or low molecular weight heparin
First pregnancy	No treatment[a]
Single pregnancy loss, <10 weeks	No treatment[a]
Recurrent fetal loss or loss after 10 weeks; history of no thrombosis	Prophylactic-dose[b] heparin with low-dose aspirin throughout the pregnancy, discontinue heparin 6–12 weeks postpartum
Recurrent fetal loss or loss after 10 weeks; history of thrombosis	Therapeutic-dose heparin[c] with low-dose aspirin throughout pregnancy, warfarin postpartum
Catastrophic antiphospholipid syndrome (APS)	Anticoagulation + corticosteroids + intravenous immunoglobulin or plasma exchange
Livedo reticularis	No treatment
Valve nodules or deformity	No known effective treatment: full anticoagulation if emboli or intracardiac thrombi are demonstrated
Thrombocytopenia, ≥30,000/mm^3	No treatment if asymptomatic with close follow-up and monitoring
Thrombocytopenia, <30,000/mm^3	Prednisone and/or intravenous immunoglobulin

[a]Aspirin 81 mg/day may be given.
[b]Prophylactic dose such as enoxaparin 30–40 mg subcutaneously (SQ) once daily.
[c]Therapeutic dose such as enoxaparin 1 mg/kg SQ twice daily or 1.5 mg/kg SQ once daily.

should be included when considering whether to use aspirin in asymptomatic aPL-positive individuals.

Avoiding unwanted pregnancies is an important part of primary prevention in aPL- positive patients. However, patients should be counseled to avoid using estrogen-containing oral contraceptives due to increased thromboembolic risk. While there are no reliable data regarding the safety of progestin-only contraception in this population, it is theoretically safer than estrogen-based contraception and is frequently used in clinical practice. A small retrospective review of women undergoing artificial reproductive technology procedures demonstrated no thrombotic events.

Venous and Arterial Thromboembolism

Anticoagulation with unfractionated heparin or low-molecular-weight heparin (LMWH) followed by a vitamin K antagonist (VKA) remains the treatment of choice for APS patients with thromboembolic events. For patients with a positive LA test that prolongs the aPTT, monitoring heparin can be accomplished by measuring anti-factor Xa levels.

Two prospective controlled studies concluded that recurrence of thromboses in APS patients could be prevented with VKA titrated to a target international normalized ratio (INR) of 2.0 to 3.0.[13,14] Although these studies provide strong evidence for standard-intensity anticoagulation after an aPL-related venous event, the management of recurrent venous thrombosis while on standard-intensity VKA is challenging. Based on limited evidence, guidelines support first ensuring VKA adherence and increasing frequency of INR monitoring. For VKA-adherent patients with recurrent thrombotic events, an increase of the INR target to 3.0 to 4.0, the addition of LDA, or a change to LMWH may be considered.[12]

The optimal management of APS with arterial thrombosis is not well established. EULAR recently recommended that patients with APS and arterial thrombosis should receive VKA, with or without LDA, to prevent recurrent events.[12] The data in support of this recommendation are limited, however, and others have suggested that aspirin alone (325 mg) could be used in these patients. Recently, a retrospective analysis compared recurrent thrombosis rates in APS patients with prior arterial events and found significant differences between patients treated with antiplatelet therapy (37.2%), anticoagulant therapy (23.7%), and anticoagulation with antiplatelet therapy (6.9%).[15] While these findings suggest that combination therapy is the optimal choice, one must take into account that anticoagulation was not standardized and that bleeding risk was not evaluated. Since the best secondary stroke prevention strategy remains unclear, treatment should be individualized, weighing the risk of recurrent thrombosis with the risk of major bleeding.

Despite the expanding use of direct oral anticoagulants (DOACs) in the treatment of patients with venous thromboembolism and the prevention of stroke in patients with atrial fibrillation, current data do not support DOAC use in APS patients. A randomized multicenter noninferiority study designed to evaluate the efficacy and safety of rivaroxaban compared with warfarin in high-risk (triple antibody positive) patients with thrombotic APS was terminated prematurely due to excessive thrombotic events in the rivaroxaban arm (19% vs. 3%).[16] A second study confirmed that recurrent thrombosis, particularly strokes, occurred more frequently in patients with thrombotic APS treated with rivaroxaban compared to VKA.[17] In addition, this study included all patients with APS, not just those with high-risk, triple-positive disease. Major bleeding events were similar between the two groups. Guidelines now recommend against the use of DOACs in APS patients with triple-positive aPL, particularly those with arterial thrombotic events.[12]

Thrombosis in aPL-positive patients typically has a high recurrence rate if anticoagulation is discontinued, and lifelong anticoagulation is usually recommended. A systematic review of the literature found that the available evidence in support of an association between the presence of aPL and risk of recurrent venous thromboembolism is of low quality, however, suggesting that indefinite anticoagulation may not be needed by all patients.[18] For example, it is unknown whether patients whose event was triggered by an acquired, reversible risk factor for thrombosis can discontinue anticoagulation or switch to aspirin when the trigger factor is eliminated. Normalization of the LA or aCL tests is *not* an indication to discontinue anticoagulation. For well-anticoagulated patients who continue to have thromboses, antiplatelet drugs, hydroxychloroquine, statins, intravenous immunoglobulin (IVIG), and plasmapheresis have theoretical bases for efficacy and have all been used.

There are no systematic studies of treatment for CAPS. Detailed reviews conclude that the most effective first-line treatment combines full-dose anticoagulation, high-dose corticosteroids, and plasmapheresis or IVIG. Concurrent treatment of the underlying precipitating factors is also recommended (e.g., infection or malignancy). Several case reports have described the successful use of complement inhibitor therapy to treat CAPS.

Pregnancy Morbidity

Pregnancy is a prothrombotic state. Management strategies in persistently aPL-positive patients should focus on the prevention of both pregnancy morbidity and maternal thrombotic complications.[5] In patients with a history of obstetric APS, a combination of heparin and low-dose aspirin is recommended.[18] Most experts in the field use LMWH (e.g., enoxaparin) due to lower risk of thrombocytopenia and osteoporosis—prophylactic doses for women with prior obstetric APS only, or full therapeutic doses for women with prior thromboses (see Table 61.6). Treatment begins after confirmation of pregnancy, continues until 48 hours before anticipated delivery (to allow epidural anesthesia), and resumes for 6 to 12 weeks postpartum (if no prior thromboembolism), or the patient is transitioned to warfarin postpartum for continued therapy. No studies unequivocally justify the treatment of women with aPL during a first pregnancy, women with only very early losses, or women whose aPL titers are low or transient. Nonetheless, it is common to offer such patients low-dose aspirin.

Other Clinical Manifestations of Antiphospholipid syndrome

There is no consensus for the treatment of patients with noncriteria and/or nonthrombotic manifestations of aPL. Corticosteroids and/or IVIG are the first-line treatments for severe thrombocytopenia. In general, patients with asymptomatic immune-mediated thrombocytopenia do not need treatment until the platelet count drops to less than 30, 000/mm^3. An open-label phase IIa pilot study of aPL found that rituximab may be effective in controlling some but not all noncriteria manifestations of APS, although aPL profiles did not change with treatment.[19]

Perioperative Management

Serious perioperative complications may occur despite prophylaxis. Patients with APS are at additional risk for

thrombosis when they undergo surgery. Thus, perioperative strategies should be clearly identified before any surgical procedure or before pharmacological and physical anti-thrombosis interventions are vigorously employed. In addition, periods without anticoagulation must be kept to an absolute minimum, and any deviation from a normal course should be considered a potential disease-related event.

Additional Therapeutic Considerations

There is experimental and clinical evidence in lupus patients that hydroxychloroquine (HCQ) might decrease the incidence of thrombosis, and *in vitro* studies have demonstrated that HCQ might protect endothelial cells and syncytialized trophoblast cell lines from the disruptive effect of antiphospholipid antibodies. In patients with systemic autoimmune diseases (particularly lupus), HCQ is commonly employed for disease control and should be considered independent of patients' aPL status. Unfortunately, a study investigating the efficacy of HCQ in primary prevention of thrombosis in aPL-positive patients without underlying autoimmune disease was terminated early due to low recruitment rate and significant cost increases of the study drug in the US. Prospective, randomized, controlled studies continue to be needed to determine the role of HCQ for the prevention and treatment of thrombosis in aPL-positive patients.

As previously discussed, initial trials of rivaroxaban for thrombotic APS were disappointing and failed to show noninferiority to VKA. At this time, providers should use caution recommending DOACs to patients with thrombotic APS, especially those with triple-positive disease or prior arterial thrombosis.

An emerging area of investigation is the role of complement inhibitors in the treatment of CAPS and thrombotic APS refractory to anticoagulation. There are several case reports describing the successful use of eculizumab in refractory APS patients.[20] However, a prospective, randomized clinical trial of complement inhibition in thrombotic APS is needed.

CONCLUSIONS AND TRANSLATIONAL RESEARCH

APS is a systemic autoimmune disease consisting of thromboses, pregnancy losses, and persistent high-titer aPL antibodies. Inflammation and complement activation are established mechanisms of aPL-related manifestations in murine models; however, definitive studies in humans do not exist. The disease is too variable clinically, and its mechanisms too diverse to expect that a single mechanism defined in a single model will apply to all aspects of this disease. Given that the mechanisms of aPL-induced thrombosis are not well understood, thrombosis is multifactorial, and that controversies exist about the strength of association between aPL and clinical events, drug development specific for aPL-positive patients has been challenging. Anticoagulation is the primary treatment today, but a future therapeutic approach will likely include immunomodulatory agents.

ON THE HORIZON

Translational Research in Antiphospholipid Syndrome

- A better understanding of the cellular mechanisms of aPL mediated clinical events that will help design better treatments that are specifically targeted.
- Identification of patients who are at risk for future/recurrent aPL-related events.
- Controlled studies of potentially useful medications such as hydroxychloroquine, complement inhibitors, or anti-B-cell therapies.

REFERENCES

1. Miyakis S, Lockshin MD, Atsumi T, et al. International consensus statement on an update of the classification criteria for definite antiphospholipid syndrome (APS). *J Thromb Haemostas*. 2006;4(2):295–306.
2. Giannakopoulos B, Krilis SA. The pathogenesis of the antiphospholipid syndrome. *N Engl J Med*. 2013;368(11):1033–1044.
3. Betapudi V, Lominadze G, Hsi L, et al. Anti-beta2GPI antibodies stimulate endothelial cell microparticle release via a nonmuscle myosin II motor protein-dependent pathway. *Blood*. 2013;122(23):3808–3817.
4. Chaturvedi S, Braunstein EM, Yuan X, et al. Complement activity and complement regulatory gene mutations are associated with thrombosis in APS and CAPS. *Blood*. 2020;135(4):239–251.
5. Roman MJ, Shanker BA, Davis A, et al. Prevalence and correlates of accelerated atherosclerosis in systemic lupus erythematosus. *N Engl J Med*. 2003;349(25):2399–2406.
6. Pengo V, Tripodi A, Reber G, et al. Update of the guidelines for lupus anticoagulant detection. Subcommittee on Lupus Anticoagulant/Antiphospholipid Antibody of the Scientific and Standardisation Committee of the International Society on Thrombosis and Haemostasis. *J Thromb Haemostas*. 2009;7(10):1737–1740.
7. Galli M, Luciani D, Bertolini G, et al. Lupus anticoagulants are stronger risk factors for thrombosis than anticardiolipin antibodies in the antiphospholipid syndrome: a systematic review of the literature. *Blood*. 2003;101(5):1827–1832.
8. Pengo V, Ruffatti A, Legnani C, et al. Incidence of a first thromboembolic event in asymptomatic carriers of high-risk antiphospholipid antibody profile: a multicenter prospective study. *Blood*. 2011;118(17):4714–4718.
9. Erkan D, Derksen WJ, Kaplan V, et al. Real world experience with antiphospholipid antibody tests: how stable are results over time? *Ann Rheum Dis*. 2005;64(9):1321–1325.
10. Erkan D, Harrison MJ, Levy R, et al. Aspirin for primary thrombosis prevention in the antiphospholipid syndrome: a randomized, double-blind, placebo-controlled trial in asymptomatic antiphospholipid antibody-positive individuals. *Arthritis Rheum*. 2007;56(7):2382–2391.
11. Arnaud L, Mathian A, Ruffatti A, et al. Efficacy of aspirin for the primary prevention of thrombosis in patients with antiphospholipid antibodies: an international and collaborative meta-analysis. *Autoimmun Rev*. 2014;13(3):281–291.
12. Tektonidou MG, Andreoli L, Limper M, et al. EULAR recommendations for the management of antiphospholipid syndrome in adults. *Ann Rheum Dis*. 2019;78(10):1296–1304.
13. Crowther MA, Ginsberg JS, Julian J, et al. A comparison of two intensities of warfarin for the prevention of recurrent thrombosis in patients with the antiphospholipid antibody syndrome. *N Engl J Med*. 2003;349(12):1133–1138.
14. Finazzi G, Marchioli R, Brancaccio V, et al. A randomized clinical trial of high-intensity warfarin vs. conventional antithrombotic therapy for the prevention of recurrent thrombosis in patients with the antiphospholipid syndrome (WAPS). *J Thromb Haemostas*. 2005;3(5):848–853.
15. Jackson WG, Oromendia C, Unlu O, et al. Recurrent thrombosis in patients with antiphospholipid antibodies and arterial thrombosis on antithrombotic therapy. *Blood Adv*. 2017;1(25):2320–2324.
16. Pengo V, Denas G, Zoppellaro G, et al. Rivaroxaban vs warfarin in high-risk patients with antiphospholipid syndrome. *Blood*. 2018;132(13):1365–1371.
17. Ordi-Ros J, Saez-Comet L, Perez-Conesa M, et al. Rivaroxaban versus vitamin K antagonist in antiphospholipid syndrome: a randomized noninferiority trial. *Ann Intern Med*. 2019;171(10):685–694.
18. Garcia D, Akl EA, Carr R, et al. Antiphospholipid antibodies and the risk of recurrence after a first episode of venous thromboembolism: a systematic review. *Blood*. 2013;122(5):817–824.
19. Erkan D, Vega J, Ramon G, et al. A pilot open-label phase II trial of rituximab for non-criteria manifestations of antiphospholipid syndrome. *Arthritis Rheum*. 2013;65(2):464–471.
20. Zikos TA, Sokolove J, Ahuja N, et al. Eculizumab induces sustained remission in a patient with refractory primary catastrophic antiphospholipid syndrome. *J Clin Rheumatol*. 2015;21(6):311–313.

62

Immunohematologic Disorders

Kristin Ammon Shimano and Pierre Noel

INTRODUCTION

Congenital (primary) or acquired (secondary) immunodeficiencies, medications, and lymphoproliferative and rheumatologic disorders are frequently associated with immune-mediated cytopenias. These processes can affect erythrocytes, leukocytes, and platelets, individually or in combination. In this chapter, we address immune-mediated cytopenias of each hematologic component, discussing pathophysiology, clinical presentation, differential diagnosis, and treatment options.

IMMUNE-MEDIATED HEMOLYTIC ANEMIA

Immune-mediated hemolysis can be autoimmune, alloimmune, idiopathic, or secondary to drugs or other diseases. Regardless of the underlying cause, immunoglobulin G (IgG) or IgM (rarely IgA) antibodies are directed against antigens on the red blood cell (RBC) membrane (Table 62.1). These disorders can be categorized on the basis of the underlying cause and the type of anti-erythrocyte antibody that mediates the process. Autoimmune hemolytic anemias (AIHAs) are usually classified as warm or cold depending on the temperature at which the antibody attaches to RBCs. The three major categories of AIHA are warm autoimmune hemolytic anemia, cold agglutinin disease, and paroxysmal cold hemoglobinuria (PCH)—although mixed antibody and direct antibody test (DAT)-negative warm AIHA can also occur.

TABLE 62.1 Classification of Immune Hemolytic Disorders

Autoimmune

Warm Antibody-Mediated

Idiopathic
Secondary
Drugs, lymphoid malignancies, infections
Other autoimmune diseases

Cold Antibody-Mediated

Cold agglutinin disease
Idiopathic
Secondary
Infection, lymphoid malignancies

Paroxysmal Cold Hemoglobinuria

Idiopathic
Secondary to infections

Alloimmune

Secondary to red cell transfusions (alloantibodies, isoantibodies)
Secondary to fetal-maternal hemorrhage
Secondary to transplanted lymphocytes

Immunopathogenesis

The antigens for anti-erythrocyte IgG are usually proteins, the most clinically important of which are the Rh-associated glycoproteins (RhAG), D, C, c, E, and e expressed on the cell membrane.[1] In contrast, anti-erythrocyte IgM is directed at polysaccharides, which include the ABO and I-antigens (I, i) found on the anion and glucose transporter proteins in the RBC membrane.[2,3] Molecular mimicry in the setting of infection, neoantigen formation in the presence of drugs, polyclonal lymphocyte activation due to viral infection, or clonal outgrowth in the setting of lymphoproliferative disease may all contribute to antibody formation in different cases.[4]

Antibodies are distinguished by being "warm" or "cold" reactive, respectively, meaning that they bind antigens at core body temperature (warm), or they bind antigens preferentially at lower temperatures (cold) in the peripheral circulation or ex vivo. This distinction results from the different thermodynamics of binding to protein (hydrophobic) and polysaccharide (electrostatic) antigens.[3]

IgG antibodies typically bind at warm temperatures and IgM antibodies typically bind at cold temperatures, although there can be overlap and exceptions (e.g., the Donath-Landsteiner IgG antibodies that are seen in PCH). IgG and IgM also differ in their ability to fix complement, and this affects the resulting mechanism of hemolysis. To attach the first component of the classical complement pathway, two IgG molecules must bind in close proximity on the RBC. However, because of its pentameric structure, a single IgM molecule can initiate complement activation.

Erythrocyte-bound IgG becomes attached to the Fc receptors of splenic macrophages, which may engulf all or part of the cell or release lysosomal enzymes that digest its membrane (antibody-dependent cell-mediated cytotoxicity [ADCC]).[5] RBC fragments escaping from this encounter lose more membrane than cytoplasm and become spherical (spherocytes) as a consequence of this change in the surface-to-volume ratio. If IgG has initiated complement activation on the cell surface, binding of C3b to splenic macrophages will augment erythrocyte phagocytosis in the spleen.[6]

When IgM fixes complement, the process begins in the cooler peripheral circulation, where IgM binds to RBCs. If the amount of IgM bound is relatively high—with at least some of it remaining on the cell at 37°C (e.g., anti-A or anti-B isoantibodies)—the cascade of complement activation goes to completion. Doughnut-shaped holes are formed in the cell membrane that allow the influx of water and sodium, inducing intravascular osmotic rupture of the cell.[5] However, if the IgM elutes from the RBC as it returns to body core temperature, the complement reaction attenuates. In this circumstance, the components remain on the cell but do not cause intravascular hemolysis. Instead, the cells are cleared by hepatic macrophages via complement-binding sites.[7]

Warm Autoimmune Hemolytic Anemia

Warm autoimmune hemolytic anemia is the most common subtype, accounting for 70% to 80% of AIHAs in adults and ~50% in children.[8] The antibody is typically a polyclonal IgG antibody directed against the Rh complex on the erythrocyte. The hemolysis is predominantly extravascular, as there is little complement activation. Laboratory manifestations include microspherocytes visible on peripheral blood smear that are the result of partial phagocytosis and removal of RBC membrane in the spleen.

Cold Agglutinin Diseases

This type of hemolytic anemia is mediated predominantly by anti-I or anti-i IgM that agglutinates red cells at temperatures well below 37°C. These antibodies engage the complement pathway, resulting in C3b-mediated RBC phagocytosis mainly by Kupffer cells, whereas membrane-associated complex is a minor mechanism at low IgM titers. The severity of the clinical illness depends on the concentration of the IgM and its "thermal amplitude." Thermal amplitude describes the temperature range over which it binds to RBCs. For example, antibodies that exclusively bind at 4°C are only active in vitro, whereas those that bind at greater than 30°C can bind to RBCs as they circulate in the periphery and begin the process of complement fixation, which can persist even as the cells return to body core temperatures. The activity of the IgM is also determined by its relative affinity for the I- and i-antigens, which varies from one individual to the next. Laboratory manifestations include rouleaux—or stacked aggregations of RBCs—seen on peripheral smear due to the agglutination of erythrocytes by the homopentameric cold agglutinins. Clinically, acrocyanosis may be seen as red cells agglutinate in the cooler extremities, and dark urine may be seen due to significant intravascular hemolysis.

Two general types of cold agglutinin disease are recognized: a chronic idiopathic disease presenting in patients over age 50 years caused by monoclonal anti-I IgM, and a transient disease secondary to certain infections (e.g., mycoplasma, Epstein-Barr virus [EBV], cytomegalovirus [CMV]) caused by polyclonal anti-i and anti-I (see Table 62.1). Avoidance of cold environments is important in both categories of cold agglutinin disease. In addition, cold agglutinin disease can be associated with B-cell lymphoproliferative disorders, and this typically is responsive to rituximab and/or rituximab combined with fludarabine.[9]

Paroxysmal Cold Hemoglobinuria

PCH is caused by anti-P IgG that is very effective in fixing complement and producing intravascular hemolysis. The polyclonal IgG is known as the Donath-Landsteiner antibody. Although rare, it is most common in children following a viral illness and can be managed by avoidance of cold. In the past, it was more commonly associated with syphilis (see Table 62.1). There is also an autoimmune variety of PCH that may require immunosuppression with corticosteroids. Splenectomy is not helpful as a consequence of the fact that the hemolysis is intravascular.

Drug-Induced Immune Hemolysis

Over 125 drugs have been associated with immune hemolysis.[10] Cefotetan, ceftriaxone, and piperacillin currently account for over 80% of cases. The prognosis of drug-related immune hemolysis is much better than that of idiopathic hemolysis because the hemolysis stops once the offending drug has been removed. The drugs most commonly associated with fatal hemolytic anemia are cefotetan (8%) and ceftriaxone (6%).[10]

Although the biochemical mechanisms of drug-related immune hemolysis are not completely clear, several hypotheses are generally accepted. Most commonly, complexes of drug and IgG and/or IgM adsorb to the RBC surface and fix complement. The resultant intravascular hemolysis is acute and often severe enough to cause renal failure from toxicities of hemoglobin to renal epithelium.

A second, less common, mechanism develops primarily in patients receiving very high doses of penicillin (rarely used) for at least 1 week. High-titer antipenicillin IgG develops and binds to penicillin that is covalently attached to the RBC membranes. The resultant hemolysis is less acute than that caused by immune complexes but can be life-threatening.

In a third mechanism, a drug stimulates the production of an antibody that reacts with the patient's RBCs independently from the drug. Serologically, this antibody is indistinguishable from an idiopathic autoantibody. This has become rare as the use of the primary causal agent, methyldopa (an antihypertensive medication), has declined. Although these autoantibodies commonly cause positive clinical antibody tests (see below), they rarely cause hemolysis in vivo, and when they do, it usually ceases within 2 weeks of discontinuing the drug.

Diagnosis

With few exceptions, if the mechanism of hemolysis is immune mediated, an anti-RBC antibody can be demonstrated, either on the RBC surface, in serum, or both. With autoimmune hemolysis, IgG or IgM and/or complement components can be identified by a DAT, originally known as a *direct Coombs test* (Fig. 62.1). For this test, a patient's RBCs are washed and suspended in buffer. Surface-bound IgG is detected by adding a divalent anti-IgG antibody, which binds to IgG on adjacent RBCs and agglutinates them into visible aggregates. Because of its pentameric structure, IgM on the cells can cause agglutination without the addition of a second antibody. Even when IgM has been previously eluted from the cell surface as a result of warming in the central circulation, its earlier presence in vivo can be detected by telltale remnants of complement that are fixed to the RBC. In this setting, detection requires the addition of anticomplement antibody (e.g., anti-C3dg).

Alloantibodies can also be detected by the DAT if allogeneic RBCs from a previous transfusion are still circulating. If these have been cleared, however, RBC antibodies can be identified in the patient's serum by adding the serum to a panel of RBCs carrying different antigens. Agglutination is detected as described above; this constitutes the indirect antibody test.

CLINICAL PEARLS

In the work-up of hemolytic anemia, the following clinical laboratory studies provide important clues as to mechanism and may lead to a diagnosis of an immune hemolytic anemia.

Direct antibody test (DAT): The presence of antibodies on the surface of red blood cells (RBCs) suggests an immune-mediated hemolysis. The presence of antibody and complement on the surface of the RBC suggests drug-related hemolysis, whereas the presence of complement alone may suggest an immunoglobulin M (IgM) or cold antibody-related hemolysis.

Peripheral blood smear: Examination of the peripheral smear and various RBC indices (chiefly the mean corpuscular volume [MCV]) give mechanistic clues to the etiology of the hemolytic process. Immune-mediated hemolysis is characterized by spherocytosis, even microspherocytosis in severe cases, as antibody-coated RBCs traversing the reticuloendothelial system assume a spherical form, rather than that of a normal biconcave disc. The appearance of other pathologic forms, such as schistocytes, sickle cells, targets, or tear drop forms (dacrocytes), suggests other causes of hemolysis, such as thrombotic thrombocytopenic purpura (TTP), or mechanical, shear-induced hemolysis (e.g., aortic stenosis), sickle cell disease, thalassemia, or extramedullary hematopoiesis from bone marrow fibrosis, metastasis, or failure. Nucleated RBCs can be seen in any form of hemolytic anemia if it is severe enough.

Reticulocytes: Evaluation of the reticulocyte count indicates whether bone marrow is capable of making new erythrocytes in response to hemolysis.

Lactic acid dehydrogenase (LDH): The LDH is typically elevated with ongoing hemolysis, since LDH is an important housekeeping enzyme found in erythrocytes. LDH is found in cells from all tissues, each having a characteristic isoenzyme form of LDH. It is rarely necessary to distinguish LDH from RBCs from other tissue sources, but LDH isoenzyme 1 is the predominant form found in RBCs.

Bilirubin: As heme is released from RBCs, it is metabolized to bilirubin, which is glycosylated and then excreted via hepatic metabolism. Initially, as large amounts of heme are released and metabolized, the bilirubin is predominantly indirect bilirubin (unconjugated), and then it is converted to direct (conjugated) bilirubin. This can be altered by cholestasis, either from biliary obstruction or hepatic disease or Gilbert disease, or hepatic immaturity in the premature infant or newborn.

Haptoglobin: Haptoglobin is extremely sensitive to even small amounts of hemolysis. Its absence merely confirms that there is a significant hemolysis, but not its extent. The presence of normal haptoglobin effectively rules out significant hemolysis, and the return of measurable amounts of haptoglobin usually signals the end of hemolysis.

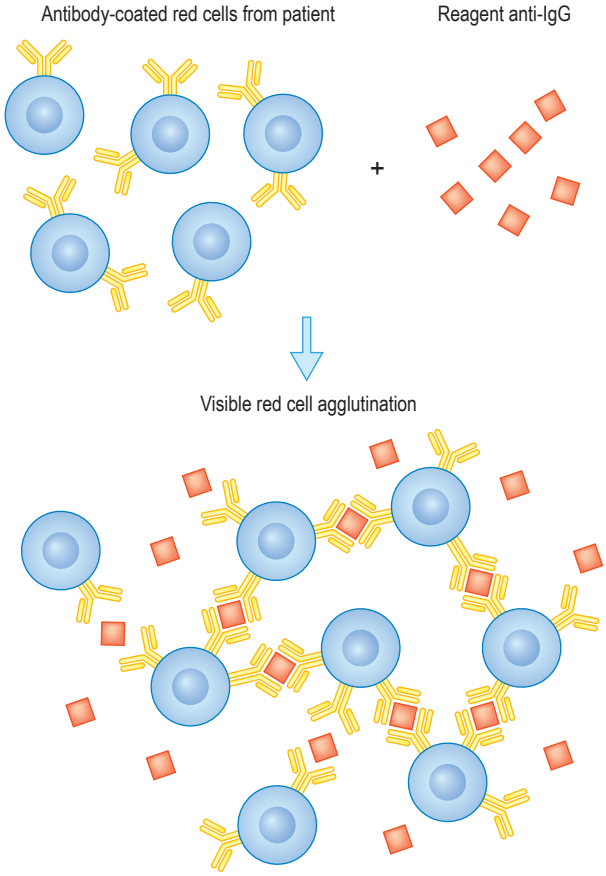


FIG. 62.1 The Direct Antibody Test (DAT). The test is positive when immunoglobulin G (*IgG*: *light blue triangles*)–coated red blood cells are cross-linked by anti-IgG antibody (*dark blue triangles*) to form visible cell aggregates. Cell-bound complement and/or IgM can be detected by using anticomplement or anti-IgM reagent antibodies.

Therapy

The first line of therapy in warm AIHA is corticosteroids, and 80% of patients achieve a partial or complete response to 1 mg/kg/day of prednisone (orally). Once a response is achieved, the prednisone dose is tapered slowly. Approximately 50% of patients require prednisone at a dose of 15 mg/day or less to maintain the hemoglobin level greater than 10 g/dL, and it may take up to 3 weeks for patients to achieve a response. Patients who do not respond in 3 weeks should be started on second-line therapy. It is estimated that long-term complete responses not requiring ongoing prednisone treatment can be achieved in 20% of patients.[11]

Transfusions are avoided except in patients with severe anemia or rapid hemolysis, in which case least incompatible RBC units are used. Splenectomy and anti-CD20 antibody (rituximab) are considered second-line therapy. Splenectomy is associated with short-term partial or complete responses in two-thirds of patients. The overall response rate to rituximab is approximately 80%, but rituximab is contraindicated in patients with untreated hepatitis B. The rare, but most severe, long-term complication of rituximab therapy is progressive multifocal leukoencephalopathy.[11]

The role of high-dose intravenous immunoglobulin (IVIG) remains controversial; its effectiveness remains to be determined in larger trials. Third-line therapy consists of other immunosuppressive agents (e.g., azathioprine, cyclophosphamide, alemtuzumab, mycophenolate mofetil, cyclosporine, sirolimus, or danazol in combination with steroids).[11,12] Plasmapheresis has been used with variable results. If the AIHA is secondary to an underlying disorder, treatment of the underlying disease may offer the most effective therapy.

The treatment of cold AIHA is largely supportive and requires avoidance of cold and pre-warming blood and fluids. Steroids are not useful in the treatment of cold AIHA. Patients with chronic monoclonal IgM-mediated disease may require therapy with rituximab, bortezomib, fludarabine, or

rituximab/bendamustine. Complement inhibition may have a role in therapy, although further study is needed.[13]

Patients with PCH are also treated with supportive therapy, warming, and transfusions as needed, although a subset of patients may respond to steroids.

IMMUNE-MEDIATED NEUTROPENIA

Immune neutropenia constitutes a heterogeneous group of acquired diseases in which the immune system responds to circulating neutrophils, selectively reducing their level to below 1500 cells/mm^3 (Table 62.2).

Neonatal Alloimmune Neutropenia

Transient neutropenia—analogous to neonatal immune hemolysis or thrombocytopenia—develops when IgG antineutrophil antibodies (ANAs) from either an allosensitized or autoimmune mother cross the placenta and destroy fetal neutrophils. The former can occur when there are discrepant maternal and paternal neutrophil-specific antigens, and the mother makes antibodies to paternal antigens that then cross the placenta. The latter occurs less commonly when antibodies from a mother with autoimmune neutropenia cross the placenta. In either case, neutropenia is typically present at birth and resolves spontaneously, as maternal antibodies usually disappear within 12 to 15 weeks, but occasionally it can persist as long as 24 weeks after delivery. It can result in serious infections in approximately 20% of cases.[14]

Primary Autoimmune Neutropenia

Primary autoimmune neutropenia, or benign neutropenia of childhood, is an antibody-mediated disease that presents commonly in early childhood.[15] Patients typically have normal blood counts at birth and develop neutropenia at 3 to 36 months of age. Children presenting at less than 2 years of age most commonly recover spontaneously within 2 to 3 years of diagnosis. The majority do not suffer from severe infections and in most cases are able to mount an appropriate neutrophil response. Nevertheless, some children, particularly those presenting at older ages or manifesting other autoimmune findings, develop a chronic neutropenia disorder. Primary immune neutropenia (in the absence of other disease manifestations) is less common in adults. Rigorous incidence data are lacking, but a small retrospective study found an annual incidence of 5 to 10 cases of primary or secondary neutropenia per 100,000 people.[16]

TABLE 62.2 Causes of Immune Neutropenia

Primary
Isoimmune neonatal neutropenia
Autoimmune neutropenia of childhood
Adult autoimmune neutropenia
Secondary
Systemic autoimmune disease
Rheumatoid arthritis (i.e., Felty syndrome)
Systemic lupus erythematosus
Sjögren syndrome
Lymphoproliferative malignancy
Large granular lymphocyte (LGL) leukemia
Lymphoma
Drug-Induced
Antiplatelets—ticlopidine
Inflammatory bowel disease drug—sulfasalazine
Antipsychotic—clozapine, phenothiazines
Antithyroid medications—propylthiouracil, methimazole
Retrovirals
Antibiotics—beta-lactams, cefepime, trimethoprim-sulfamethoxazole, vancomycin, rifampicin, quinine/quinidine
Diuretics—furosemide, spironolactone
Antiepileptic—lamotrigine
Rituximab, infliximab, etanercept

Neutropenia Associated With Systemic Autoimmune or Lymphoproliferative Diseases

Autoimmune neutropenia is more commonly found secondary to other disorders, such as systemic autoimmune disease or lymphoproliferative diseases, in adults. Secondary autoimmune neutropenia may also be seen in children in the setting of an underlying immunodeficiency or autoimmune disease, or in the setting of other immune cytopenias (Evans syndrome, or multilineage immune cytopenias). Most patients with active systemic lupus erythematosus (SLE) develop neutropenia as part of a more global leukopenia (Chapter 53).[17] Separately, a smaller subset develops severe neutropenia, presumably mediated by anti-neutrophil antibodies. Sjögren syndrome (Chapter 55) and other systemic autoimmune diseases are also sometimes complicated by immune neutropenia. Immune neutropenia develops in about 1% of patients with rheumatoid arthritis, most commonly (but not exclusively) in association with splenomegaly, a combination designated as Felty syndrome.[18] Patients typically express human leukocyte antigen-D related 4 (HLA-DR4) and have long-standing seropositive rheumatoid arthritis complicated by erosive joint disease, subcutaneous nodules, and/or leg ulcers. The natural history is often marked by repeated infection, with 5-year mortality rates of greater than 30% reported in some studies. The complex pathophysiology of this disorder is discussed elsewhere in detail (Chapter 52). Lymphoproliferative disorders, such as chronic lymphocytic leukemia (Chapter 77), can occasionally be complicated by immune neutropenia as well.

T-Cell Large Granular Lymphocyte Leukemia

Large granular lymphocytes (LGLs) are medium to large lymphocytes recognizable on light microscopy by their distinctive azurophilic granules (Fig. 62.2). These cells normally constitute

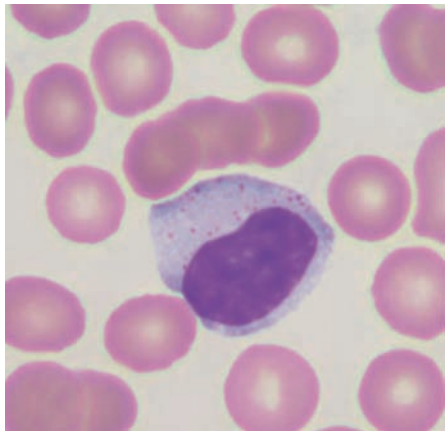


FIG. 62.2 The Large Granular Lymphocyte (LGL) Is a Moderate-Sized Lymphocyte Containing Several Distinctive Azurophilic Granules. Additional immunophenotyping is required to distinguish the CD3$^+$, CD8$^+$, CD57$^+$ T-LGL associated with neutropenia from CD16$^+$, CD56$^+$ natural killer (NK)-LGL.

less than 15% of circulating leukocytes and are composed of two major subsets. One is the natural killer (NK)-LGLs that express CD2, CD16, and CD56 and are not linked to neutropenia. The other, T-cell (T) LGLs, express CD2, CD3, CD8, and CD57 with or without CD16, a phenotype typical of antigen-stimulated mature CD8 effector T cells. Polyclonal and transient monoclonal expansions of these cells sometimes appear in response to viral infection or other immune stimuli without adverse effect. However, some patients develop an indolent lymphoproliferative disease characterized by the accumulation of an autonomous T-LGL clone in blood and other lymphoid organs, particularly bone marrow, the liver, and/or the spleen. Patients with this disease have a remarkably high incidence of immune neutropenia. Even in the absence of gross marrow involvement, over 80% have a neutrophil count of less than 2000/mm^3 at presentation, and at some point, 30% to 40% develop severe neutropenia with less than 500 neutrophils/mm^3.[19,20]

The pathophysiology resembles Felty syndrome in many respects. At the extremes, these syndromes are easily separable. Patients with classical Felty syndrome have severe rheumatoid arthritis, usually requiring antinflammatory therapy, incidentally complicated by late neutropenia. This is quite different from the pattern in patients with isolated T-LGL leukemia with neutropenia in the absence of clinical autoimmune disease. Although some clinicians have attempted to develop criteria for distinguishing Felty syndrome from T-LGL with pseudo-Felty syndrome, there is now substantial evidence that clonal T-LGL disorders are commonly found in rheumatology patients and that patients with clonal disorders seldom develop a progressive, neoplastic disorder. Conversely, although patients with T-LGL leukemia have a malignancy, it is typically quite indolent—and in these cases, the clinical course is often dominated by rheumatologic complications and/or neutropenia and not by progressive neoplastic disease.

Drug-Induced Immune Neutropenia

A wide variety of medications cause neutropenia. In some cases, it may be a direct toxic effect on marrow precursors, but in others it may be antibody-mediated (see Table 62.2 for common examples). Idiosyncratic drug-induced neutropenia is thought to be due to immune-mediated destruction of neutrophils. While the mechanism is not fully understood, antibodies to hapten-carrier complexes or inflammasome activation may play a role in the pathogenesis. Timing of the neutropenia may be months after initiation of the offending medication, or even after discontinuation in some cases, and mortality is around 5%.[21]

Rituximab (anti-CD20) is occasionally associated with late-onset neutropenia (LON). LON is typically self-limiting and of no significant clinical consequence; the occurrence and severity of LON may be associated with the total dose of rituximab and the myelotoxicity of the accompanying chemotherapy administered. LON appears to coincide with post-rituximab B-cell recovery. In varying studies the incidence of LON was 3% to 27%, the median time of onset was 38 to 191 days after rituximab, and the median duration was 5 to 17 days.[22]

Immunopathogenesis

Regulation of Antineutrophil Antibody Production

Isoimmune ANA production in pregnant women represents an "appropriate" immune response to foreign antigens (paternal neutrophil antigens). The production of autoimmune ANAs in other settings reflects immune dysregulation. Older studies attributed primary immune neutropenia of childhood to delayed maturation of T cells responsible for regulating B-cell responses. In this view, the spontaneous recovery usually observed reflects the eventual appearance of mature regulatory or inhibitory T cells. However, this hypothesis is not well supported with clinical laboratory data. Another hypothesis is that ANA production is triggered by molecular mimicry in the setting of a viral infection.[14] The cause of antibody production in other autoimmune diseases remains unclear. A significant proportion of patients with Sjögren syndrome will have neutropenia, with an increased association with anti-Ro/La antibodies.[23] In patients with SLE, as well, anti-SSA and -SSB antibodies are correlated with autoimmune neutropenia, possibly due to similarity between Ro/SSA and neutrophil antigens.[24]

Antibody Specificity

The most important target for antibody responses is the Fcγ receptor IIIb (FcγRIIIb), a low-avidity granulocyte-specific Fcγ receptor that binds IgG immune complexes. This cell surface protein—a glycosyl phosphatidylinositol-linked variant of CD16 selectively expressed on neutrophils—contains several highly immunogenic polymorphisms.[25] Human neutrophil antigen 1a (HNA-1a) and HNA-1b (previously designated NA1 and NA2) are two related polymorphic forms of FcγRIIIb.

Isoimmune neonatal neutropenia is associated with maternal IgG isoantibodies or autoantibodies that can be generated in response to polymorphic alloantigens, particularly polymorphisms affecting FcγRIIIb. Immune neutropenia in childhood is most commonly associated with IgG directed against the autoantigens HNA-1a and/or HNA-1b.[26] Sera from affected patients often also bind (albeit more weakly) to neutrophils expressing the alternative allele, and in some series, more than half the patients with this entity were shown to produce antibodies capable of binding to nonpolymorphic elements within FcγRIIIb. In one study, pan-specific FcγRIIIb antibodies were detected early in the disease course, and more specific HNA-1a and -1b antibodies were detected later.[27] In adults, it is often difficult to distinguish clinically immune neutropenia from nonimmune idiopathic neutropenia. Consequently, the sensitivity and specificity of antibody assays in this setting are uncertain. In general, antibodies against HNA-1a or HNA-1b are less common, and antibodies against surface receptors, such as CD11b/CD18 (CR3), are more common in older children and adults than in young children. Sera from patients with Felty syndrome and T-LGL leukemia are often positive in ANA assays. Interpretation of these results is complicated by the high incidence of nonspecific immune complexes in these populations, which may bind nonspecifically through Fc and complement receptors to the neutrophil surface. Indeed, because it is difficult to distinguish these two types of binding, the incidence of "true" ANAs in these syndromes remains uncertain. Detectable ANAs are low in titer or absent in most patients carefully studied with either diagnosis.[19,20]

Impact of Antibodies and Immune Complexes on Neutrophil Survival

There is quality experimental evidence that both ANAs and immune complexes can induce neutropenia in vivo. The relative importance of reversible sequestration and neutrophil destruction in inducing neutropenia varies with the experimental model, the character of the antibody/immune complex, spleen size, and presumably other factors as well.

The detection of ANAs in serum, however, does not automatically predict accelerated immune clearance of neutrophils in vivo. Some antibodies bind well to neutrophils under in vitro assay conditions without provoking neutrophil destruction in vivo. In part, this reflects the inability of these crude assays to distinguish effective from ineffective binding of immunoproteins.

Myelopoiesis in Immune Neutropenia

In primary immune neutropenia, bone marrow is typically normocellular or mildly hypercellular with an increased proportion of early myeloid forms (particularly myelocytes and promyelocytes) and decreased mature forms (neutrophils, bands, and metamyelocytes), a pattern designated "maturation arrest." Although maturation arrest can also be seen in a number of other diseases, in this setting it suggests an expansion in immature precursors with early release of mature components into blood. Rigorous kinetic studies in children with primary neutropenia are not available, but the available data suggest myelopoiesis in this setting is increased.

The findings are more complex in Felty syndrome and T-LGL leukemia. In vivo neutrophil kinetic studies and in vitro assays of marrow function often document reduced myelopoiesis in these settings.[19,20] This has been attributed to T-cell–mediated and cytokine-mediated suppression. T-LGL leukemia cells constitutively express Fas ligand on their surface and also release significant quantities of soluble Fas ligand into plasma in vivo. Reduced myelopoiesis in T-LGL leukemia and some patients with Felty syndrome may be linked to apoptosis instigated by the binding of Fas ligand expressed on the abnormal cells to Fas expressed on the surface of myeloid precursors.[20] Whatever the precise mechanism, reductions in myelopoiesis appear to be a common element in patients with these forms of secondary immune neutropenia.

Diagnosis

Clinical Presentation

Isoimmune neutropenia presents at birth and may persist for up to 6 months. Patients may have severe infections. Self-limiting primary autoimmune neutropenia typically presents in early childhood, with a median time to resolution of 2 years. Patients rarely have severe infections, and the occurrence of severe infection should prompt evaluation for other causes of neutropenia, such as severe congenital neutropenia or cyclic neutropenia. In older children and adults, neutropenia is more commonly associated with other systemic autoimmune disease, especially rheumatoid arthritis and SLE or T-LGL leukemia. Drug-induced neutropenia must always be considered in patients taking medications.

Laboratory Findings

Blood counts typically demonstrate isolated neutropenia, sometimes with monocytosis. More generalized leukopenia, anemia, and/or thrombocytopenia suggest concurrent SLE or a primary bone marrow disorder, especially aplastic anemia or myelodysplasia.

Examination of the blood film for evidence of abnormalities in other cell lines or increased numbers of LGL is essential. The persistent presence of greater than 2000 LGL/mm^3 for 6 months in itself is diagnostic of T-LGL leukemia; however, normal LGL counts in blood do not preclude this diagnosis. Perhaps a quarter of patients with T-LGL leukemia and immune neutropenia have fewer than 500 monoclonal LGL/mm^3 in blood.[20] The evaluation of patients with small T-LGL clones detected in blood on flow cytometric or molecular testing without clear tissue infiltration remains problematic. At least some of these patients probably have self- limiting "T-cell gammopathies of unknown origin" unassociated with overt lymphoproliferation or autoimmunity.

Bone marrow findings in immune neutropenias (as briefly reviewed above) can vary substantially. Perhaps the most important function of the bone marrow examination is to rule out hypoplasia/aplasia, myelocathexis, marked megaloblastic dysplastic changes, or abnormal infiltration with nonhematopoietic cells, each of which might suggest an alternative diagnosis. The marrow examination may also be helpful in confirming T-LGL leukemia.

Detection of Antineutrophil Antibodies. Antibodies are assayed clinically using indirect assays (i.e., by measuring the binding of antibodies from patient sera to fixed granulocytes from unrelated individuals). The granulocyte immunofluorescence test (GIFT), which exploits flow cytometry for detection, is most commonly used because of its high sensitivity. The granulocyte agglutination test (GAT) is less sensitive but is particularly valuable used in conjunction with GIFT to detect antibodies against HNA-3a or HNA-1b. Once the presence of an antibody has been confirmed, a monoclonal antibody-specific immobilization of granulocytes assay (MAIGA) is a valuable technique in identifying the target molecule recognized by the antibody, information that may be very helpful in identifying antibody specificity and in distinguishing granulocyte-specific antibodies from alloantibodies directed against HLA determinants. More precise epitope typing still requires a panel of granulocytes of varied phenotype. Unfortunately, to date, granulocyte panels are both difficult to prepare and impossible to store. Consequently, antibody typing remains a laborious task. At the second international granulocyte serology workshop, 12 centers independently tested a series of unknown sera. Many laboratories could detect strong HNA-1a antibodies, but the success rate was much lower in defining HNA-1b or HNA-2a antibodies, and individual laboratories varied greatly in their proficiency.[28]

Clinical Use of Antineutrophil Antibody Studies. Because of the lack of both sensitivity and specificity of ANA assays, primary autoimmune neutropenia in young children with neutropenia remains a clinical diagnosis. Using GIFT or GAT assays, ANAs can be detected in more than 70% of children with primary immune neutropenia. When both are used in tandem, the yield increases further. A strong positive result strongly supports the diagnosis of immune neutropenia. However, a negative result does not exclude the diagnosis.[29]

Primary autoimmune neutropenia in adults is difficult to distinguish from the ill-defined entity chronic idiopathic neutropenia. Because there is no "gold standard" for distinguishing immune disease from nonimmune disease in this setting, the diagnostic sensitivity and specificity of the ANA assays are unclear. Assays are positive in perhaps a third of adults referred with chronic neutropenia, and a positive result in the absence of other systemic autoimmune disease certainly supports a diagnosis of immune neutropenia. A negative result does not preclude an immune etiology, but it is more consistent with chronic idiopathic neutropenia.

In patients with systemic autoimmune disease or T-LGL leukemia, hyperglobulinemia and circulating immune complexes greatly complicate laboratory evaluation. ANA assays are frequently positive, even in the absence of neutropenia. Since the specificity of a positive result is low, its diagnostic value is very limited, and the clinician must be vigilant for other possible causes, especially drug-induced neutropenia.

Therapy

Overview

All patients with neutrophil counts below 1000/mm^3 have some increased risk of infection, but some remain asymptomatic even with absolute neutrophil counts of 500/mm^3 or less. Growth factors can usually improve neutropenia and reduce the risk of infection, but given their expense, inconvenience, and possible side effects, they should be reserved for use in patients with a very low count or a previous pattern of frequent infection. The indications for immunosuppressants, steroids, and splenectomy are more complex.

THERAPEUTIC PRINCIPLES

Immune Neutropenia

- Palliative treatment of neutropenia is reserved for patients with a neutrophil count below 500/mm^3 or recurrent infection.
- Recombinant granulocyte colony-stimulating factor (G-CSF) is the most effective single agent for palliating neutropenia.
- Immunosuppressive agents, steroids, and splenectomy are reserved for patients with persistent or refractory neutropenia or with other detrimental manifestations of systemic autoimmunity.

Colony-Stimulating Factors

Controlled trials are lacking in this disease setting, but granulocyte colony-stimulating factor (G-CSF) or granulocyte–macrophage colony-stimulating factor (GM-CSF) usually enhances neutrophil counts in each of the clinical groups discussed. Because of their safety, speed, and efficacy, they have replaced steroids and splenectomy as first-line symptomatic therapy. They should be used at the lowest effective dose.

Infants with alloimmune neutropenia may develop severe infections, and in those cases, G-CSF is indicated until resolution of the neutropenia. Because patients with primary autoimmune neutropenia typically have a benign course, they rarely require growth factor support. For the rare patients who do develop severe infections, or in the perioperative setting, however, G-CSF may be indicated, starting at a dose of 1 to 2 mcg/kg/day.[30]

Immunosuppressive Agents

Because disease is usually self-limiting and responsive to G-CSF when necessary, immunosuppressive agents are not used in children with primary immune neutropenia. Pediatric patients with secondary immune neutropenia will typically also respond to G-CSF, although immunosuppressants may be indicated if they have multilineage cytopenias or other autoimmune phenomena. Chronic low-dose methotrexate is considered first-line therapy for patients with Felty syndrome or LGL leukemia–associated neutropenia. Cyclophosphamide and cyclosporine are considered second-line therapy for LGL leukemia, and purine analogues, splenectomy, and alemtuzumab are considered third-line therapy for LGL leukemia.

Other Therapy

Each of the following treatments can be effective in reversing neutropenia, but their use has diminished considerably in recent years. IVIG can temporarily reverse neutropenia, particularly in children, most likely by blocking Fc receptors responsible for triggering neutrophil destruction. However, G-CSF, which is more convenient to administer and at least as effective, has largely replaced IVIG.

Splenectomy and steroids can each reduce immune destruction by suppressing the body's capacity to clear IgG- and complement-coated cells. Over a longer time frame, these treatments can also suppress antibody production—in the first case by removing a major site of production and in the second by reducing ANA production and blocking T-cell–mediated myelosuppression. Steroid therapy can reverse neutropenia in many patients, but steroids' long-term impact on outcome remains unclear. Given their risks and side effects, both splenectomy and steroids are generally reserved for patients resistant to CSFs and low-dose immunosuppressants.

Prophylactic Antibiotics

Where recurrent infection is a problem, oral trimethoprim–sulfamethoxazole (TMP-SMX) is commonly used for prophylaxis, particularly in children. This approach is very reasonable, given its success in other immunocompromised groups, but it has not been tested in a controlled trial. Immunization with pneumococcal and meningococcal vaccines is also recommended in situations where therapeutic splenectomy has been used or is being planned.

IMMUNE-MEDIATED THROMBOCYTOPENIAS

Immune Thrombocytopenia

KEY CONCEPTS

Immune Thrombocytopenic Purpura

- Antibody-mediated destruction or decreased production of platelets
- Both B and T cells important in etiology
- Clearance of platelets mediated by Fcγ receptors
- Diagnosis depends on:
  - Clinical presentation of low platelets
  - Exclusion of other causes of thrombocytopenia

Immune thrombocytopenia (ITP) is an autoimmune syndrome involving antibody and cell-mediated destruction of platelets. ITP can occur in the absence of an identified predisposing factor (primary ITP) or in the setting of an underlying immunodeficiency, immune dysregulatory syndrome, or autoimmune disease, drug exposure, infection, or other identified cause (secondary ITP). Definitions from an international working group recommend that a platelet count of $<100 \times 10^9$/L be required for diagnosis.[31]

Pathogenesis

In 1951, Dr. William Harrington infused plasma from patients with chronic ITP into himself and other healthy volunteers, resulting in transient thrombocytopenia in most of them, implicating a plasma-derived factor, subsequently identified by Shulman as IgG, the causative agent.[32] In the majority of patients, the underlying defects leading to autoantibody production

remain unclear. In some patients, ITP follows exposure to viral or bacterial antigens. Molecular mimicry appears to play a role in the development of self-reactive platelet antibodies following infections. Human immunodeficiency virus (HIV), hepatitis C virus (HCV), and *Helicobacter pylori* infections have been associated with ITP. *H. pylori* CagA antigen appears to cross-react with platelet antigens,[33] which may explain the association with ITP and *H. pylori* infection in adults. The most common epitopes for platelet antibodies in ITP are the platelet GPIIb/IIa and GPIb-IX receptors.[34] The autoantibodies serve as opsonins, resulting in the clearance of platelets by FcγR-bearing cells in the reticuloendothelial system through Syk-mediated phagocytosis (Fig. 62.3).[35] Antibodies are identified in only a subset of patients with ITP, however, and other immune processes also contribute to the pathogenesis of platelet destruction.[36] There is upregulation of genes involved in cell-mediated cytotoxicity via $CD3^+CD8^+$ T lymphocytes,[37] with T-helper 1 (Th1)–associated cytokines predominating.[38] Regulatory T cells (Tregs) are decreased,[39] and B-cell activation is increased.[40] There is evidence of suppression of megakaryopoiesis by both T lymphocytes[41] and ITP plasma/IgG.[42] In addition, megakaryocyte and platelet production are dependent on thrombopoietin signaling through binding to the Mpl receptor, and patients with ITP have reduced thrombopoietin levels despite the presence of low platelet counts.[43]

Laboratory Diagnosis

Immune-mediated thrombocytopenia is a diagnosis of exclusion and is typified by isolated thrombocytopenia. The presence of abnormalities in other cell lines should prompt investigation for other etiologies, including marrow failure syndromes, myelodysplastic syndrome, or leukemia. Peripheral blood smear examination is important to evaluate for the presence of schistocytes, leukocyte adhesion bodies in MYH-9-related disease, small platelets in Wiskott-Aldrich syndrome, and giant platelets in inherited macrothrombocytopenias and to exclude ethylenediaminetetraacetic acid (EDTA)–dependent platelet agglutination.

The diagnosis of ITP may be made based upon history, physical exam, blood counts, and peripheral smear exam. Bone marrow aspirate, biopsy, flow cytometry, and cytogenetics should be considered in patients older than 60 years of age and in patients with systemic symptoms. The detection of *H. pylori* with the urea breath test or stool antigen tests should be considered in adults. Routine serologic evaluation for HIV and HCV is recommended in adults. Baseline Igs (IgG, IgA, IgM) should be measured in adults to diagnose such conditions as common variable immunodeficiency (CVID) and selective IgA deficiency (Chapter 33) in which ITP is a common complication. For children who go on to have chronic ITP, a broader immune evaluation, including quantitative immunoglobulins, should be investigated.[44]

Antiplatelet antibody testing to specific platelet glycoproteins is not routinely recommended; the test is neither sensitive nor specific for ITP. Routine testing for anticardiolipin antibodies is not recommended in the absence of symptoms of antiphospholipid syndrome. DAT should be obtained if hemolytic anemia is suspected and in patients for whom anti-D antiglobulin treatment is considered. Blood group Rh(D) typing should be obtained if anti-D antiglobulin treatment is considered.

Where a microangiopathic process is evident—especially with concomitant renal failure, fever, or cognitive impairment—measurement of ADAMTS13 (A Disintegrin-like And Metalloprotease with ThromboSpondin type 13 motifs) is warranted to rule out TTP.

Therapy

In pediatric patients, the decision to treat ITP is guided primarily by the presence or absence of bleeding symptoms. Most pediatric patients without bleeding may be safely managed with close observation, while in adult patients, in most circumstances, the goals of therapy are to keep the platelet count above 30×10^9/L and to minimize toxicity. In the absence of hemostatic comorbidity, trauma, or surgery, intracranial hemorrhage is rare in patients with a platelet count above 20×10^9/L. Treatment of ITP should be individualized, but first-line therapies include corticosteroids, IVIG, or anti-Rh(D) in Rh(D)-positive, nonsplenectomized patients. Intravascular hemolysis, disseminated intravascular coagulation, and renal failure have been reported with the use of anti-Rh(D). Anti-Rh(D) should be avoided in patients with underlying hemolysis or a positive result of the DAT that is not the result of prior therapy.[36,44,45]

In adult patients, second-line therapy should be considered if there is absence of a robust response at 3 months or once significant steroid-related toxicity supervenes. Second-line therapy should be considered in pediatric patients who do not respond to first-line therapies or who have chronic ITP. The thrombopoietin receptor agonists (TPO-RAs) avatrombopag, eltrombopag, and romiplostim have shown significant sustained activity in patients with ITP. Romiplostim is administered subcutaneously and avatrombopag and eltrombopag orally. Increased bone marrow reticulin fibrosis has been described in some patients receiving TPO-RAs, though this is typically reversible upon discontinuation. Two-thirds of patients obtain a durable long-term remission following splenectomy. Patients should receive immunizations with the pneumococcal, *Haemophilus influenzae* type b, and the quadrivalent meningococcal vaccines before splenectomy. A single course of rituximab (anti-CD20) (375 mg/m^2 weekly for 4 weeks) is associated with 40% complete remission at 1 year and 15% to 20% complete remission at 5 years. Patients who relapse after an initial response usually respond to a second course. Rituximab is contraindicated in patients with active hepatitis B. Cases of multifocal leukoencephalopathy have been reported in HIV-negative patients treated with rituximab. Fostamatinib is an oral spleen tyrosine kinase (Syk) inhibitor recently approved as a second-line agent for use in adults with ITP.[46]

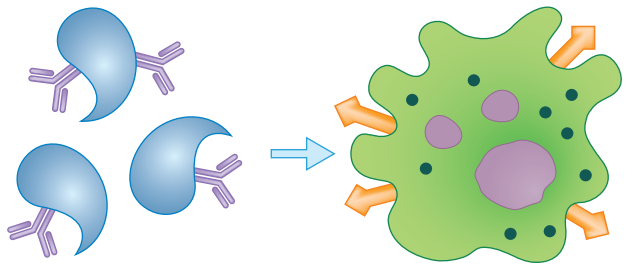


FIG. 62.3 Antibody-Bound Platelets Are Cleared From the Circulation by Binding of the Fc Domain of Immunoglobulin G (IgG) to FcγIIIA Receptors on Macrophages and Other Cells. The cross-linking of the macrophage receptors sets off a cascade of internal signaling that leads to increased expression of the inhibitory FcγIIB receptors (not shown).

A number of other therapies have been used with varying success in patients who are unresponsive to or ineligible for first- and second-line therapies. These agents include azathioprine, cyclophosphamide, cyclosporine, danazol, dapsone, mycophenolate mofetil, vinca alkaloids, and other immunosuppressants or combination therapies.

Drug-Induced Thrombocytopenia

Drug-induced thrombocytopenia (DITP) is an idiosyncratic immune-mediated reaction. The drug-dependent antibodies bind to specific epitopes on platelet surface glycoproteins only in the presence of the sensitizing drug. The drugs bind noncovalently and reversibly to platelets, commonly to GPIIb-IIIa and GPIb-V-IX, and also to the antibody. The Fab domains of the antibodies bind to the drug-platelet epitope. Drug-dependent antibodies inducing thrombocytopenia typically develop 1 to 2 weeks after exposure to a drug; exceptions to this rule include eptifibatide, tirofiban, and abciximab, as naturally occurring antibodies to these drugs can cause thrombocytopenia within a few hours of the first exposure. Thrombocytopenia with platelet counts frequently below 20×10^9/L develops acutely, recovery occurs 1 to 2 days after discontinuation of the drug and is usually complete after 1 week, but rarely thrombocytopenia persists for several weeks. Quinidine, quinine, rifampin, tegretol, TMP-SMX, vancomycin, danazol, acetaminophen, abciximab, eptifibatide, tirofiban, and gold salts are the most common culprits. Treatment consists of discontinuing the offending drug; platelet transfusions are sometimes necessary.[47]

Heparin-induced thrombocytopenia (HIT) is a special case that is caused by antibodies to platelet factor 4 (PF4)–heparin complexes. It can be associated with life-threatening thrombosis. The antibody-PF4-heparin complex activates platelets, resulting in a high risk of both arterial and venous thrombotic events. Thrombocytopenia occurs as a result of clearance of platelet aggregates induced by the antibody and usually appears 5 to 7 days after treatment with heparin (or low-molecular-weight heparin) unless a patient has been previously exposed to heparin. In the event of prior exposure, especially within the last 100 days, thrombocytopenia can occur within 1 day of heparin administration. Even small doses of heparin given as "flushes" to maintain intravenous catheter patency can be sufficient to cause HIT with thrombosis. Suspected HIT can be evaluated using the 4Ts clinical score (timing, degree of thrombocytopenia, presence of thrombosis, other etiologies for thrombocytopenia), an immunoassay, and a functional assay.[48] In cases of suspected HIT, all heparins should be stopped, and an alternative anticoagulant agent, such as the direct thrombin inhibitors argatroban or bivalirudin, should be used. Fondaparinux is an anticoagulant synthesized from the pentasaccharide core of the heparin molecule and may be another option for the treatment of HIT, since it binds to HIT antibodies but does not activate platelets and cause thrombosis. In cases of confirmed or strongly suspected HIT, anticoagulation should be continued at least until platelet recovery and for up to 3 months, because the risk of thrombosis persists in this group of patients. Direct oral anticoagulants that target thrombin or activated factor X or vitamin K antagonists may be used for long-term anticoagulation.

Neonatal Alloimmune Thrombocytopenia

Neonatal alloimmune thrombocytopenia (NAIT) is caused by maternal antibodies against the HPAs that the fetus carries but which the mother lacks (most commonly HPA-1a). NAIT caused by antihuman platelet antigen 1a (HPA-1a) antibodies occurs in 1 in 1250 pregnancies in the Caucasian population. Severe hemorrhage occurs in 1 in 12,500 to 1 in 25,000 pregnancies. NAIT is caused by maternal antibodies against paternally derived antigens on fetal platelets, most commonly HPA-1a. These antibodies cross the placenta and sensitize fetal HPA-1a–positive platelets, which are then removed in the spleen. Two percent of Caucasian women carry the less frequent HPA-1b and can be immunized against HPA-1a during pregnancy (25%) or at the time of delivery (75%). In most circumstances, first-time cases of NAIT are generally identified following the birth of a markedly thrombocytopenic neonate; antenatal management is, thus, only possible in subsequent pregnancies. There is risk of severe bleeding, including intracerebral hemorrhage in 10% to 20% of cases, which may occur in utero or postnatally, and it is very likely to recur if it has occurred in a previous pregnancy. Screening all pregnancies for NAIT is under evaluation in several countries. Antenatal management of NAIT consists of maternal administration of high doses of IVIG and corticosteroids.[49] NAIT may be diagnosed by parental platelet antigen typing and detection of alloantibodies in maternal serum that bind to specific paternal human platelet antigens not present on maternal platelets. Treatment of infants with NAIT includes transfusions with typed platelets if available, random-donor platelets, or administration of IVIG.[50]

AUTOIMMUNE MULTILINEAGE CYTOPENIAS

Some patients develop either simultaneous or sequential cytopenias of more than one cell lineage. Most frequently, this includes AIHA + ITP, but can also include AIN. This entity was described in 1949 by Evans and Duane[51] and has historically been referred to as Evans syndrome. The presence of multi-lineage cytopenias should prompt an evaluation for an underlying immunodeficiency or immune dysregulation syndrome. In a study of patients presenting in childhood, over half of patients with combined autoimmune cytopenias had evidence of secondary disease.[52] Autoimmune lymphoproliferative syndrome (ALPS) or ALPS-like disorder, combined immunodeficiency (CID), CVID, and other humoral defects were most common. A significant number also had non-hematologic manifestations of autoimmune disease.

The implications for treatment are salient as well, as targeted therapies may be warranted for treatment of immune cytopenias in specific underlying disorders. Sirolimus and mycophenolate mofetil are often effective for chronic control of cytopenias in ALPS. Costimulatory blockade with CTLA4-Ig may be used in patients with CTLA4 haploinsufficiency or LRBA deficiency, and interleukin-6 (IL-6) blockade or JAK inhibitors may be used in patients with gain-of-function mutations in the STAT pathway. As the genetic underpinnings of immune dysregulatory disorders are uncovered, it is likely that the therapies available to manage immune cytopenias in these settings will continue to evolve as well.

REFERENCES

1. Avent ND, Reid ME. The Rh blood group system: a review. *Blood*. 2000;95(2):375–387.
2. Mollison PL, Engelfriet CP, Contreras M. Immunology of red cells. In: Mollison PL, Engelfriet CP, Contreras M, eds. *Blood Transfusion in Clinical Medicine*. 10 ed. London: Blackwell Science; 1997:60.
3. Hughes-Jones NC. Red-cell antigens, antibodies and their interaction. *Clin Haematol*. 1975;4(1):29–43.

4. Barcellini W. New insights in the pathogenesis of autoimmune hemolytic anemia. *Transfus Med Hemother*. 2015;42(5):287–293.
5. Petz LD, Garratty G. Mechanisms of immune hemolysis. In: Petz LD, Garratty G, eds. *Immune Hemolytic Anemias*. Philadelphia, PA: Churchill Livingstone; 2004:133.
6. Kurlander RJ, Rosse WF, Logue GL. Quantitative influence of antibody and complement coating of red cells on monocyte-mediated cell lysis. *J Clin Invest*. 1978;61(5):1309–1319.
7. Frank MM, Schreiber AD, Atkinson JP, et al. Pathophysiology of immune hemolytic anemia. *Ann Intern Med*. 1977;87:210–222.
8. Kalfa TA. Warm antibody autoimmune hemolytic anemia. *Hematology Am Soc Hematol Educ Program*. 2016;2016(1):690–697.
9. Berentsen S, Randen U, Vagan AM, et al. High response rate and durable remissions following fludarabine and rituximab combination therapy for chronic cold agglutinin disease. *Blood*. 2010;116(17):3180–3184.
10. Garratty G. Immune hemolytic anemia associated with drug therapy. *Blood Rev*. 2010;24(4–5):143–150.
11. Lechner K, Jager U. How I treat autoimmune hemolytic anemias in adults. *Blood*. 2010;116(11):1831–1838.
12. Barros MM, Blajchman MA, Bordin JO. Warm autoimmune hemolytic anemia: recent progress in understanding the immunobiology and the treatment. *Transfus Med Rev*. 2010;24(3):195–210.
13. Gertz MA. How I treat cold agglutinin hemolytic anemia. *Clin Adv Hematol Oncol*. 2019;17(6):338–343.
14. Farruggia P. Immune neutropenias of infancy and childhood. *World J Pediatr*. 2016;12(2):142–148.
15. Capsoni F, Sarzi-Puttini P, Zanella A. Primary and secondary autoimmune neutropenia. *Arthritis Res Ther*. 2005;7(5):208–214.
16. Beatty PA, Stroncek DF. Autoimmune neutropenia in Sheboygan County, Wisconsin. *J Lab Clin Med*. 1992;119(6):718–723.
17. Starkebaum G. Chronic neutropenia associated with autoimmune disease. *Semin Hematol*. 2002;39(2):121–127.
18. Balint GP, Balint PV. Felty's syndrome. *Best Pract Res Clin Rheumatol*. 2004;18(5):631–645.
19. Mohan SR, Maciejewski JP. Diagnosis and therapy of neutropenia in large granular lymphocyte leukemia. *Curr Opin Hematol*. 2009;16(1):27–34.
20. Lamy T, Loughran Jr. TP. How I treat LGL leukemia. *Blood*. 2011;117(10):2764–2774.
21. Curtis BR. Non-chemotherapy drug-induced neutropenia: key points to manage the challenges. *Hematology Am Soc Hematol Educ Program*. 2017;2017(1):187–193.
22. Wolach O, Bairey O, Lahav M. Late-onset neutropenia after rituximab treatment: case series and comprehensive review of the literature. *Medicine (Baltimore)*. 2010;89(5):308–318.
23. Brito-Zeron P, Soria N, Munoz S, et al. Prevalence and clinical relevance of autoimmune neutropenia in patients with primary Sjögren's syndrome. *Semin Arthritis Rheum*. 2009;38(5):389–395.
24. Afzal W, Owlia MB, Hasni S, Newman KA. Autoimmune neutropenia updates: etiology, pathology, and treatment. *South Med J*. 2017;110(4):300–307.
25. Bux J. Molecular nature of granulocyte antigens. *Transfus Clin Biol*. 2001;8(3):242–247.
26. Bruin MC. von dem Borne AE, Tamminga RY, et al. Neutrophil antibody specificity in different types of childhood autoimmune neutropenia. *Blood*. 1999;94(5):1797–1802.
27. Bruin M, Dassen A, Pajkrt D, et al. Primary autoimmune neutropenia in children: a study of neutrophil antibodies and clinical course. *Vox Sang*. 2005;88(1):52–59.
28. Bux J, Chapman J. Report on the second international granulocyte serology workshop. *Transfusion*. 1997;37(9):977–983.
29. Kobayashi M, Nakamura K, Kawaguchi H, et al. Significance of the detection of antineutrophil antibodies in children with chronic neutropenia. *Blood*. 2002;99(9):3468–3471.
30. Fioredda F, Calvillo M, Bonanomi S, et al. Congenital and acquired neutropenias consensus guidelines on therapy and follow-up in childhood from the Neutropenia Committee of the Marrow Failure Syndrome Group of the AIEOP (Associazione Italiana Emato-Oncologia Pediatrica). *Am J Hematol*. 2012;87(2):238–243.
31. Rodeghiero F, Stasi R, Gernsheimer T, et al. Standardization of terminology, definitions and outcome criteria in immune thrombocytopenic purpura of adults and children: report from an international working group. *Blood*. 2009;113(11):2386–2393.
32. Imbach P, Kuhne T, Signer E. Historical aspects and present knowledge of idiopathic thrombocytopenic purpura. *Br J Haematol*. 2002;119(4):894–900.
33. Takahashi T, Yujiri T, Shinohara K, et al. Molecular mimicry by *Helicobacter pylori* CagA protein may be involved in the pathogenesis of *H. pylori*-associated chronic idiopathic thrombocytopenic purpura. *Br J Haematol*. 2004;124(1):91–96.
34. McMillan R. Antiplatelet antibodies in chronic adult immune thrombocytopenic purpura: assays and epitopes. *J Pediatr Hematol Oncol*. 2003;25(Suppl 1):S57–S61.
35. Newland A, Lee EJ, McDonald V, Bussel JB. Fostamatinib for persistent/chronic adult immune thrombocytopenia. *Immunotherapy*. 2018;10(1):9–25.
36. Cooper N, Ghanima W. Immune Thrombocytopenia. *N Engl J Med*. 2019;381(10):945–955.
37. Olsson B, Andersson PO, Jernas M, et al. T-cell-mediated cytotoxicity toward platelets in chronic idiopathic thrombocytopenic purpura. *Nat Med*. 2003;9(9):1123–1124.
38. Panitsas FP, Theodoropoulou M, Kouraklis A, et al. Adult chronic idiopathic thrombocytopenic purpura (ITP) is the manifestation of a type-1 polarized immune response. *Blood*. 2004;103(7):2645–2647.
39. Yu J, Heck S, Patel V, et al. Defective circulating CD25 regulatory T cells in patients with chronic immune thrombocytopenic purpura. *Blood*. 2008;112(4):1325–1328.
40. Emmerich F, Bal G, Barakat A, et al. High-level serum B-cell activating factor and promoter polymorphisms in patients with idiopathic thrombocytopenic purpura. *Br J Haematol*. 2007;136(2):309–314.
41. Olsson B, Ridell B, Carlsson L, et al. Recruitment of T cells into bone marrow of ITP patients possibly due to elevated expression of VLA-4 and CX3CR1. *Blood*. 2008;112(4):1078–1084.
42. McMillan R, Wang L, Tomer A, et al. Suppression of in vitro megakaryocyte production by antiplatelet autoantibodies from adult patients with chronic ITP. *Blood*. 2004;103(4):1364–1369.
43. Aledort LM, Hayward CP, Chen MG, et al. Prospective screening of 205 patients with ITP, including diagnosis, serological markers, and the relationship between platelet counts, endogenous thrombopoietin, and circulating antithrombopoietin antibodies. *Am J Hematol*. 2004;76(3):205–213.
44. Provan D, Arnold DM, Bussel JB, et al. Updated international consensus report on the investigation and management of primary immune thrombocytopenia. *Blood Adv*. 2019;3(22):3780–3817.
45. Neunert C, Terrell DR, Arnold DM, et al. American Society of Hematology 2019 guidelines for immune thrombocytopenia. *Blood Adv*. 2019;3(23):3829–3866.
46. Bussel J, Arnold DM, Grossbard E, et al. Fostamatinib for the treatment of adult persistent and chronic immune thrombocytopenia: Results of two phase 3, randomized, placebo-controlled trials. *Am J Hematol*. 2018;93(7):921–930.
47. Aster RH, Curtis BR, McFarland JG, Bougie DW. Drug-induced immune thrombocytopenia: pathogenesis, diagnosis, and management. *J Thromb Haemost*. 2009;7(6):911–918.
48. Cuker A, Arepally GM, Chong BH, et al. American Society of Hematology 2018 guidelines for management of venous thromboembolism: heparin-induced thrombocytopenia. *Blood Adv*. 2018;2(22):3360–3392.
49. Vinograd CA, Bussel JB. Antenatal treatment of fetal alloimmune thrombocytopenia: a current perspective. *Haematologica*. 2010;95(11):1807–1811.
50. Baker JM, Shehata N, Bussel J, et al. Postnatal intervention for the treatment of FNAIT: a systematic review. *J Perinatol*. 2019;39(10):1329–1339.
51. Evans RS, Duane RT. Acquired hemolytic anemia; the relation of erythrocyte antibody production to activity of the disease; the significance of thrombocytopenia and leukopenia. *Blood*. 1949;4(11):1196–1213.
52. Al Ghaithi I, Wright NA, Breakey VR, et al. Combined autoimmune cytopenias presenting in childhood. *Pediatr Blood Cancer*. 2016;63(2):292–298.

63

Bullous Diseases of the Skin and Mucous Membranes

Ralf J. Ludwig and Enno Schmidt

INTRODUCTION

Autoimmune blistering diseases (AIBDs) comprise a heterogeneous group of disorders that are characterized by autoantibodies deposited in the skin and adjacent mucous membranes. Autoantibody deposition leads to blisters and/or erosions in these tissues. AIBDs can be divided into two subgroups according to the anatomy of split formation. Pemphigus diseases demonstrate intraepidermal/epithelial splitting and pemphigoid disorders and dermatitis herpetiformis are characterized by subepidermal blistering.[1] This distinction dates back to 1953 when Walter Lever described intraepidermal split formation and loss of cell adherence between neighboring keratinocytes (acantholysis) as histopathological hallmarks of pemphigus, and subepidermal splitting disorders as *pemphigoid*.[2] In the 1960s, serum and skin-bound autoantibodies were described in pemphigus and pemphigoid diseases. In 1975, dermatitis herpetiformis was distinguished from the pemphigoid disease linear immunoglobulin A (IgA) dermatosis based on the pattern of IgA deposit in the skin.[1] Building on these observations, additional pemphigoid and pemphigus entities have been described, the target antigens of most AIBDs identified, and the pathogenic relevance of autoantibodies determined in various in vitro and mouse models.[3–5]

In this chapter, the target antigens, genetics, risk factors, comorbidities, and diagnosis of the three main pemphigus and six main pemphigoid diseases as well as dermatitis herpetiformis will be discussed. Photographs of these entities will be presented and approaches to management outlined. Characteristics of the rarer AIBD not detailed below are summarized in Table 63.1.

TARGET ANTIGENS AND DISEASE ENTITIES

Pemphigus Diseases

Pemphigus diseases comprise three major forms: pemphigus vulgaris, pemphigus foliaceus, and paraneoplastic pemphigus. Pemphigus vulgaris and pemphigus foliaceus account for more than 90% of pemphigus cases. Immunopathologically, pemphigus disorders are characterized by autoantibodies against structural proteins of the epidermal/epithelial desmosomes. Desmosomes are adherence structures that connect neighboring cells, including keratinocytes, to withstand mechanical forces. Desmoglein (Dsg) 1 and 3 are in integral parts of epidermal/epithelial desmosomes and have been identified as the major target antigens in pemphigus (Fig. 63.1).[6] Dsgs contain five extracellular domains (EC1 to EC5), a single transmembrane domain, and a cytoplasmic domain containing plakoglobin and plakophilin binding sites. The Dsg molecules are interconnected with each other via their N-terminal EC1 and EC2 domains, which are preferential targets of pemphigus autoantibodies.

Interestingly, in the great majority of pemphigus vulgaris and pemphigus foliaceus patients, the clinical phenotype is mirrored by the targeted Dsg molecule. While in pemphigus foliaceus, lesions are limited to the skin and the autoantibody response is restricted to DSG1, patients with the mucosal-dominant type of pemphigus vulgaris preferentially generate anti-DSG3 autoantibodies. In patients with the mucocutaneous type of pemphigus vulgaris. which presents with both mucosal and skin lesions, autoantibodies against both DSG1 and DSG3 are present. The remarkable relation of the autoantibody specificity and the clinical phenotype reflects the differences in the expression

TABLE 63.1 Characteristics of Very Rare Autoimmune Blistering Diseases

Disease	Main Target Antigen	Characteristics
Pemphigus Diseases		
Neonatal pemphigus	DSG3	Transient flaccid blisters and erosions of the newborn caused by the transplacental passage of anti-Dsg IgG from a mother with pemphigus
IgA pemphigus	DSC1, 2, 3, DSG1, 3	Also called intercellular IgA dermatosis; flaccid pustules; intercellular IgA staining of the epithelium by direct IF
Pemphigus herpetiformis	DSG1	Pruritic grouped vesicles, papules, and erythema; spared mucous membranes
Pemphigoid Diseases		
Lichen planus pemphigoides		Preceding and concomitant lichen planus; blisters and erosions independent of lichen planus lesions; treatment needs also to be directed against lichen planus that triggers the AIBD
Cicatricial pemphigoid	BP180 laminin 332	Mucous membranes not predominantly affected; lesions heal with scarring
Bullous systemic lupus erythematosus (SLE)	Type VII collagen	Preceding and concomitant SLE; blisters and erosions not restricted to LE lesions; typically in sun-exposed areas in woman with African background in the 3rd decade; rapid response to dapsone

AIBD, Autoimmune blistering disease; *BP180 NC16A*, extracellular part of the 16th non-collagenous domain of the 180 kDa bullous pemphigoid antigen; *Dsc*, desmocollin; *Dsg*, desmoglein; *IF*, immunofluorescence; *SLE*, systemic lupus erythematodes.
See references 1, 4, and 5.

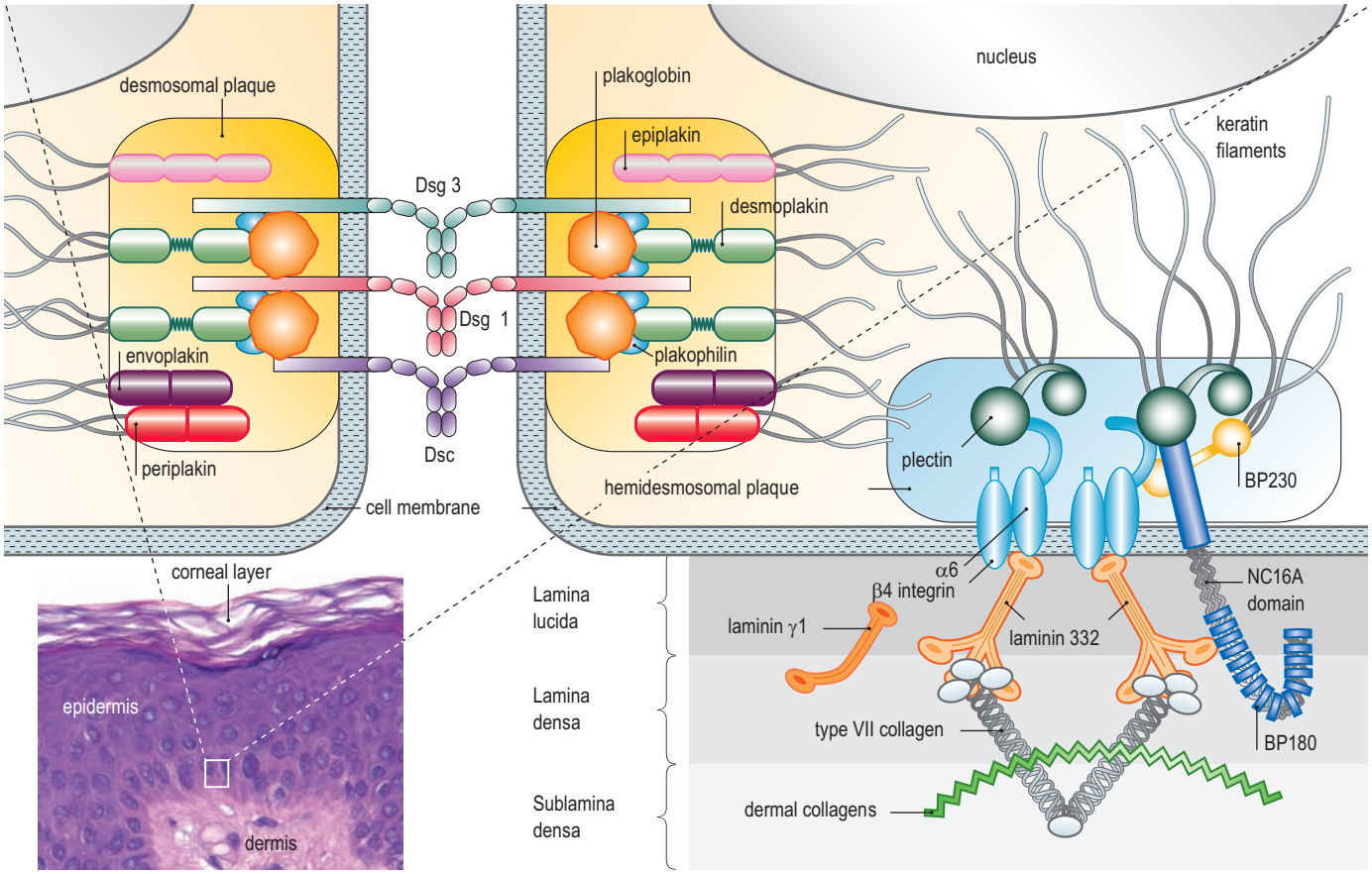


FIG. 63.1 Schematic diagram of the desmosome *(left)* and the dermal–epidermal junction (right). Only antigens that are targeted in autoimmune blistering diseases are shown. While only homophilic transinteractions between desmogleins *(Dsg)* and desmocollins *(Dsc)* are depicted, heterophilic interactions have also been described. Modified from Reference 31.

of the two Dsgs in the epidermis and surface-close mucosal epithelia. The Dsg compensation theory holds that when the adhesive property is compromised in one of them and they are expressed together in the same cell, the normal Dsg isoform can compensate for the other.[4,7]

In pemphigus vulgaris, in addition to antibodies against DSG1 and DSG3, reactivity against desmocollins (see Fig. 63.1), and various other molecules such as muscarinic and nicotinic acetylcholine receptors, pemphaxin, mitochondrial proteins, and thyroid peroxidase have been detected.[8,9]

In paraneoplastic pemphigus, autoantibodies mainly target DSG3 and the plakin proteins envoplakin and periplakin (see Fig. 63.1). In addition, autoantibodies against DSG1, desmocollins, the plakin proteins, desmoplakin I/II, plectin, epiplakin, and BP230 as well as the hemidesmosomal protein BP180 (type XVII collagen) and the 170 kDa proteinase inhibitor α_2-macroglobulin-like 1 have been reported with varying frequencies (see Fig. 63.1, Table 63.2).[4,10,11]

Pemphigoid Diseases and Dermatitis Herpetiformis

Six main pemphigoid diseases can be distinguished. In each, a target antigen has been identified on the molecular level, The autoantibodies recognize proteins at the dermal–epidermal junction (DEJ). The different diseases are distinguished by clinical appearance (e.g., mucous membrane pemphigoid [MMP] and pemphigoid gestationis), by the target antigens (e.g., anti-p200 pemphigoid and epidermolysis bullosa acquisita), and by the autoantibody isotype (e.g., linear immunoglobulin [IgA] disease). In bullous pemphigoid (BP), diagnosis is based on the combination of the three factors (see Table 63.2).

The DEJ, the cutaneous basement membrane zone of the skin (Chapter 24), connects the epidermis with the dermis. Four subregions can be distinguished by transmission electron microscopy: the basal cell plasma membranes of basal keratinocytes and the hemidesmosomal plaques, the lamina lucida, the lamina densa, and the sublamina densa. Hemidesmosomes are multiprotein focal adherens structures within the DEJ and contain the intracellular hemidesmosomal plaque and the anchoring filaments of the lamina lucida. Hemidesmosomes connect the intermediate filaments of the cytoskeleton of the basal keratinocyte with the dermal collagens via anchoring fibrils that mainly consist of type VII collagen (see Fig. 63.1).[12] Traditionally, pemphigoid diseases have been regarded as diseases with anti-hemidesmosomal antibodies, excluding epidermolysis bullosa acquisita, which is characterized by reactivity against type VII collagen. Here, we follow the more recent concept of pemphigoid disorders as diseases with autoantibodies against structural proteins of the DEJ, and thus include epidermolysis bullosa acquisita (see Table 63.2).[5]

A central molecule of the DEJ and main target antigen of different pemphigoid diseases is BP180, also termed BP antigen-2 and type XVII collagen. The 180 kDa BP180 is a transmembrane

TABLE 63.2 Autoantibody Specificities and Diagnostic Clues

Disease	Target Antigens	Diagnostic Clues
Pemphigus Diseases		
Pemphigus vulgaris	***DSG3***, *DSG1*	Mucosal lesions, positive Nikolsky sign, positive direct IF, serum antibodies against DSG3
Pemphigus foliaceus	***DSG1***	Crusted erosions, *no* mucosal lesions, positive Nikolsky sign, positive direct IF, serum antibodies against DSG1
Paraneoplastic pemphigus	***envoplakin***, **periplakin**, ***DSG3***, desmoplakin I/II, plectin, epiplakin, *BP230*, *BP180*, DSC1, 2, 3, *DSG1*, α_2-macroglobulin-like 1	Severe stomatitis, neoplasm, dyskeratosis and interface dermatitis, serum antibodies against plakins
Pemphigoid Diseases		
Bullous pemphigoid	BP180 ***NC16A***, *BP230*	Intense pruritus, tense blisters and erosions without predominant mucosal involvement, old age, serum IgG against BP180 NC6A; non-bullous variants may arise in up to 20% of patients
Mucous membrane pemphigoid	**BP180**, ***laminin 332***, *BP230*, laminin 311,[a] ($\alpha_6\beta_4$ integrin)[b]	predominant mucosal involvement
Pemphigoid gestationis	BP180 ***NC16A***, *BP230*	Pregnancy or postpartum period
Linear IgA disease	**LAD-1**, BP230 (IgA reactivity)	Tense blisters and erosions without predominant mucosal involvement; linear IgA staining at the DEJ without reactivity against type VII collagen; most frequent AIBD in children
Anti-p200 pemphigoid	p200 protein, laminin γ1	Tense blisters and erosions without predominant mucosal involvement, serum antibodies agaisnt the p200 protein
Epidermolysis bullosa acquisita	type VII ***collagen***	Mechanobullous and inflammatory variants; no predominant mucosal involvement, serum antibodies against type VII collagen, u-sertaed pattern by direct IF
Dermatitis herpetiformis	***transglutaminase 3***, *transglutaminase 2*	Intense pruritus, erythematous papules and vesicles; granular IgA deposits in dermal papillae by direct IF

[a]Laminin 311 shares the α_3 chain with laminin 332; antibodies against the β_1 and γ_1 chains have not been described.
[b]There is insufficient evidence for the recognition of $\alpha_6\beta_4$ integrin as target antigen in mucous membrane pemphigoid except for individual patients.
Main target antigens are indicated in **bold**. For target antigens in *italics*, commercial detection systems are available. Adopted from references 4 and 5.
AIBD, Autoimmune blistering disease; *DEJ*, dermal–epidermal junction; *Dsc*, desmocollin; *Dsg*, desmoglein; *IF*, immunofluorescence; *IgA*, immunoglobulin A; *LAD-1*, linear IgA disease antigen 1 (identified as soluble ectodomain of BP180).

glycoprotein of about 1500 amino acids that spans the lamina lucida before kinking back from the lamina densa into the lamina lucida (see Fig. 63.1).

In dermatitis herpetiformis, autoantibodies are directed against two enzymes, epidermal transglutaminase (TG3) and tissue type TG (TG2), with TG3 being the autoantigen.

KEY CONCEPTS

Target Antigens

- Most target antigens of AIBD have been described on the molecular level.
- The main target antigens of pemphigus are desmoglein 1 and 3 (DSG1 and 3).
- The main target antigens of pemphigoid diseases are BP180 (type XVII collagen), laminin 332, type XVII collagen, and an incompletely defined p200 protein.

EPIDEMIOLOGY

Incidence

The incidence of AIBD differs between populations. The most frequent AIBD in Central Europe and North America is BP. In central Europe and Scotland, the incidence of BP was calculated to be between 12 and 22 per million per year. A higher incidence of about 70 per million per year has been reported in the UK and Sweden. Moreover, in the UK, Germany, and France, the incidence of BP has at least doubled within the last 10 years, which may reflect the increasing age of the general population, the improved availability and quality of diagnostic assay, and the high association of BP with neurological and psychiatric diseases that have also been increasing.[5] The annual incidences of MMP and pemphigoid gestationis are approximately 2 per million and 1 per million, respectively. Linear IgA disease, anti-p200 pemphigoid, and epidermolysis bullosa acquisita are even rarer.

The incidence of pemphigus ranges from less than 1 per million per year in Switzerland and Finland to about 10 per million per year in Greece and Iran. The Jewish population has the highest incidence.[4] Pemphigus vulgaris arises three to five times more frequently than pemphigus foliaceus. In some rural areas of South America and North Africa much higher incidences of pemphigus foliaceus have been observed. During the 20th century, the prevalence of the so-called endemic pemphigus foliaceus reached 3% to 5% of the population in certain areas of South America, but more recently has considerably declined.[13]

Incidences of dermatitis herpetiformis are even more diverse in different populations. While extremely rare in individuals of Asian or African origin, the incidence in the UK and Scandinavia is 10 per million per year, with approximately 10% of patients with celiac diseases (Chapter 75) suffering from dermatitis herpetiformis. More recently, the incidence of dermatitis herpetiformis has decreased even though the incidence of celiac disease has increased. This may be explained by the increased awareness and use of screening tests for the latter disease leading to earlier treatment and thus reduced occurrence of dermatitis herpetiformis.[14]

Genetics

In contrast to pemphigoid diseases, a strong genetic susceptibility is seen with certain Human Leukocyte Antigen (HLA) class II

alleles (Chapter 5) in pemphigus and dermatitis herpetiformis.

*DRB1*04:02* and *DQB1*05:03* have been described as risk alleles for pemphigus vulgaris across several populations,[15] with the majority of pemphigus vulgaris patients expressing one of the two alleles. In pemphigus foliaceus, HLA DRB1*04 was the most frequently reported susceptibility allele.[4]

An additional four non-HLA genes have also been associated with pemphigus vulgaris: *DSG3* encoding for the pemphigus vulgaris autoantigen DSG3, *TAP2* encoding for an ATP-binding cassette transporter involved in antigen presentation (Chapter 6), *IL6* encoding for the pleiotropic cytokine interleukin-6 (IL-6) (Chapter 14), and *ST18* encoding a transcription factor involved in inflammation and apoptosis that is overexpressed in pemphigus vulgaris skin.[4,15]

In BP, only few genetic studies have been performed, reporting weak associations with *DQB1*03:01*, *DRB1*04:03*, and *DQA1*05:01*. A recent study from Germany found significant associations with *DQA1*05:05, DQA1*02:01*, and *DQB1*03:01*, as well as with the *ZNF385D* and *FGF14* genes that encode two transcription factors (unpublished). Susceptibility to MMP has been described with *DQB1*03:01*, *DRB*04*, and *DRB*11*, and with *GALC* encoding for the lysosomal enzyme β-galactocerebrosidase.[16] In linear IgA disease, associations with HLA *B8*, *Cw7*, *DR3*, *DR2*, and *TNF2* were found. In pemphigoid gestationis, there are associations with *DR3, DR4*, and the so-called C4 null allele. And in epidermolysis bullosa acquisita, there are associations with *DRB1*15:03*.[1]

Of all AIBD, dermatitis herpetiformis has the highest genetic association, with 5% to 10% of patients with first-degree relatives affected by celiac disease or dermatitis herpetiformis. In nearly all patients, either HLA-*DQ2* (combination of *DQA1*05:01* and *DQB1*02*) or *DQ8* (combination of *DQA1*03* and *DQB1*03:02*) can be found.[14]

Risk Factors

The high prevalence of pemphigus foliaceus in certain areas of South America and North Africa suggests the existence of specific triggering factors. In endemic pemphigus foliaceus, the salivary protein LJM11 of the sand fly *Lutzomyia longipalpis* was found to be cross-reactive with DSG1.[17] In non-endemic pemphigus, drugs such as penicillamine and captopril, exposure to pesticides, metal vapor, ultraviolet light, ionizing radiation, burns, surgery, and stressful life events are risk factors for pemphigus.[4]

In BP, a strong association with the use of the oral anti-diabetics dipeptidyl-peptidase IV inhibitors, in particular with vildagliptin, is observed.[18,19] The association with spironolactone and phenothiazines with aliphatic side chains is less pronounced. Not unsurprisingly, BP has developed in patients receiving immune checkpoint inhibitors. Trauma, burns, radiotherapy, UV radiation, and vaccination have also been associated with disease onset.[5] Linear IgA disease and epidermolysis bullosa acquisita can also be triggered by drugs, most frequently vancomycin and penicillins.[1]

Comorbidities

Pemphigus vulgaris and pemphigus foliaceus are associated with a variety of diseases. These include autoimmune disorders (e.g., autoimmune thyroid diseases and rheumatoid arthritis), neurological disorders, psoriasis, and hematological and solid malignancies.[4]

A strong association in BP is seen with neurological or psychiatric diseases. Indeed, between a third and half of all BP patients suffer from latter disorders, which include major cognitive impairment, Parkinson disease, stroke, epilepsy, and multiple sclerosis.[20] These findings are particularly intriguing since in the great majority of patients the neurological disease precedes BP and both target antigens, BP180 and BP230, are expressed in the central nervous system. This suggests that CNS diseases may trigger the skin manifestations.[5,20] Furthermore, an association has been shown between hematological malignancies and BP and, to a lesser extent, epidermolysis bullosa acquisita.[21] Concomitant Crohn disease has been reported in 10% to 20% of AIBD patients.[1] Psoriasis occurs in about a third of Japanese patients with anti-p200 pemphigoid.[22] Of note, about 25% of anti-laminin 332-reactive MMP patients develop a solid malignancy. Thus MMP warrants a throughout tumor search. Since laminin 332 is overexpressed in many solid tumors and involved in tumor spreading one may hypothesize that the tumor can trigger the AIBD.[1]

Mortality

Before the introduction of corticosteroids in the early 1950s, mortality in pemphigus vulgaris and pemphigus foliaceus was about 75%. At present, the risk of death is about two- to threefold higher than control populations. Infections, in particular pneumonia and septicemia, are the most frequent causes of death, followed by cardiovascular diseases and peptic ulcer disease. In paraneoplastic pemphigus, the 5 years mortality is two to three times higher than pemphigus vulgaris.[4] In BP, the 1-year mortality rate of between 15% and 40% is two- to threefold higher than age- and sex-matched controls and ranges.[5]

KEY CONCEPTS

Epidemiology

- Bullous pemphigoid is the most frequent AIBD.
- The incidence of bullous pemphigoid has increased fourfold over the last 20 years.
- Pemphigus and bullous pemphigoid are associated with other autoimmune disease and hematological malignancies (in pemphigus) and with neurological diseases (in bullous pemphigoid). Pemphigus vulgaris and dermatitis herpetiformis have a strong genetic background, with associations to *DRB1*04:02* and *DQB1*05*:03 in the former and HLA-*DQ2* and *DQ8* in the latter disease.

CLINICAL PRESENTATIONS

Clinical presentations may vary considerably between the different AIBDs and even between patients with the same disease. The unifying clinical hallmarks are blisters and/or erosions on the skin and/or surface-close epithelia. In some patients, in particular with BP, pemphigoid gestationis, and dermatitis herpetiformis, blisters and erosions can be absent and pruritic papules, urticarial erythema, and eczematous lesions prevail. The sex distribution is roughly equal in all AIBD with the obvious exception of pemphigoid gestations.[1,4,5] Children are rarely affected.[23] In-depth descriptions of the broad range of clinical pictures can be found in standard dermatological textbooks.[1,23,24]

Pemphigus

Pemphigus Vulgaris

The oral cavity, which shows enanthema, erosions, and ulcers (Fig. 63.2, A), is the main affected site in pemphigus vulgaris.

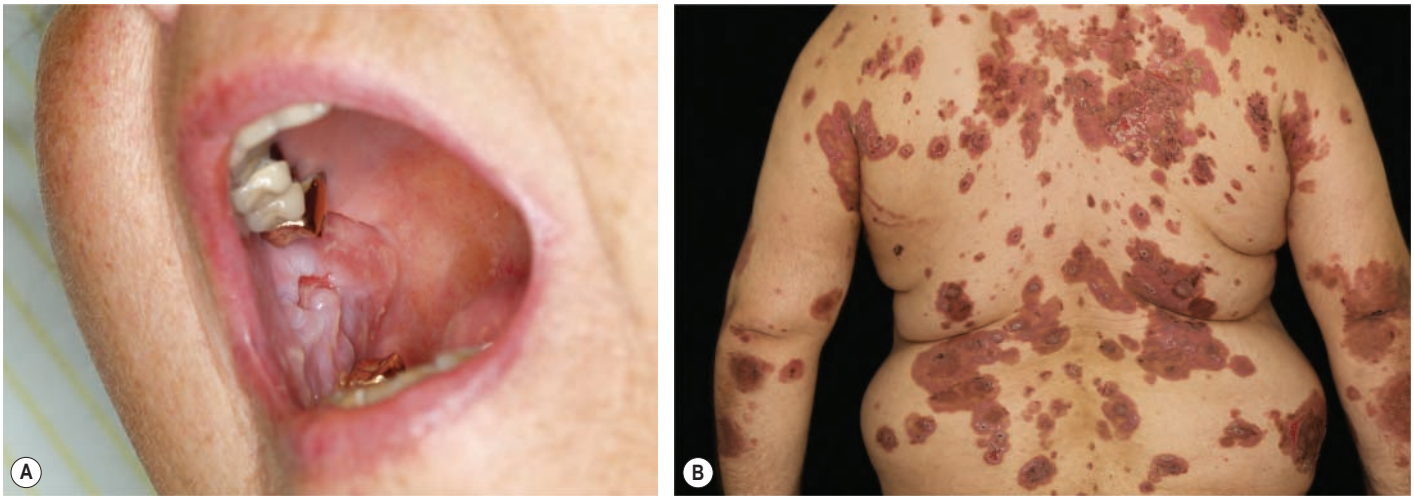


FIG. 63.2 Pemphigus Vulgaris. Erosions and a deep ulcer on the oral mucosa (A). Erosions and partly crusted erythematous plaques on the back and arms (B).

Nearly all patients will develop oral lesions. Pain is highly variable between patients. It ranges from minor to severe discomfort and may prevent food intake and thus be accompanied by rapid weight loss. All surface-close mucosal tissues may become involved. This includes the nose (e.g., hemorrhagic crusts), pharynx, larynx (e.g., hoarseness), urethra, glans penis, vulva, cervix, and the perianal region. In the mucocutaneous variant of pemphigus vulgaris, which occurs in about half of the patients, skin lesions arise in parallel with mucosal lesions. Skin lesions begin as flaccid blisters, erosions, and crusts (see Fig. 63.2, B). Subsequently, large areas may denude. Both mucosal and skin lesions heal without scarring. However, in patients with darker skin, hyperpigmentation of affected skin areas may be visible for many months.[1,4,24]

Pemphigus Foliaceus

In pemphigus foliaceus, mucosal surfaces are spared and lesions are restricted to the skin. They may present as erythema, "puff pastry–like" scales, and crusts, preferentially arising on the face, scalp, and seborrheic areas of the upper trunk (Fig. 63.3). Intact blisters are rare, due to the superficial splitting directly below the stratum corneum, which leads to rapid destruction of the blister roof upon mechanical stress.[1,4,24]

Paraneoplastic Pemphigus

The clinical picture is polymorphous with flaccid blisters, pustules, tense blisters as in BP, erythema multiforme–like erythema, and lichenoid lesions.[3,4,25] A severe stomatitis is nearly always present. Genital, nasal, and ocular mucosal tissues can also be affected. In almost all patients, a neoplasm is present at the time of diagnosis. It can be found by an extensive tumor search, or arise within the following months. Lymphoproliferative disorders can be detected in 70% to 80% of European patients. Others include thymoma, malignant solid tumors, and Castleman tumor. The latter tumor is the most frequent associated neoplasm in Asia. Bronchiolitis obliterans, explained by the ectopic expression of DSG3 in the lung, was found in 6% of European patients and 20% of Japanese patients.[4]

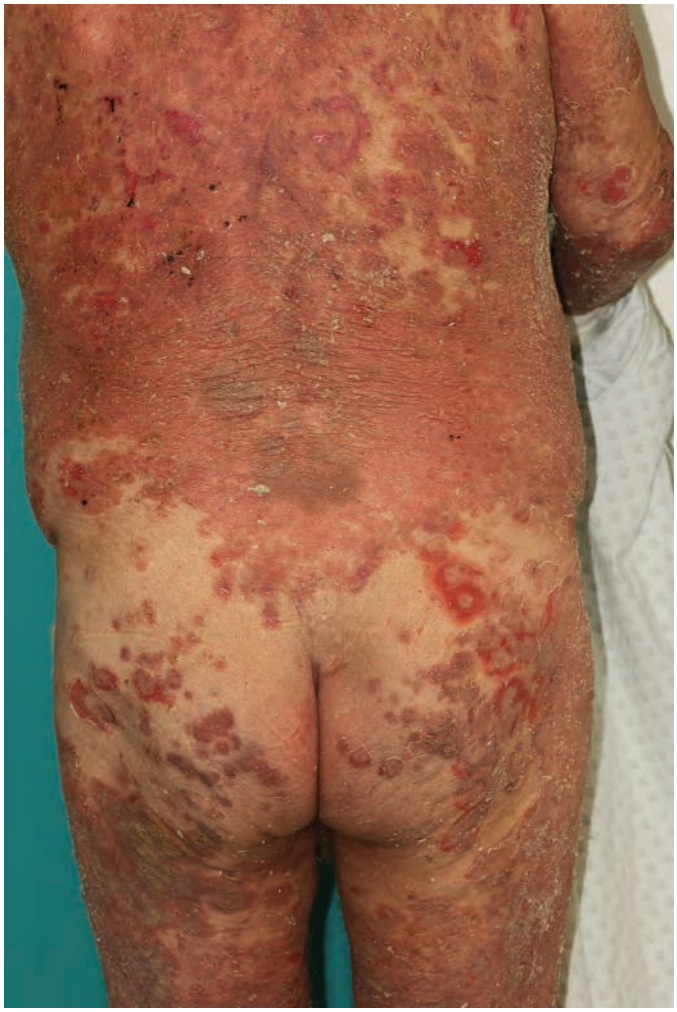


FIG. 63.3 Pemphigus Foliaceus. Widespread erythema, erosions, and "puff pastry–like" scaling on the back.

Pemphigoid Diseases

Bullous Pemphigoid

BP is a disease of the elderly, with a mean age at disease onset of 75 to 80 years. Nearly all patients with BP suffer from intense pruritus. In the classical bullous form, widespread tense blisters of variable sizes, erosions, and crusts appear on apparently normal or erythematous skin (Fig. 63.4, A). Urticarial erythematosus plaques may also develop (see Fig. 63.4, B). The flexural aspects of the limbs and abdomen are most commonly affected. Oral lesions can be found in 10% to 20% of cases. In the absence of severe superinfection, lesions heal without scarring. In a majority of patients, a non-bullous prodromal phase is seen with intense pruritus and unspecific erythematous lesions.[1,5,26] About 20% of patients present with non-bullous BP, including a large variety of clinical variants.[1,5,27] In some patients, the disease is limited to certain body parts, especially the pretibial area. Non-bullous or localized forms may remain as such or develop into classical BP.

Mucous Membrane Pemphigoid

MMP affects the oral cavity in 85% of patients (Fig. 63.5, A) followed by conjunctivae (65%), skin (25% to 30%), nasal mucosa (20% to 40%) (see Fig. 63.5, B), ano-genital area (20%), pharynx (20%), larynx (5% to 10%), and esophagus (5% to 15%). Except the oral cavity, lesions at all affected body sites tend to heal with scarring. Nasal lesions may present as hemorrhagic crusts and epistaxis, pharyngeal lesions are painful on swallowing, and the initial involvement of the larynx manifests as hoarseness. Esophageal disease may become symptomatic with dysphagia and heartburn. Ocular lesions usually start unilaterally with burning and foreign-body sensation. They may proceed to shortening of the inferior fornix, symblepharon, trichiasis, neovascularization, and, finally, blindness (see Fig. 63.5, C). In patients with only one affected mucosal site, terms related to location are applied (e.g., oral, ocular, or vulvar MMP). A thorough tumor search is required in patients with anti-laminin 332 MMP, since about 25% of patients develop a solid cancer.[1,5,26,26a,65]

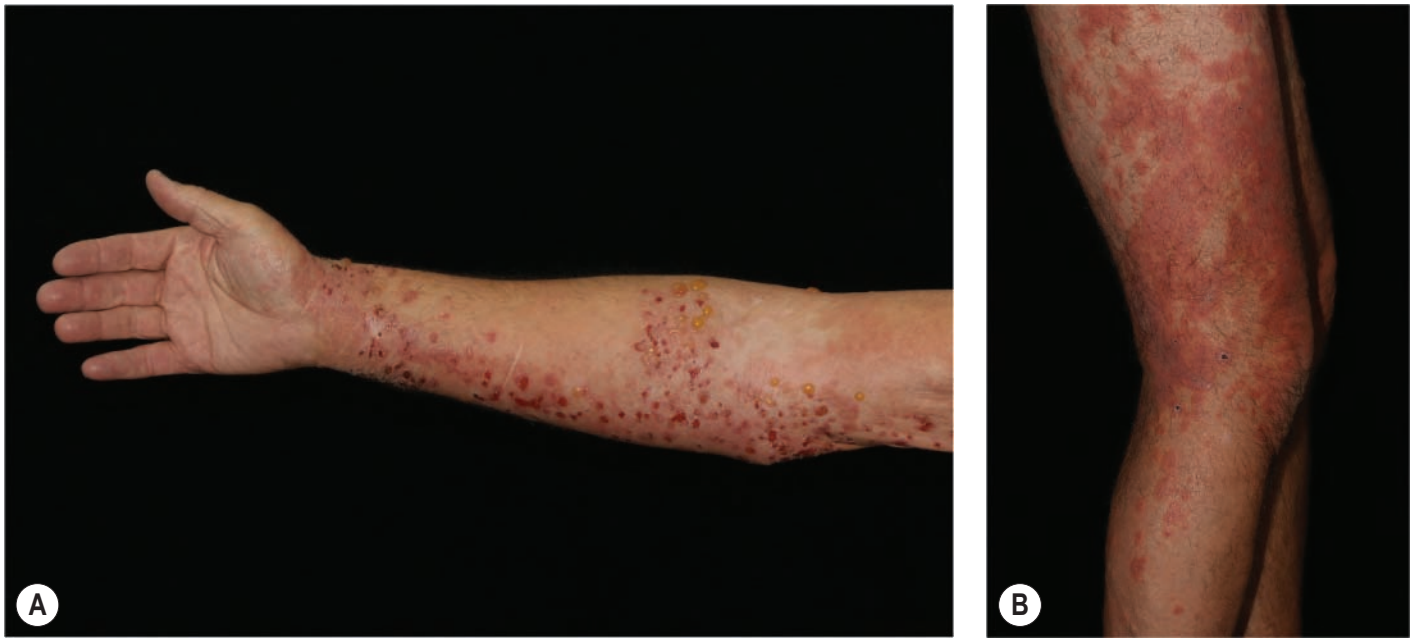


FIG. 63.4 Bullous Pemphigoid. Tense blisters and vesicles, erythematous plaques, and erythema on the right arm in the classical variant of bullous pemphigoid (A). Erythematous plaques, erythema, and a single vesicle on the right leg (B).

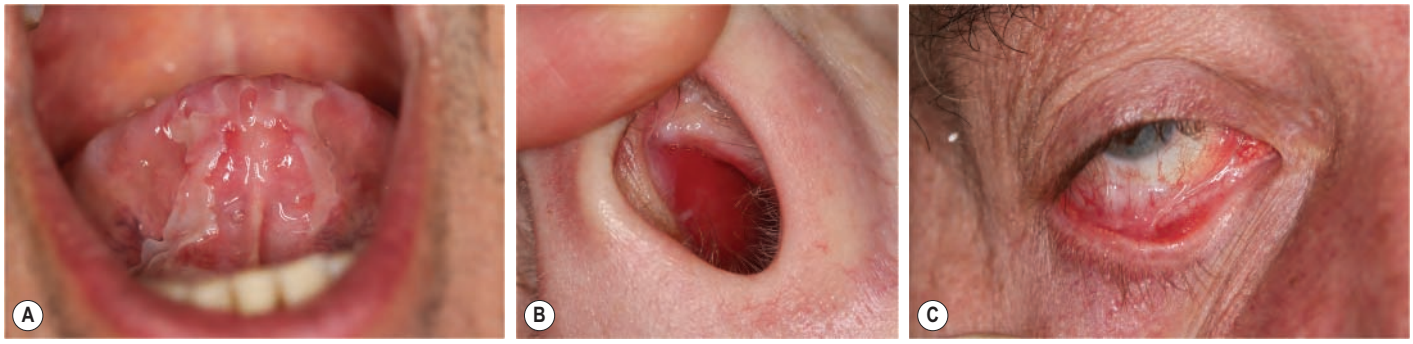


FIG. 63.5 Mucous Membrane Pemphigoid. Fibrin-covered erosions on the tongue (A), erosion on the nasal mucosa (B), conjunctival injections, shortening of the lower fornix, and symblepharon of the right eye (C).

Linear Immunoglobulin A Disease

Tense blisters, vesicles, urticated plaques, erosions, and erythema are the clinical hallmarks of linear IgA disease, or dermatosis. The clinical picture is virtually indistinguishable from BP and anti-p200 pemphigoid. However, blisters and vesicles tend to arise in an annular pattern in linear IgA disease, with blistering along the edge of lesions forming the so-called "string of pearls" or "cluster of jewels" sign. This sign is not exclusive to linear IgA disease. While the individual lesion does not differ between adults and children, in children lesions arise more abruptly and tend to involve the perioral area and perineum in addition to the other predilection sites, trunk and limbs. Mucosal lesions, mostly oral, nasal, and genital, occur in about 70% of patients. In patients with ocular scarring, the diagnosis of MMP is appropriate.[1,5,28]

Pemphigoid Gestationis

Frank blistering is rare. Erythematous papules, urticarial erythema, and papulovesicles predominantly develop on the periumbilical area in the second and third trimester as well as post partum. As in BP, intense pruritus is nearly always present. Disease onset in the first or second trimester. Blisters are associated with adverse pregnancy outcomes.[29]

Anti-p200 Pemphigoid

Most patients present with tense blisters on erythematosus or normal skin resembling BP (Fig. 63.6). Hands and feet appear to be favored, and oral and/or genital lesions may be present. Lesions usually heal without scarring.[22]

Epidermolysis Bullosa Acquisita

Two main clinical forms have been described: the classical mechanobullous variant in about a third of patients, and the inflammatory form (Fig. 63.7). In the classical mechanobullous variant, skin fragility, erosions, blisters, crusts, and scars on trauma-prone areas arise, and scarring alopecia, nail loss, and milia formation may occur. The inflammatory variant resembles other pemphigoid diseases such as BP, MMP, and linear IgA disease, and may arise together in the same patient.[1,5,26,30]

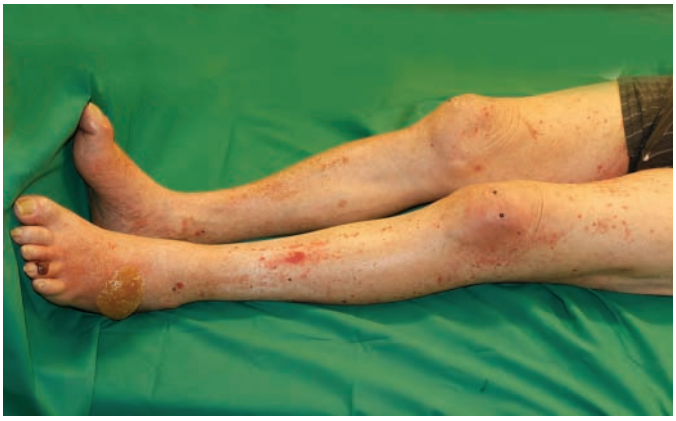


FIG. 63.6 Anti-p200 Pemphigoid. Tense blisters on the left foot and erythema, as well as erythematous plaques and papules on both legs.

Dermatitis Herpetiformis

Pruritic vesicles, erosions, and erythematous papules arranged in a grouped, so-called herpetiform pattern, are the clinical hallmarks of dermatitis herpetiformis (Fig. 63.8). Lesions tend to be symmetrical, predominantly affecting the extensor sites, in particular the nates, and heal without scarring. Some patients develop punctate purpura on the palms and soles.[28]

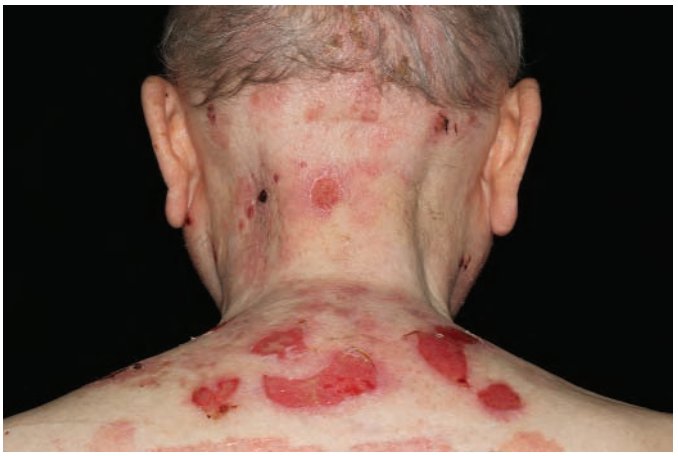


FIG. 63.7 Epidermolysis Bullosa Acquisita. Erosions, crusts, and erythema on the head, neck, and upper back in a patient with the inflammatory variant.

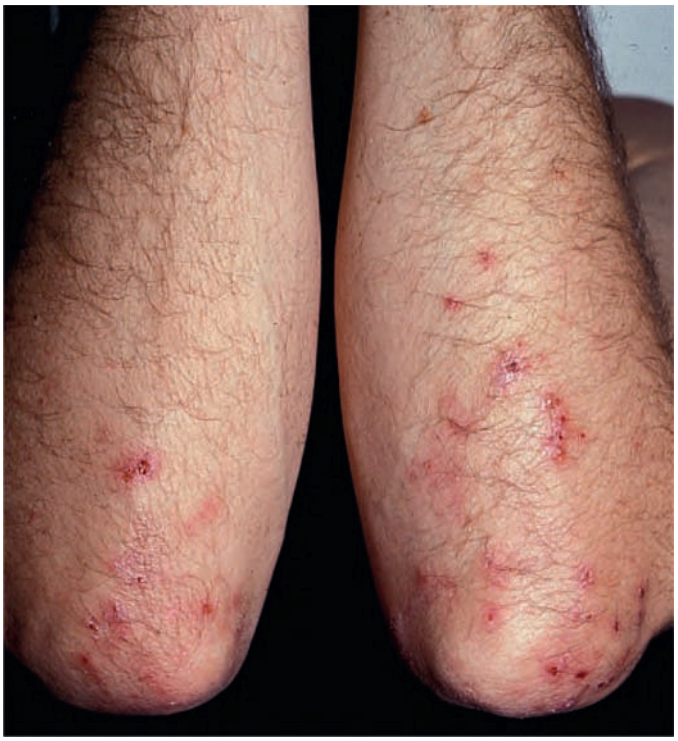


FIG. 63.8 Dermatitis Herpetiformis. Erythematous, partly excoriated or crusted papules and rare vesicles on the elbows.

KEY CONCEPTS

Clinical Presentation

- In pemphigus vulgaris, erosions arise on surface-close mucosal tissue. The oral cavity is nearly always affected and about half of the patients develop additional skin lesions.
- Lesions in pemphigus foliaceus are restricted to the skin.
- Pemphigoid diseases present with tense blisters and erosions and cannot be differentiated from each other based on the clinical picture alone.
- Bullous pemphigoid is a disease of the elderly with a mean age at disease onset of 75–80 years.
- In mucous membrane pemphigoid, mucosal lesions predominate.
- Patients with dermatitis herpetiformis rarely show frank blistering but extensively pruritic erythematosus papules and erosions.

DIAGNOSIS

Some clinical signs are typical of specific immunobullous diseases. For example, old age, tense blisters, and severe pruritus characterize BP, and flaccid blisters and a positive Nikolsky sign (induction of erosion by mechanical friction) characterize pemphigus vulgaris. Nevertheless, an AIBD cannot be diagnosed by clinical signs alone and requires the detection of tissue-bound and circulating autoantibodies.[4,5,31]

The diagnostic gold standard is the detection of tissue-bound autoantibodies by direct immunofluorescence (IF) microscopy of a perilesional biopsy. In pemphigus diseases, intercellular deposits of IgG and/or C3 (IgA in IgA pemphigus) are seen between neighboring keratinocytes/epithelial cells (Fig. 63.9). In pemphigoid disorders, direct IF reveals linear staining of IgG, IgA, and/or C3 at the DEJ. Within the linear staining, two patterns can be distinguished: a n-serrated pattern with arches closed at the top and a u-serrated pattern with arches closed at the bottom (Fig. 63.10). While a u-serrated pattern is restricted to an autoimmune reaction to type VII collagen (e.g., epidermolysis bullosa acquisita and bullous systemic lupus erythematosus), all other pemphigoid diseases show a n-serration pattern.[32] In dermatitis herpetiformis, granular deposits of IgA are seen in the dermal papillae (Fig. 63.11) and sometimes along the DEJ.

With the discovery of the molecular identity of most AIBD target antigens (see Tables 63.1 and 63.2) and advances in assays for detecting serum autoantibodies, most AIBD patients can be diagnosed by serology alone.

The most sensitive screening substrate for pemphigus antibodies is monkey esophagus (Fig. 63.12) and for pemphigoid diseases, human salt-split skin (Fig. 63.13). In the latter substrate, an artificial split within the lamina lucida of the DEJ is generated by incubation of normal human skin with 1 M NaCl solution. In this substrate, and depending on the target antigen, autoantibodies label the epidermal or dermal side of the artificial split (see Fig. 63.13). Standardized enzyme-linked immunosorbent assay (ELISA) systems detect antibodies against major AIBD target antigens. DSG1, DSG3, BP180, BP230, type VII collagen, and envoplakin are widely available and allow both the diagnosis and monitoring of serum autoantibody levels during the course of the disease (see Table 63.2).

Alternatively, a CE-certified multivariant indirect IF test that is based on Biochip technology can be applied. Here, various tissue substrates are placed as 1×1 mm miniature substrates, so-called biochips, in one incubation field on a routine laboratory slide (Fig. 63.14). This method allows the

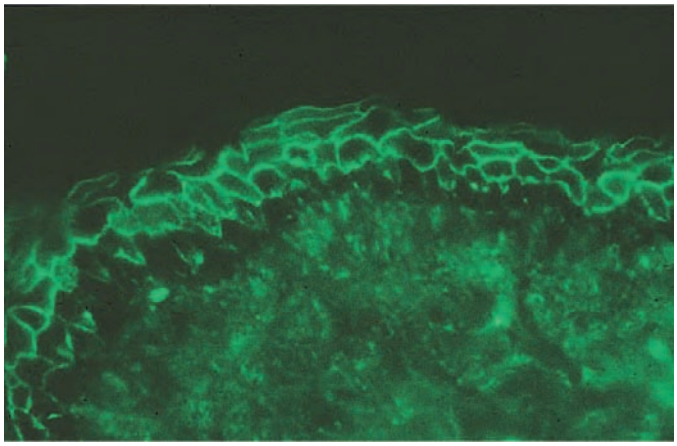


FIG. 63.9 Direct Immunofluorescence Microscopy in Pemphigus. A perilesional biopsy from a patient with pemphigus foliaceus shows intercellular deposits of immunoglobulin G in the epidermis in the typical rete-like pattern. A similar pattern is seen in pemphigus vulgaris.

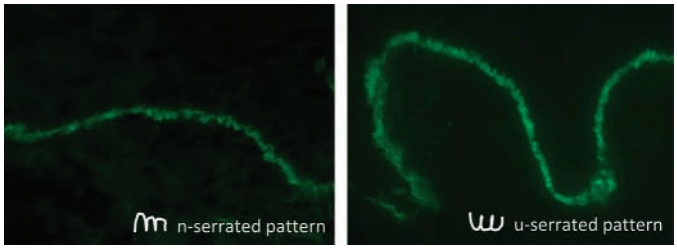


FIG. 63.10 Direct Immunofluorescence Microscopy in Pemphigoid Diseases. In perilesional biopsies, linear deposits of immunoglobulin G at the dermal–epidermal junction. In autoimmunity against type VII collagen (i.e., epidermolysis bullosa acquisita and systemic lupus erythematosus), a u-serrated pattern is seen with arches open at the top *(right)*. In all other pemphigoid diseases, an n-serrated pattern is observed with arches closed at the top *(left)*.

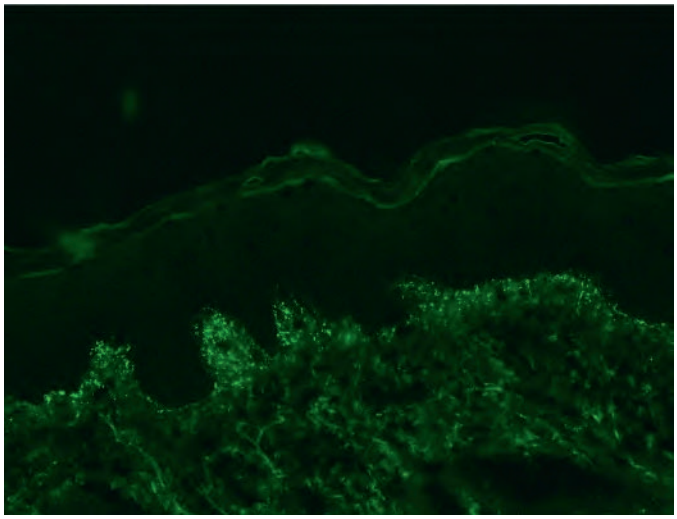


FIG. 63.11 Direct Immunofluorescence Microscopy in Dermatitis Herpetiformis. A perilesional biopsy reveals granular deposits of immunoglobulin A in the dermal papillae and along the dermal–epidermal junction.

simultaneous incubation of multiple miniature substrates and a standardized and rapid serological diagnosis of the major AIBD.[31] Substrates include monkey esophagus, salt-split skin, and rat bladder, recombinant BP180 NC16A, and human cells expressing recombinant forms of AIBD target antigens such as DSG1, DSG3, BP230, type VII collagen, and laminin 332. Detection of serum IgG antibodies against the BP180 ectodomain (in MMP), the p200 protein, and laminin γ1, as well as against IgA autoantibodies in pemphigoid and pemphigus diseases, are still restricted to specialized laboratories.

Traditionally, lesional histopathology has been the third diagnostic column for AIBDs. Histopathology allowed differentiation between subepidermal AIBD (pemphigoid and dermatitis herpetiformis) and pemphigus disorders.[2] Since histopathology cannot separate the different pemphigoid diseases, its diagnostic impact has been largely replaced by direct IF and serology. However, a lesional biopsy is still recommended in any patients with suspected AIBD in order to allow consideration of differential (alternative) diagnoses when direct IF and serology are negative or inconclusive. Differential diagnoses of AIBDs are detailed in the respective textbooks.[1,24,26,28]

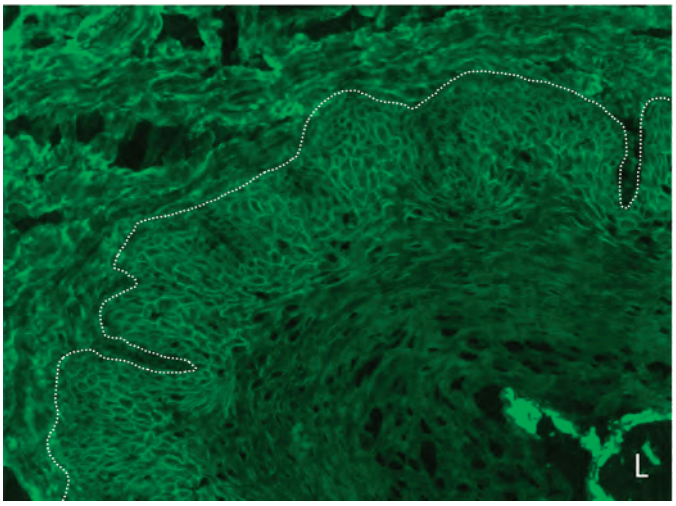


FIG. 63.12 Intercelluar staining of immunoglobulin G in the epithelium of monkey esophagus by indirect immunofluorescence microscopy of pemphigus serum. The basement membrane zone is indicated by a *dotted line*. *L*, Lumen of the esophagus.

KEY CONCEPTS

Diagnosis

- Direct IF microscopy of a perilesional biopsy remains the diagnostic gold standard of AIBD.
- Standardized, highly sensitive and specific ELISA systems and indirect IF assays based on recombinant forms of the major target antigens are widely available for most AIBD.
- At present, about 90% of AIBD can be diagnosed based on the clinical picture and serology.

PATHOPHYSIOLOGY

Pemphigus

In genetically susceptible individuals, the autoimmune reaction in pemphigus[3,4,6,33,34] is driven by autoreactive T and B lymphocytes. Autoreactive T cells are educated by antigen-presenting cells that present specific Dsg peptides via their HLA class II molecules encoded by the HLA class II risk haplotypes described above (Chapters 6 and 9). These Dsg-specific CD4 and IL-10-producing autoreactive T lymphocytes then drive the generation of Dsg-specific antibodies by B cells.[3,33]

The impact of autoantibodies against DSG1 and DSG3 in the pathogenesis of pemphigus is derived from several clinical and experimental observations (Fig. 63.15): (i) the transplacental transfer of maternal autoantibodies from mothers suffering from pemphigus can cause transient blistering in newborns, (ii) in nearly all patients with pemphigus foliaceus, and in the majority of pemphigus vulgaris patients, anti-Dsg IgG serum levels closely correlate with the extent of lesions, (iii) when cultured keratinocytes are treated with IgG from pemphigus patients or anti-Dsg3 IgG, degradation of the desmosomes is observed, (iv) incubation of sheets of cultured keratinocyte with pemphigus IgG leads to fragmentation of sheets mimicking acantholysis, (v) mice injected with pemphigus serum or anti-Dsg IgG develop intraepidermal splitting and macroscopic blistering,[35] (vi) blistering in mice is prevented after injection with pemphigus IgG depleted from anti-Dsg1/Dsg3 IgG, (vii) co-application of pemphigus IgG with a peptide that mediates Dsg3 crosslinking does not result in blistering of mice, and (viii) anti-Dsg3 IgG is induced leading to microscopic and macroscopic blistering in immunodeficient mice after adoptive transfer of lymphocytes from Dsg3-deficient mice that had been immunized with recombinant murine Dsg.[3,4,36]

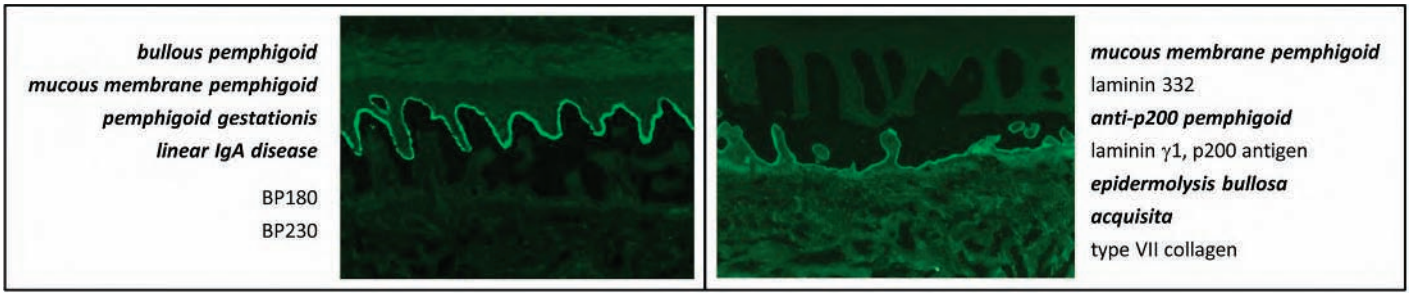


FIG. 63.13 Indirect Immunofluorescence Microscopy on Normal Human Salt-Split Skin. Serum autoantibodies against BP180 and BP230 in bullous pemphigoid, mucous membrane pemphigoid, pemphigoid gestationis, and linear immunoglobulin A (*IgA*) disease label the roof of the artificial split, while autoantibodies against laminin 332, the p200 antigen, and type VII collagen in mucous membrane pemphigoid, anti-p200 pemphigoid, and epidermolysis bullosa acquisita bind along the blister floor.

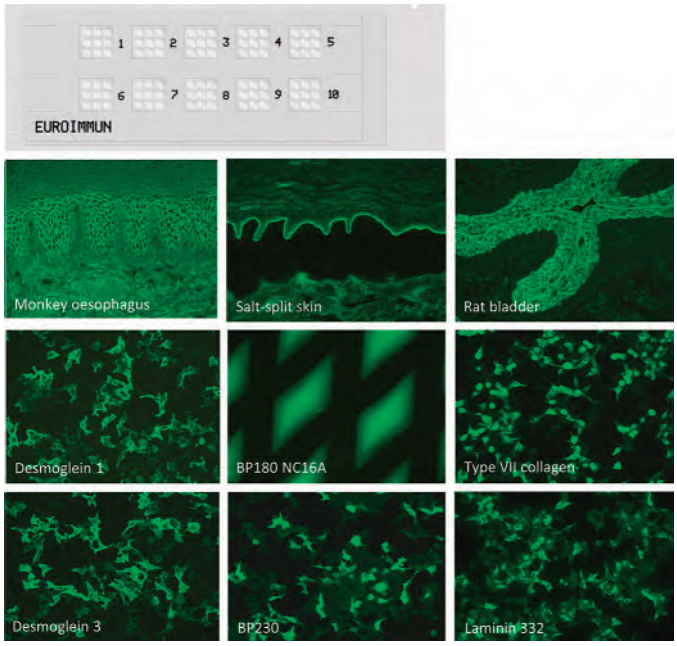


FIG. 63.14 Multivariant Indirect Immunofluorescence Microscopy Based on the Biochip Technology. Miniature substrates placed in an incubation field of a standard-sized laboratory slide allow the simultaneous detection of serum auto-antibodies with different specificities. Here, positive reactivities against monkey esophagus (intercellular epithelial staining in pemphigus), normal salt-split skin (epidermal staining in bullous pemphigoid), rat bladder (urothelium staining in paraneoplastic pemphigus), the human cell line HEK293 expressing recombinant desmoglein 1, desmoglein 3 (in pemphigus), type VII collagen (in epidermolysis bullosa acquisita), BP230 (in bullous pemphigoid), and laminin 332 (in mucous membrane pemphigoid) on the cells surface, respectively, as well as recombinant BP180 NC16A (in bullous pemphigoid) are compiled. (Courtesy of Euroimmun, Lübeck, Germany.)

However, not all anti-Dsg antibodies are pathogenic. This may explain the observation that serum anti-DSG3 levels in some pemphigus vulgaris patients do not correlate with disease activity and that lesions may persist when patients are in remission.[8] Pathogenic epitopes on DSG1 and DSG3 are Ca^{2+}- and conformationally dependent. They are mainly clustered in the EC1 and EC2 domain but can be found on the entire Dsg ectodomains. Interestingly, while the majority of pemphigus vulgaris patients with an initial mucosal variant and exclusive anti-DSG3 autoimmunity will eventually develop additional antibodies against DSG1 and skin lesions, epitope spreading within the Dsg molecules is rarely seen during the course of the disease.[3,4]

In pemphigoid diseases, a cascade of events including complement activation and Fc-receptor-mediated effects are pivotal for autoantibody-induced subepidermal blister formation (see below).[5] Monovalent fragments of anti-Dsg antibodies that lack the Fc portion can also cause acantholysis in vitro and in vivo.[4] The exact sequence of events in anti-Dsg antibody-mediated acantholysis has not yet been fully clarified. Three major mechanisms leading to acantholysis upon binding of anti-Dsg IgG have been identified: (i) direct interference with Dsg transinteraction, a phenomenon termed steric hindrance, (ii) remodeling of Dsg expression on the cell surface leading to internalization and depletion of Dsg from the cell membrane, and (iii) signaling events within the targeted keratinocytes.[3,4] The latter include activation of the p38 mitogen-activated protein kinase (MAPK), epidermal growth factor receptor, RHO GTPases, MYC, and caspases that interfere with the cytoskeletal architecture.[3,33,34,37] While these three mechanisms are sufficient for anti-Dsg IgG-mediated acantholysis, additional events mediated by soluble Fas ligand and non-desmoglein antibodies may also contribute to the pemphigus phenotype at later stages of the disease.[4]

The potential pathogenic role of non-desmoglein antibodies is exemplified by the finding of severe erosions and suprabasal splitting in mice deficient in epidermal Dsc1 as well as the pathogenic effects of anti-Dsc3 IgG in vitro and in a recent mouse model.[38] The contribution of antibodies against non-desmosomal antigens to the pathophysiology of pemphigus is suggested by the correlation of serum levels of anti-muscarinic acetylcholine receptor IgG with disease severity, and the co-pathogenic effect of antimitochondrial antibodies in addition to anti-Dsg IgG in vitro.[4,39]

KEY CONCEPTS

Pathophysiology of Pemphigus

- Anti-Dsg antibodies can induce acantholysis in vitro and in vivo independent of Fc-fragment-mediated mechanisms.
- Steric hindrance, remodeling of Dsg expression, and signal-transducing events are the three main pathogenic mechanisms leading to acantholysis after the binding of pemphigus autoantibodies. The in vivo relevance of non-desmoglein autoantibodies in the pathophysiology of pemphigus awaits further clarification.

Pemphigoid Diseases

The pathogenesis of pemphigoid diseases can be divided into three distinct steps (Fig. 63.16): (1) Loss of tolerance to pemphigoid disease autoantigens and production of autoantibodies, (2) circulation of autoantibodies, and (3) autoantibody-mediated tissue pathology. Insights into these steps has been obtained from genetic studies and the use of pemphigoid disease model systems.[40]

Loss of Tolerance to Pemphigoid Disease Autoantigens and Production of Autoantibodies

Loss of tolerance and production of pemphigoid disease–specific autoantibodies is associated with genes located within and outside the Major Histocompatibility Complex (MHC)/HLA-locus. In most pemphigoid disease, an association with certain HLA loci has been described.[41] Immunization-induced pemphigoid in mice also shows a strong association with the MHC region as well as outside the MHC.[41] The metagenome also has a significant impact on the generation of the autoantibody response in pemphigoid disease. In BP patients, the skin microbiome (Chapter 23) is distinct from that of healthy controls.[42] This difference is of functional relevance, at least in mice, as a low diversity of the skin microbiome prior to induction of experimental pemphigoid disease by immunization is associated with a high risk of developing skin lesions, while a high diversity of the skin microbiome confers protection from clinical disease manifestations.[43]

Regarding the cellular requirements that lead to the loss of tolerance to pemphigoid disease antigens and autoantibody production (see Fig. 63.16, A), data from experimental murine pemphigoid disease identified B cells, dendritic cells, and macrophages as key antigen-presenting cells that induce the CD4 T cell-dependent autoantibody production of B cells. Interestingly,

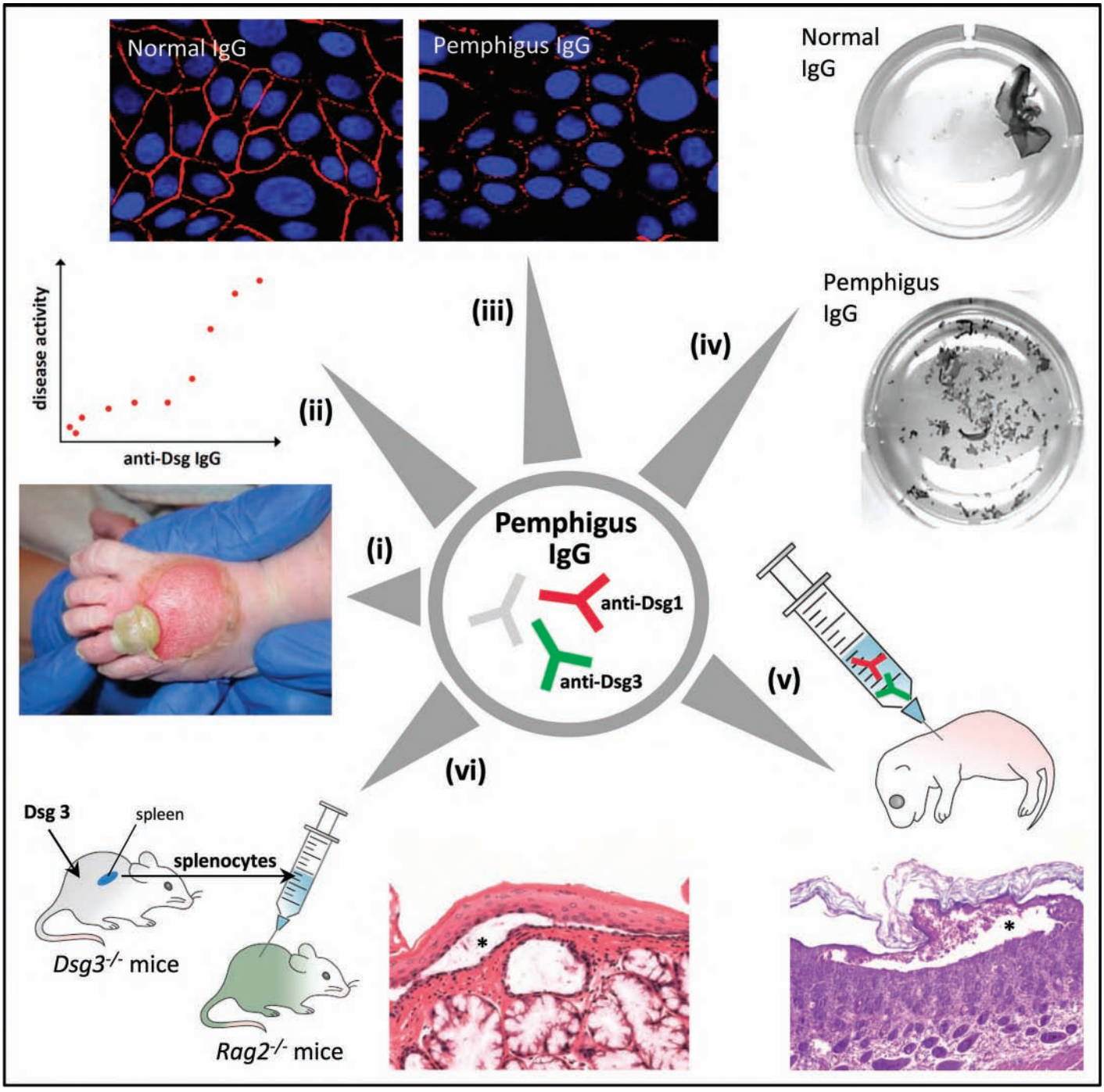


FIG. 63.15 Pathogenic Potential of Pemphigus Autoantibodies Based on Different Clinical and Experimental Observations. (i) The transplacental transfer of maternal autoantibodies from mothers with pemphigus may cause transient blistering in newborns, (ii) close correlation between the extent of lesions and anti-desmoglein (*Dsg*) immunoglobulin G (*IgG*) serum levels in most pemphigus patients, (iii) desmosome degradation in cultured keratinocytes after incubation with pemphigus IgG, (iv) acantholysis of cultured keratinocyte sheets upon treatment with pemphigus IgG, (v) intraepidermal splitting (*) and macroscopic blistering (not shown) in mice injected with pemphigus serum or anti-Dsg IgG,[35] and (vi) induction of anti-Dsg3 antibody production and microscopic (*) and macroscopic (not shown) blistering in immunodeficient *Rag2*$^{-/-}$ mice after adoptive transfer of lymphocytes from Dsg3-deficient (*Dsg3*$^{-/-}$) mice after immunization with recombinant Dsg3.[36] (Clinical picture in (i) courtesy Susann Ott, Department of Pediatrics, Klinikum Bayreuth, Bayreuth, Germany. Histology image in (vi) courtesy Hayato Takahashi and Masayuki Amagai, Department of Dermatology, Keio University School of Medicine, Tokyo, Japan. Modified from Schmidt E, Kasperkiewicz M, Joly P. Pemphigus. *Lancet*. 2019;394[10201]:882–894. Epub 2019/09/10.)

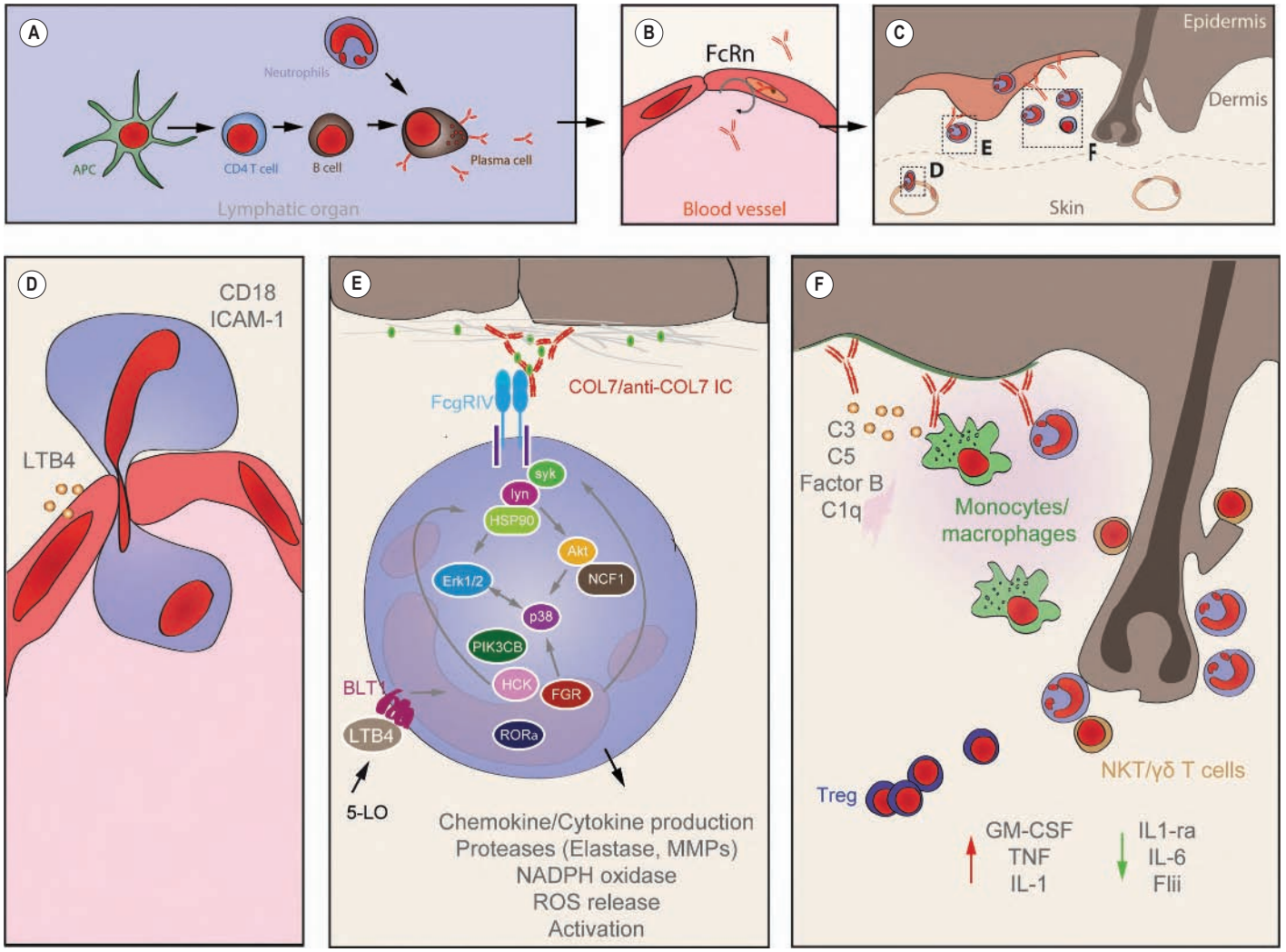


FIG. 63.16 Pathogenesis of Pemphigoid Diseases. Pemphigoid disease pathogenesis can be divided into three distinct pathways: loss of tolerance and autoantibody production, circulation of autoantibodies, and autoantibody-induced tissue pathology. (A) Antigen-presenting cells *(APC)* and CD4 T cells promote expansion of autoreactive B cells that develop into autoantibody-producing plasma cells. This process is enhanced by the presence of B-helper neutrophils. In mice, this process almost exclusively occurs in peripheral lymph nodes, where the majority of pemphigoid disease-specific autoreactive B/plasma cells are found. (B) Once in the circulation, the half-life of immunoglobulin G (IgG) autoantibodies is maintained by the neonatal Fc receptor *(FcRn)*, which protects all IgG from proteolysis after uptake by endothelial cells. (C). In the skin, neutrophils bind to the immune complexes located at the dermal–epidermal junction, become activated and release reactive oxygen species (*ROS*) and specific proteases, which induce subepidermal blistering and cutaneous inflammation. (D) Extravasation of neutrophils (and other effector cells) from the bloodstream into the skin is mediated by an interaction of CD18 and ICAM-1. (E) Binding of neutrophils to the immune complexes at the dermal–epidermal junction is mediated by activating FcγR. This triggers a complex intracellular signaling cascade, which ultimately leads to mediator release by the neutrophils. Among those are the blister- and inflammation-driving ROS and specific proteases, as well as cytokine release. The latter enhances the extravasation and activation of neutrophils. In addition to engagement of activating FcγR, soluble mediators, such as leukotriene B4 *(LTB4)*, further promote neutrophil activation. (f) In addition to neutrophils, monocytes/macrophages, NK- and γδ T cells promote inflammation in pemphigoid diseases, while regulatory T cells *(Treg)* dampen neutrophil extravasation into the skin. *C1q*, Complement component 1q; *Erk*, extracellular-signal regulated kinases; *FGR*, Tyrosine-protein kinase; Fgr *Fliii*, Flightless I; *HCK*, Tyrosine-protein kinase HCK; *HSP90*, heat shock protein 90; *NCF1*, neutrophil cytosolic factor 1; *PIK3CB*, Phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit beta; *RORa*, RAR Related Orphan Receptor A.

neutrophils act as B helper neutrophils in these models, as their depletion leads to reduced autoantibody concentrations. Very little has been published on the molecular requirements of this phase of pemphigoid disease pathogenesis. So far, only Granulocyte–macrophage colony-stimulating factor (GM-CSF) is known to promote autoantibody production in experimental pemphigoid, presumably through the recruitment of B helper neutrophils into peripheral lymph nodes, where antigen-specific B/plasma cells are found.[41]

Circulation of Autoantibodies

Once generated, autoantibodies are released into the bloodstream (see Fig. 63.16, B). With few exceptions (e.g., linear IgA disease),

the majority of autoantibodies in pemphigoid diseases are IgG. The relative long half-life of IgG antibodies, as opposed to the other Ig isotypes, is due to the constant rescue of IgG from proteolysis by the neonatal Fc receptor (FcRn). Hence, blockade of FcRn has emerged as a potential therapeutic target for the treatment of pemphigoid disease.[44] When observed by multiphoton microscopy, autoantibodies rapidly bind to their target skin antigen.

Autoantibody-Mediated Tissue Pathology

After binding to their target skin antigens, pemphigoid disease autoantibodies form immune complexes at the DEJ (see Fig. 63.16, C–F). In most cases, this leads to inflammation and subsequent inflammation-dependent subepidermal blistering. Recent evidence also implicates inflammation-independent pathways that induce blisters.[45] In inflammation-dependent blistering, the formation of immune complexes promotes a proinflammatory milieu. This process is dependent on both the Fab and the Fc components of the autoantibodies. Binding of the autoantibodies to their target antigen induces the release of proinflammatory cytokines from the keratinocytes. This is best described for BP, where incubation of keratinocytes with anti-BP180 NC16A IgG induces time- and dose-dependent release of IL-8. In mice, several cytokines (e.g., GM-CSF, IL-1, Tumor necrosis factor [TNF], LTB4) and anaphylatoxins (e.g., C5a) can promote the influx of immune cells into the skin. In contrast, other cytokines with increased expression in experimental pemphigoid disease demonstrate anti-inflammatory activities. For example, IL-6 induces expression of IL-1 receptor antagonist, which counteracts the disease-promoting effects of IL-1. Ultimately, this leads to a CD18/Intercellular adhesion molecule-1 (ICAM-1)-dependent extravasation of effector leukocytes into the skin. Among these, neutrophils are the major disease-driving cell type.

In the skin, neutrophils bind to the immune complexes located at the DEJ. In mice, this is mediated by the activating FcγR III and IV. In humans, it is facilitated by FcγR IIA and IIIB. FcγR-dependent binding of neutrophils to the immune complexes induces an intracellular signaling cascade involving SYK, p38 MAPK, ERK1/2, AKT, PI3Kβ/δ, Hsp90, RORα, PDE4, Src kinases, and CARD9. Ultimately, this results in the release of specific proteolytic enzymes and reactive oxygen species (ROS), which induce cutaneous inflammation and subepidermal blistering. In addition to neutrophils, monocytes/macrophages, NK-, and γδ T cells promote blistering in pemphigoid diseases, while regulatory T cells dampen skin inflammation in pemphigoid disease.[46]

The role of mast cells is controversial: while mast cells are reportedly essential for the induction of experimental pemphigoid in neonatal mice, their depletion had no effect on clinical disease manifestation in antibody transfer–induced epidermolysis bullosa acquisita in adult mice. Based on critical evaluation of the mast cell–deficient mouse strains in these different studies, these observations suggest that mast cells are activated but not essential for pathogenesis in/of pemphigoid disease.[41]

The mechanisms underlying the resolution of inflammation in pemphigoid disease[47] are incompletely understood. In addition to C5aR2, IL-6, IL10, proresolving lipid mediators, and regulatory T cells, flightless I (Flii) has been implicated in the resolution in pemphigoid.[47,48]

Inflammation-independent pathways to blister formation have also been identified. For example, the binding of autoantibodies targeting the BP180 induces internalization of BP180, which leads to a decrease in the strength of adhesion to the underlying matrix.[49] Autoantibody binding also impairs pro-adhesive protein–protein interactions in the DEJ, which further weakens the adhesion of keratinocytes to the underlying matrix.

Experimental Models

Most of the above insights have been obtained by using pemphigoid disease animal models (Fig. 63.17, Table 63.3). So far,

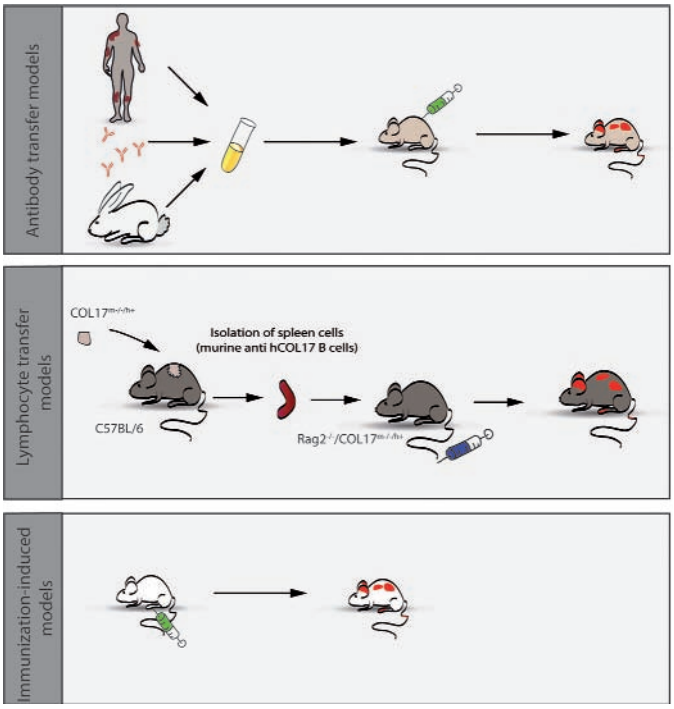


FIG. 63.17 Mouse Models of Pemphigoid. (A) In (auto)antibody transfer-induced pemphigoid, (auto)antibodies targeting a specific autoantigen in pemphigoid—that is, type XVII collagen (BP180)—are obtained from patients, generated recombinantly or obtained from rabbits immunized with the murine autoantigen. These (auto)antibodies are then injected into mice, which develop disease within a relatively short timeframe (days). When using these antibody transfer–induced models, reactivity of the (auto) antibodies with the skin of the recipient mouse has to be considered. These models reflect key pathogenic events of the autoantibody-induced tissue damage in pemphigoid, while not duplicating events leading to the generation of autoantibodies. (B) In the lymphocyte transfer-induced pemphigoid mouse models, wild-type C57Bl6 mice are immunized against human (h) type XVII collagen *(COL17)* by transplantation of skin from a murine (m) COL17-deficient / hCOL17 transgenic mouse. This leads to an adaptive immune response against human type 17 collagen (hCOL17). When lymphocytes from these mice are injected into immunodeficient *Rag2*$^{-/-}$, m*COL17*-deficent/hCOL17 transgenic mice, experimental bullous pemphigoid is induced. This model is ideal to investigate the interplay between T and B cells, as well as antibody transfer–induced tissue damage in pemphigoid. (C) In the immunization-induced pemphigoid disease models, pemphigoid is induced by immunization of susceptible mouse strains with the murine autoantigen. This leads to the development of an adaptive immune response against the used autoantigen, with subsequent development of clinical disease in the majority of the mice. This model reflects all aspects of pemphigoid disease pathogenesis, and allows therapeutic interventions for drug testing.

TABLE 63.3 Mouse Models of Pemphigoid Diseases and Dermatitis Herpetiformis

PD	Principle of Model	Variations of Model	Reference(s)
BP	Antibody transfer	Rabbit anti-mouse Col17 IgG into neonatal mice	*J Clin Invest.* 92: 2480
		Rabbit anti-mouse Col17 IgG into adult mice	*Am J Pathol.* 184:2185
		IgG from BP patients into Col17-humanized mice	*Nat Med.* 13:378
		Monoclonal human anti-human Col17 IgG1 into COL17-humanized mice	*J Immunol.* 185:7746
		Anti-human NC16A IgE purified from BP patients into humanized NC16A neonatal mice	*J Invest Dermatol.* 138:2032
		IgG from BP patients into humanized NC16A neonatal mice	*J Autoimmun.* 31:331
	Immunization	4× immunization with mouse COL17, female SJL/J mice	*J Immunol.* 187:1176
	Lymphocyte transfer	**Donor:** Immunization of wild-type mice by transplantation of skin from $mCOL17^{ko}$/$hCOL17^{tg}$ mice. **Recipient**: $hCOL17^{tg}$	*J Immunol.* 184:2166
EBA	Antibody transfer	Rabbit anti-mouse Col7 IgG into adult mice	*J Clin Invest.* 115:870; *J Cell Mol Med.* 18:1727
		Rabbit anti-human Col7 IgG into adult mice	*J Invest Dermatol.* 124:958
		Human anti-human Col7 IgG into adult mice	*J Invest Dermatol.* 126:1323
		Human anti-human Col7 IgG into adult mice	*Am J Pathol.* 170:2009
		Rabbit anti-human Col7 IgG into COL7-humanized mice	*J Invest Dermatol.* 135:1565
		Affinity-purified rabbit anti-mouse COL7 into adult mice	*Sci Rep.* 10:4509
	Immunization	4× immunization with murine $COL7^{mCOL7C}$, SJL/J mice	*J Immunol.* 177:3461
		4× immunization with different epitopes within murine COL7, SJL/J mice	*J Cell Mol Med.* 18:1727
		1× immunization with murine $COL7^{mCOL7C}$, SJL/J mice	*J Invest Dermatol.* 131:167
		1× immunization with murine COL^{vWFA2}, SJL/J and B6.SJL-H2s mice	*J Immunol.* 191:2978
MMP	Antibody transfer	Rabbit anti-human laminin-332 IgG into neonatal mice	*J Clin Invest.* 98:1509
		Rabbit anti-human laminin-332 Fab into neonatal mice	*Clin Immunol.* 95:26
		MMP patient IgG injected into skin grafts on SCID mice	*J Invest Dermatol.* 114:178
		Rabbit anti-mouse LAM α3 IgG into adult mice	*J Invest Dermatol.* 137:1709
DH	Immunization	Immunization of HLA-DQ8 transgenic NOD mice with gluten, followed by repetitive gluten feeding	*J Clin Invest.* 114:1090

BP, Bullous pemphigoid; *Col17*, type XVII collagen; *DH*, dermatitis herpetiformis; *EBA*, epidermolysis bullosa acquisita; *IgG*, immunoglobulin G; *LAM α3*, α3 chain of laminin 332; *MMP*, mucous membrane pemphigoid; *PD*, pemphigoid diseases. References are indicated by Journal. Volume: first page.

there are three principal mechanisms by which experimental pemphigoid can be induced in mice: First, by transfer of (auto) antibodies targeting specific pemphigoid autoantigens, such as type VII collagen or BP180. Second, by transfer of pemphigoid disease–specific lymphocytes into immunodeficient mice that express the corresponding autoantigen. Third, by immunization of susceptible mouse strains with the respective pemphigoid disease autoantigen.

The in vivo pathogenicity of anti-BP180 autoantibodies was demonstrated in 1993, when Liu and colleagues immunized rabbits with the immunodominant stretch of murine BP180 and transferred the rabbit anti-mouse BP180 IgG into neonatal mice.[50] For epidermolysis bullosa acquisita, the injection of anti-type VII collagen IgG in adult mice is currently the most frequently used pemphigoid mouse model.[40] The injection of anti-laminin 332 IgG into adult mice results in the replication of the clinical characteristics of the human disease (e.g., oral, ocular, and skin lesions).[51] These models allow in-depth investigations solely focused on autoantibody-induced tissue damage in pemphigoid. There are many pitfalls that need to be considered (such as altered or absent cross-reactivity of the human autoantibodies with the mouse antigens). Recently, a detailed protocol has been published that describes the step-by-step induction of experimental epidermolysis bullosa acquisita.[40]

The lymphocyte transfer–induced model of pemphigoid is based on the principle that mice that do not express the (auto) antigen develop an adaptive immune response toward this particular antigen (e.g., BP180). This immune response can be enhanced by immunization of these "autoantigen-deficient" mice with the autoantigen. Lymphocytes from these mice are then transferred into immunodeficient mice that express the autoantigen. The "auto"-reactive lymphocytes are stimulated by the presence of autoantigen, and experimental pemphigoid disease develops. This principle has been established for BP:[52] Here, wild-type C57Bl/6 mice were immunized against human BP180 by the transplantation of skin from mouse *COL17*-deficient, human COL17 transgenic mice. Consequently, these mice develop an adaptive immune response against hBP180. Transfer of lymphocytes from wild-type mice transplanted with hCOL17 transgenic mice into *mCOL17*-deficient, hCOL17-transgenic immunodeficient RAG2-deficient mice induces experimental BP in the recipient mice.[52]

Experimental pemphigoid disease in mice can also be induced by immunization of susceptible mouse strains. In these models, susceptible mouse strains are immunized with pemphigoid disease antigens (e.g., BP180 or type VII collagen). A specific autoantibody response is observed in most strains. However, only very few inbred strains (e.g., SJL/J) develop clinical disease.[53] The immunization-induced epidermolysis bullosa acquisita model is very robust, with a disease penetrance of 60% to 80% within 10 weeks after immunization. Disease persists over weeks after a single immunization. Hence, this model is also well-suited to evaluate the potential therapeutic effects of innovative treatment strategies.

In addition to experimental pemphigoid in mice, in vitro model systems mirror several aspects of pemphigoid disease pathogenesis. Among disease-specific in vitro model systems are: (1) immune complex-induced neutrophil activation, (2) ex vivo dermal–epidermal separation in cryosections of skin incubated with pemphigoid autoantibodies and neutrophils, and (3) the indirect complement fixation assay.[54,55]

KEY CONCEPTS

Pathophysiology of Pemphigoid Diseases

- Loss of tolerance to pemphigoid disease antigens is a complex interplay of (meta)genetics with environmental factors (drugs), and requires cells of the adaptive and innate immune system.
- FcRn maintains the long half-life of pathogenic pemphigoid disease-specific autoantibodies of the IgG subclass. Pemphigoid disease autoantibodies promote tissue pathology primarily through inflammation-dependent pathways, although inflammation-independent pathways can also play a role.

Dermatitis Herpetiformis

Dermatitis herpetiformis is regarded as the cutaneous manifestation of celiac disease (Chapter 75). In basically all patients with dermatitis herpetiformis, some degree of gluten-sensitive enteropathy can be found in small bowel biopsies. In genetically susceptible individuals (i.e., individuals with the MHC class II DQ2 or DQ8 variants, which are shared by patients with dermatitis herpetiformis and celiac disease), exposure to gluten initiates an autoimmune response against gliadin. Gliadin is the alcohol-soluble fraction of gluten, a family of grain proteins present in wheat, rye, and barley, but not oats. Once gliadin is absorbed via the lamina propria, glutamine residues within gliadin are deaminated by tissue transglutaminase (TG2), a process that may lead to optimal antigen presentation to T cells by HLA-DQ2-positive antigen-presenting cells. Subsequently, IgA antibodies against gliadin, deaminated gliadin, gliadin crosslinked to TG2, and TG2 are generated. TG2-specific IgA antibodies are the serological hallmarks of celiac disease, while IgA reactivity against epidermal transglutaminase (TG3), characteristic for dermatitis herpetiformis, may develop during the continued exposure to gliadin via epitope spreading later in the disease course. Deposition of anti-TG3 IgA in the dermal papillae attracts neutrophils and their release of ROS and enzymes disrupts the DEJ.[28,56]

KEY CONCEPTS

Pathophysiology of Dermatitis Herpetiformis

- Dermatitis herpetiformis has a strong genetic disposition. While in most patients both anti-TG2 and TG3 reactivity develops, TG3 is the autoantigen of dermatitis herpetiformis.

TREATMENT

Pemphigus

Pemphigus Vulgaris und Pemphigus Foliaceus

Since the 1950s, oral corticosteroids have been the therapeutic backbone of pemphigus. The initial dosage is usually 0.5 to 1.0 mg/kg/day predniso(lo)ne in patients with moderate pemphigus (pemphigus disease area index >15 and ≤45), and 1.0 to 1.5 mg/kg/day in patients with severe pemphigus (pemphigus disease area index >45). In addition, potentially corticosteroid-sparing agents are applied. This can include azathioprine, mycophenoles, and cyclophosphamide,[57] although cyclophosphamide is less commonly used due to its toxicity. These regimens led to complete remissions in 15% to 20% of patients.[4] Since 2002, an increasing number of patients with severe and refractory pemphigus have been treated off-label with rituximab, a monoclonal antibody against the CD20 antigen on B lymphocytes that depletes CD20-positive B cells from the circulation for 6 to 12 months (Chapters 7 and 85). In a recent randomized controlled trial with newly diagnosed pemphigus vulgaris and pemphigus foliaceus patients, a clear superiority of rituximab over a standard corticosteroid regimen was demonstrated.[58] Subsequently, rituximab was approved by the Food and Drug Administration (FDA) (in 2018) and the European Medicines Agency (EMA) (in 2019) for the treatment of moderate to severe pemphigus vulgaris, and is recommended as first-line therapy for this group of patients.[57,59] In refractory patients, immunoadsorption, high-dose intravenous immunoglobulins, or intravenous corticosteroid pulses are recommended.[57,59]

Paraneoplastic Pemphigus

This pemphigus disease is most likely triggered by the underlying neoplasm. Consequently, oncological therapy is paramount. Systemic corticosteroids, rituximab, immunoadsorption, and high-dose intravenous immunoglobulins have been successfully used as treatment for the AIBD.[4]

THERAPEUTIC PRINCIPLES

Pemphigus

- Rituximab, or its biosimilar, is recommended as first-line therapy for moderate and severe pemphigus vulgaris and foliaceus. After repeated infusions, complete remission off therapy occurs in about 90% of patients.
- Multiple rituximab infusions are required in about a quarter of patients.
- Treated patients may develop severe adverse events including hypogammaglobulinemia, pneumonia, septicemia, and/or reactivation of chronic infections (e.g., herpes simplex, herpes zoster, viral hepatitis, or HIV). When rituximab is not available, the combination of long-term high-dose corticosteroids with potentially corticosteroid-sparing agents such as azathioprine or mycophenoles is recommended.

Pemphigoid Diseases

Bullous Pemphigoid

Systemic (usually prednis[ol]one, 0.5 mg/kg/day) or superpotent topical corticosteroids are the basis of BP treatment.[60–62] If applied as recommended, over 90% of the patients go into remission within 4 weeks.[60] If remission has been achieved, corticosteroids are tapered over several months to prevent relapse. Despite this, relapse occurs in 30% to 50% of the patients while tapering or after stopping corticosteroid treatment. The use of potentially corticosteroid-sparing adjuvant agents is still mainly based on few randomized controlled trials or case series and includes, with regional preferences, dapsone, doxycycline, azathioprine, mycophenoles, and methotrexate. In refractory or severe BP, anti-CD20, intravenous immunoglobulin (IVIG) infusions, and immunoadsorption have been applied successfully.[61–63] Based on a recent prospective and controlled clinical trial, doxycycline is an alternative to first-line corticosteroid treatment. Compared to corticosteroids, doxycycline treatment has fewer adverse events but is also less effective.[64]

Pemphigoid Gestationis

Since pemphigoid gestationis is self-limiting in nearly all patients, the treatment aims to relieve the symptoms and is usu-

ally based on oral (predniso[lo]ne, 0.25 to 0.5 mg/kg/day) and/or lesional topical corticosteroids. Since pemphigoid gestationis has been associated with prematurity and fetal growth restriction, a close cooperation with gynecologists is recommended.[29]

Anti-p200 Pemphigoid

In anti-p200 pemphigoid, similar therapeutics as in BP are applied. Usually, anti-p200 pemphigoid responds better than BP and prednis(ol)one (0.5 mg/kg/day) with or without dapsone, doxycycline or azathioprine are given. However, no data from randomized controlled trials are available.[22]

Mucous Membrane Pemphigoid

Treatment of MMP depends on affected mucosal sites as well as disease severity. The guideline of the European Academy of Dermatology and Venereology recommends treating mild MMP (e.g., limited to oral and nasal mucosa) with dapsone or tetracycline with or without topical corticosteroids, and in case of ineffectiveness, with oral corticosteroids combined with mycophenoles. In severe MMP (e.g., extended oral lesions, severe genital lesions, laryngeal, tracheal, or esophageal involvement), cyclophosphamide with or without oral corticosteroids is recommended as first-line approach followed by rituximab and high-dose IVIG.[65] In MMP with ocular involvement, a step-ladder approach including topical agents and various immunomodulants and immunosuppressants is recommended.[65]

Epidermolysis Bullosa Acquisita

Like severe MMP or its ocular involvement, epidermolysis bullosa acquisita is notoriously difficult to treat. In a retrospective case series, on average 9 months were required to induce remission despite generalized immunosuppressive therapy. Basically, similar treatments to those used in BP and pemphigus are employed, with systemic corticosteroids being the therapeutic backbones. In addition, colchicine may be used in mild cases. A recent meta-analysis of treatment outcomes in over 1000 patients showed that rituximab or high-dose IVIG were significantly more often associated with remission than other treatments.[66]

THERAPEUTIC PRINCIPLES

Pemphigoid Diseases

- Corticosteroids remain the backbone of pemphigoid disease treatment.
- The corticosteroid-sparing effect of the currently applied adjuvant immunomodulants and immunosuppressants has not been clearly established.
- Prevention of relapse after stopping corticosteroid treatment is challenging in bullous pemphigoid.
- Safer treatments of bullous pemphigoid are required. More effective treatments for mucous membrane pemphigoid and epidermolysis bullosa acquisita are needed.

Dermatitis Herpetiformis

Treatment consists of a gluten-free diet and dapsone. Dapsone resolves the usually considerable pruritus within 2 to 3 days, an effect so impressive that it has been employed for the diagnosis of the disease (*ex juvantibus*) before immunopathological tests were available. Dapsone, however, does not affect the intestinal pathology. Thus, initial therapy is recommended with a gluten-free diet in combination with dapsone, and tapering of dapsone after several months of strict gluten-free diet according to clinical symptoms.[28]

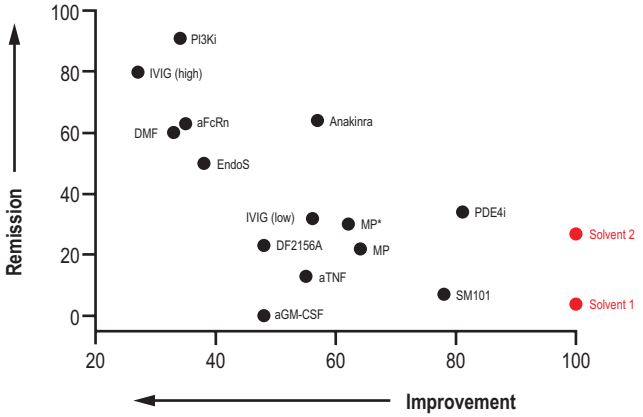


FIG. 63.18 Therapeutic Approaches in Experimental Pemphigoid Disease. Clinical responses to antiinflammatory agents applied in immunization-induced epidermolysis bullosa acquisita. In all experiments, epidermolysis bullosa acquisita was induced by immunization and treatments were started when the mice had 2% or more of their body surface area affected by skin lesions. Treatment periods vary from 2 to 6 weeks. In some experiments, SJL/J mice were used, while in others, B6.s mice were included. Remission is defined as the proportion of mice that had a lower affected body surface area at the end of the treatment period compared to the affected area at randomization. Improvement is defined as clinical disease severity of treated mice in relation to solvent-treated animals. If more than one dose of the drug was used, the most effective concentration is shown. *aFcRn*, anti-FcRn antibody; *aGM-CSF*, anti-GM-CSF antibody; *aTNF*, eternacept; *DF2156A*, allosteric CXCR1/2 inhibitor; *DMF*, dimethyl fumarate; *EndoS*, endoglycosidase derived from *Streptococcus pyogenes*; *IVIG (high)*, intravenous IgG administered every 3rd day; *IVIG (low)*, administered once per week; *MP**, MP from an independent experiment; *MP*, methylprednisolone; *PDE4i*, PDE4 inhibitor roflumilast; *PI3Ki*, phosphatidylinositol-3-kinase δ inhibitor LAS191954; *SM101*, recombinant soluble non-glycosylated version of the FcγRIIb. (Copyright [image and text]: Ralf Ludwig, used with permission.)

THERAPEUTIC PRINCIPLES

Dermatitis Herpetiformis

A gluten-free diet, initially combined with dapsone, will lead to complete remission of dermatitis herpetiformis.

ON THE HORIZON

In pemphigus, randomized controlled trials with inhibitors of neonatal FcR (FcRn), Bruton tyrosinkinase (BTK), and B-cell activating factor (BAFF or BLYS) are currently active. Phase Ib studies with DSG3-containing nanoparticles and DSG3-specific chimeric autoantigen receptor T cells are being conducted.[4,67] Study of pemphigoid disease model systems has led to the emergence of several novel therapeutic targets, including cytokines (e.g., TNF, GM-CSF, IL-17), signal transduction molecules (e.g., SYN, PDE4, PI3K), and several distinct targets linked to the complement system. New drugs for already well-known targets in pemphigoid disease (e.g., complement system) are being developed for the use in pemphigoid disease (Fig. 63.18). Ongoing clinical trials for BP are evaluating the efficacy and safety of ixekizumab (anti-IL-17, phase II), NPB-1 and efgartigimod (anti-FcRn, phase III), AC-203 (inflammasome inhibitor, phase II), bertilimumab (anti-eotaxin, phase II), benralizumb (anti-IL-5R, phase III), nomacopan (anti-C5a/ LTB4, phase III), and rituximab (anti-CD20, phase III).

ACKNOWLEDGMENTS

We thank Beke Linnemann, Lübeck, for help with the clinical pictures and Marina Kongsbak-Reim and Dr. Maike Holtsche, Lübeck, for the immunofluorescence photographs. We are indebted to Carolin Mahlerwein, Lübeck, for the graphical design of Fig. 63.1 and parts of Fig. 63.15, as well as to Katja Bieber, Lübeck, who generated Figs. 16 and 17.

REFERENCES

1. Schmidt E, Groves R. Immunobullous diseases. 9th ed. In: Griffith C, Barker J, Chalmers, Bleiker T, Creamer D, eds. *Rook's Textbook of Dermatology.* Chichester: Wiley-Blackwell; 2016:1–56.
2. Lever WF. Pemphigus. *Medicine.*1953;32:1–123.
3. Kasperkiewicz M, Ellebrecht CT, Takahashi H, et al. Pemphigus. *Nat Rev Dis Primers.* 2017;3:17026. Epub 2017/05/12.
4. Schmidt E, Kasperkiewicz M, Joly P. Pemphigus. *Lancet.* 2019;394(10201):882–894. Epub 2019/09/10.
5. Schmidt E, Zillikens D. Pemphigoid diseases. *Lancet*. 2013;381(9863): 320–332. Epub 2012/12/15.
6. Stanley JR, Amagai M. Pemphigus, bullous impetigo, and the staphylococcal scalded-skin syndrome. *N Engl J Med*. 2006;355(17):1800–1810.
7. Mahoney MG, Wang Z, Rothenberger K, et al. Explanations for the clinical and microscopic localization of lesions in pemphigus foliaceus and vulgaris. *J Clin Invest*. 1999;103(4):461–468. Epub 1999/02/18.
8. Ahmed AR, Carrozzo M, Caux F, et al. Monopathogenic vs multipathogenic explanations of pemphigus pathophysiology. *Exp Dermatol*. 2016;25(11):839–846. Epub 2016/10/30.
9. Sajda T, Hazelton J, Patel M, et al. Multiplexed autoantigen microarrays identify HLA as a key driver of anti-desmoglein and -non-desmoglein reactivities in pemphigus. *Proc Natl Acad Sci U S A*. 2016;113(7):1859–1864. Epub 2016/02/03.
10. Anhalt GJ, Kim SC, Stanley JR, et al. Paraneoplastic pemphigus. An autoimmune mucocutaneous disease associated with neoplasia. *N Engl J Med*. 1990;323(25):1729–1735.
11. Schepens I, Jaunin F, Begre N, et al. The protease inhibitor alpha-2-macroglobulin-like-1 is the p170 antigen recognized by paraneoplastic pemphigus autoantibodies in human. *PLoS One*. 2010;5(8):e12250. Epub 2010/09/02.
12. Goletz S, Zillikens D, Schmidt E. Structural proteins of the dermal-epidermal junction targeted by autoantibodies in pemphigoid diseases. *Exp Dermatol*. 2017;26(12):1154–1162. Epub 2017/09/10.
13. Warren SJ, Lin MS, Giudice GJ, et al. The prevalence of antibodies against desmoglein 1 in endemic pemphigus foliaceus in Brazil. Cooperative Group on Fogo Selvagem Research. *N Engl J Med*. 2000;343(1):23–30. Epub 2000/07/07.
14. Antiga E, Maglie R, Quintarelli L, et al. Dermatitis herpetiformis: novel perspectives. *Front Immunol*. 2019;10:1290. Epub 2019/06/28.
15. Vodo D, Sarig O, Sprecher E. The genetics of pemphigus vulgaris. *Front Med*. 2018;5:226. Epub 2018/08/30.
16. Sadik CD, Bischof J, van Beek N, et al. Genomewide association study identifies GALC as susceptibility gene for mucous membrane pemphigoid. *Exp Dermatol*. 2017;26(12):1214–1220. Epub 2017/11/15.
17. Qian Y, Jeong JS, Maldonado M, et al. Cutting edge: Brazilian pemphigus foliaceus anti-desmoglein 1 autoantibodies cross-react with sand fly salivary LJM11 antigen. *J Immunol.* 2012;189(4):1535–1539. Epub 2012/07/17.
18. Kridin K, Cohen AD. Dipeptidyl-peptidase IV inhibitor-associated bullous pemphigoid: a systematic review and meta-analysis. *J Am Acad Dermatol.* 2018. Epub 2018/10/09.
19. Kridin K, Ludwig RJ. The growing incidence of bullous pemphigoid: overview and potential explanations. *Front Med*. 2018;5:220. Epub 2018/09/05.
20. Forsti AK, Huilaja L, Schmidt E, Tasanen K. Neurological and psychiatric associations in bullous pemphigoid—more than skin deep? *Exp Dermatol.* 2017;26(12):1228–1234. Epub 2017/07/06.
21. Schulze F, Neumann K, Recke A, et al. Malignancies in pemphigus and pemphigoid diseases. *J Invest Dermatol.* 2015;135(5):1445–1447. Epub 2015/01/07.
22. Goletz S, Hashimoto T, Zillikens D, et al. Anti-p200 pemphigoid. *J Am Acad Dermatol.* 2014;71(1):185–191. Epub 2014/04/29.
23. van Beek N, Schmidt E. Autoimmune bullous diseases. In: Höger P, Kinsler V, Yan A, eds. *Harper's Textbook of Pediatric Dermatology*. 4th ed. Chichester: Wiley-Blackwell; 2020:868–897.
24. Amagai M. Pemphigus. In: Bologna JL, Schaffer JL, Cerroni L, eds. *Dermatology*. 4th ed. Amsterdam: Elsevier; 2018:494–509.
25. Kim JH, Kim SC. Paraneoplastic pemphigus: paraneoplastic autoimmune disease of the skin and mucosa. *Front Immunol*. 2019;10:1259. Epub 2019/06/20.
26. Bernard P, Borradori L, Pemphigoid group. In: Bologna JL, Scha$er JL, Cerroni L, eds. *Dermatology*. 4th ed. Amsterdam: Elsevier; 2018:510–526.

26a. Rashid H, Lamberts A, Borradori L, et al. European guidelines (S3) on diagnosis and management of mucous membrane pemphigoid, initiated by the European Academy of Dermatology and Venereology - Part I. *J Eur Acad Dermatol Venereol*. 2021 https://doi.org/10.1111/jdv.17397.

27. Lamberts A, Meijer JM, Jonkman MF. Nonbullous pemphigoid: a systematic review. *J Am Acad Dermatol.* 2018;78(5):989–995. e2. Epub 2017/11/06.
28. Hull C, Zone J. Dermatitis herpetiformis and linear IgA bullous dermatosis. 4th ed. In: Bologna JL, Schaffer JL, Cerroni L, eds. *Dermatology*. Amsterdam: Elsevier; 2018:527–537.
29. Huilaja L, Makikallio K, Tasanen K. Gestational pemphigoid. *Orphanet J Rare Dis.* 2014;9:136. Epub 2014/09/03.
30. Ludwig RJ. Clinical presentation, pathogenesis, diagnosis, and treatment of epidermolysis bullosa acquisita. *ISRN Dermatol*. 2013;2013:812029. Epub 2013/08/21.
31. Beek NV, Zillikens D, Schmidt E. Bullous Autoimmune Dermatoses: Clinical Features, Diagnostic Evaluation, and Treatment Options. *Dtsch Arztebl Int*. 2021;118(Forthcoming). https://doi.org/10.3238/arztebl.m2021.0136. arztebl.m2021.0136.
32. Terra JB, Meijer JM, Jonkman MF, et al. The n- vs. u-serration is a learnable criterion to differentiate pemphigoid from epidermolysis bullosa acquisita in direct immunofluorescence serration pattern analysis. *Br J Dermatol.* 2013;169(1):100–105. Epub 2013/03/16.
33. Pollmann R, Schmidt T, Eming R, et al. Pemphigus: a comprehensive review on pathogenesis, clinical presentation and novel therapeutic approaches. *Clin Rev Allergy Immunol.* 2018;54(1):1–25. Epub 2018/01/10.
34. Spindler V, Eming R, Schmidt E, et al. Mechanisms causing loss of keratinocyte cohesion in pemphigus. *J Invest Dermatol.* 2018;138(1):32–37. Epub 2017/10/19.
35. Anhalt GJ, Labib RS, Voorhees JJ, et al. Induction of pemphigus in neonatal mice by passive transfer of IgG from patients with the disease. *N Engl J Med*. 1982;306(20):1189–1196.
36. Amagai M, Tsunoda K, Suzuki H, et al. Use of autoantigen-knockout mice in developing an active autoimmune disease model for pemphigus. *J Clin Invest*. 2000;105(5):625–631.
37. Waschke J, Spindler V. Desmosomes and extradesmosomal adhesive signaling contacts in pemphigus. *Med Res Rev*. 2014;34(6):1127–1145. Epub 2014/02/20.
38. Lotti R, Atene CG, Marconi A, et al. Development of a desmocollin-3 active mouse model recapitulating human atypical pemphigus. *Front Immunol*. 2019;10:1387. Epub 2019/07/06.
39. Amber KT, Valdebran M, Grando SA. Non-desmoglein antibodies in patients with pemphigus vulgaris. *Front Immunol*. 2018;9:1190. Epub 2018/06/20.
40. Kasprick A, Bieber K, Ludwig RJ. Drug discovery for pemphigoid diseases. *Curr Protoc Pharmacol*. 2019;84(1):e55. Epub 2019/02/21.
41. Ludwig RJ, Vanhoorelbeke K, Leypoldt F, et al. Mechanisms of autoantibody-induced pathology. *Front Immunol*. 2017;8:603. Epub 2017/06/18.
42. Miodovnik M, Kunstner A, Langan EA, et al. A distinct cutaneous microbiota profile in autoimmune bullous disease patients. *Exp Dermatol.* 2017;26(12):1221–1227. Epub 2017/04/19.
43. Ellebrecht CT, Srinivas G, Bieber K, et al. Skin microbiota-associated inflammation precedes autoantibody induced tissue damage in experimental epidermolysis bullosa acquisita. *J Autoimmun.* 2016;68:14–22. Epub 2015/09/06.

44. Kasprick A, Hofrichter M, Smith B, et al. Treatment with anti-neonatal Fc receptor (FcRn) antibody ameliorates experimental epidermolysis bullosa acquisita in mice. *Br J Pharmacol.* 2020;177(10):2381–2392. Epub 2020/01/25.
45. Kamaguchi M, Iwata H, Nishie W, et al. The direct binding of collagen XVII and collagen IV is disrupted by pemphigoid autoantibodies. *Lab Invest.* 2019;99(1):48–57. Epub 2018/08/10.
46. Koga H, Prost-Squarcioni C, Iwata H, et al. Epidermolysis bullosa acquisita: the 2019 update. *Front Med.* 2018;5:362. Epub 2019/01/29.
47. Sadik CD, Schmidt E. Resolution in bullous pemphigoid. *Semin Immunopathol.* 2019;41(6):645–654. Epub 2019/11/17.
48. Kopecki Z, Ruzehaji N, Turner C, et al. Topically applied flightless I neutralizing antibodies improve healing of blistered skin in a murine model of epidermolysis bullosa acquisita. *J Invest Dermatol.* 2013;133(4): 1008–1016. Epub 2012/12/12.
49. Iwata H, Kamaguchi M, Ujiie H, et al. Macropinocytosis of type XVII collagen induced by bullous pemphigoid IgG is regulated via protein kinase C. *Lab Invest.* 2016;96(12):1301–1310. Epub 2016/10/25.
50. Liu Z, Diaz LA, Troy JL, et al. A passive transfer model of the organ-specific autoimmune disease, bullous pemphigoid, using antibodies generated against the hemidesmosomal antigen, BP180. *J Clin Invest.* 1993;92(5):2480–2488.
51. Heppe EN, Tofern S, Schulze FS, et al. Experimental laminin 332 mucous membrane pemphigoid critically involves C5aR1 and reflects clinical and immunopathological characteristics of the human disease. *J Invest Dermatol.* 2017;137(8):1709–1718. Epub 2017/05/01.
52. Nishie W, Sawamura D, Goto M, et al. Humanization of autoantigen. *Nat Med.* 2007;13(3):378–383.
53. Hirose M, Recke A, Beckmann T, et al. Repetitive immunization breaks tolerance to type XVII collagen and leads to bullous pemphigoid in mice. *J Immunol.* 2011;187(3):1176–1183. Epub 2011/06/28.
54. Kasprick A, Holtsche MM, Rose EL, et al. The anti-C1s antibody TNT003 prevents complement activation in the skin induced by bullous pemphigoid autoantibodies. *J Invest Dermatol.* 2018;138(2):458–461. Epub 2017/09/14.
55. Kasperkiewicz M, Sadik CD, Bieber K, et al. Epidermolysis bullosa acquisita: from pathophysiology to novel therapeutic options. *J Invest Dermatol.* 2016;136(1):24–33. Epub 2016/01/15.
56. Sardy M, Karpati S, Merkl B, et al. Epidermal transglutaminase (TGase 3) is the autoantigen of dermatitis herpetiformis. *J Exp Med.* 2002;195(6):747–757.
57. Joly P, Horvath B, Patsatsi A, et al. Updated S2K guidelines on the management of pemphigus vulgaris and foliaceus initiated by the european academy of dermatology and venereology (EADV). *J Eur Acad Dermatol Venereol.* 2020;34(9):1900–1913. https://doi.org/10.1111/jdv.16752. Epub 2020 Aug 24.
58. Joly P, Maho-Vaillant M, Prost-Squarcioni C, et al. First-line rituximab combined with short-term prednisone versus prednisone alone for the treatment of pemphigus (Ritux 3): a prospective, multicentre, parallel-group, open-label randomised trial. *Lancet.* 2017;389(10083):2031–2040. Epub 2017/03/28.
59. Murrell DF, Peña S, Joly P, et al. Diagnosis and management of pemphigus: Recommendations of an international panel of experts. *J Am Acad Dermatol.* 2020 Mar;82(3). https://doi.org/10.1016/j.jaad.2018.02.021. 575–585.e1. Epub 2018 Feb 10.
60. Joly P, Roujeau JC, Benichou J, et al. A comparison of oral and topical corticosteroids in patients with bullous pemphigoid. *N Engl J Med.* 2002;346(5):321–327.
61. Feliciani C, Joly P, Jonkman MF, et al. Management of bullous pemphigoid: the European Dermatology Forum consensus in collaboration with the European Academy of Dermatology and Venereology. *Br J Dermatol.* 2015;172(4):867–877. Epub 2015/04/02.
62. Eming R, Sticherling M, Hofmann SC, et al. S2k guidelines for the treatment of pemphigus vulgaris/foliaceus and bullous pemphigoid. *J Dtsch Dermatol Ges.* 2015;13(8):833–844. Epub 2015/07/28.
63. Sticherling M, Franke A, Aberer E, et al. An open, multicentre, randomized clinical study in patients with bullous pemphigoid comparing methylprednisolone and azathioprine with methylprednisolone and dapsone. *Br J Dermatol.* 2017;177(5):1299–1305. Epub 2017/05/12.
64. Williams HC, Wojnarowska F, Kirtschig G, et al. Doxycycline versus prednisolone as an initial treatment strategy for bullous pemphigoid: a pragmatic, non-inferiority, randomised controlled trial. *Lancet.* 2017;389(10079):1630–1638. Epub 2017/03/11.
65. Schmidt E, Rashid H, Marzano AV, et al. European Guidelines (S3) on diagnosis and management of mucous membrane pemphigoid, initiated by the European Academy of Dermatology and Venereology - Part II. *J Eur Acad Dermatol Venereol.* 2021 https://doi.org/10.1111/jdv.17395.
66. Iwata H, Vorobyev A, Koga H, et al. Meta-analysis of the clinical and immunopathological characteristics and treatment outcomes in epidermolysis bullosa acquisita patients. *Orphanet J Rare Dis.* 2018;13(1):153. Epub 2018/09/06.
67. Ellebrecht CT, Bhoj VG, Nace A, et al. Reengineering chimeric antigen receptor T cells for targeted therapy of autoimmune disease. *Science.* 2016;353(6295):179–184. Epub 2016/07/02.

64

Immunology of Psoriasis

Cristina Albanesi

CLINICAL AND HISTOLOGIC FEATURES OF PSORIASIS

Psoriasis is a common chronic, relapsing immune-mediated disease that involves the skin and small joints of genetically predisposed individuals. It affects approximately 2% of the general population, with more than half of patients presenting in the first three decades of life.

There is a wide spectrum of cutaneous manifestations. The clinical spectrum of psoriasis includes plaque, guttate, small plaque, inverse, erythrodermic, and pustular variants.[1] The most common and well-recognized morphologic presentation of psoriasis is the plaque type, which can vary from pinpoint to large lesions (Fig. 64.1).

The disease is characterized by the formation of demarked erythematous plaques with large scaling. The scales are a result of a hyperproliferative epidermis with premature maturation of keratinocytes and incomplete cornification. There is retention of nuclei in the stratum corneum (parakeratosis). The mitotic rate of the basal keratinocytes is increased when compared with normal skin. This leads to a thickening of the epidermis (acanthosis), which is marked by elongated rete ridges that form fingerlike protrusions into the dermis. The granular layer of the epidermis (the starting site of terminal keratinocyte differentiation) is strongly reduced or missing. The epidermis is infiltrated by neutrophils and activated CD8 T lymphocytes. Within the dermis, there is an inflammatory infiltrate composed mainly of $CD3^+$ T cells, dendritic cells (DCs), macrophages, mast cells, and neutrophils. Elongated and dilated blood vessels in the dermal papillae represent a further histologic hallmark of psoriatic skin lesions (Fig. 64.2).[1]

KEY CONCEPTS

- Psoriasis is a common chronic-relapsing immune-mediated skin disease affecting approximately 2% of the general population.
- A complex set of gene variants, rather than a single gene, predispose the patient to an aberrant response to environmental factors.
- There is infiltration of effector immune cells into both the epidermis and dermis, which promotes hyperproliferation of the epidermis, premature maturation of keratinocytes, and incomplete cornification.
- The epidermis is thickened, with elongated rete ridges forming protrusions into the dermis.
- Primary effector cells are dermal dendritic cells (DCs), in particular plasmacytoid DCs (pDCs), whose activation can depend on DNA-LL37 or RNA-LL37 complexes released by injured keratinocytes. This leads to a massive production of interferon (IFN)-α.
- pDC-released IFN-α or RNA-LL37 complexes released by keratinocytes activate myeloid DCs (mDCs), which in turn induces type 1 and type 17 T-cell responses.
- A T helper 22 (Th22) response is also pathogenetically induced.
- Pathologic cytokines include T cell–derived lymphokines, such as IFN-γ, tumor necrosis factor (TNF)-α, interleukin (IL)-17, IL-22, and antigen-presenting cell–derived cytokines, such as IL-12 and IL-23.
- T lymphocytes and DC establish the inflammatory cytokine milieu influencing keratinocyte proliferation and immune responses.
- Intrinsic alterations of keratinocytes in the activation of signal transduction pathways (i.e., STAT3, RAS) further amplify psoriatic processes.

IMMUNE-RELATED GENETIC FACTORS PREDISPOSING TO PSORIASIS

The genetic basis of psoriasis has long been recognized, because family members of patients with psoriasis are at greater risk of developing the disease. The concordance rate of psoriasis is approximately 70% in monozygotic twins and 20% in dizygotic twins, depending on the study and population.[2] The mode of inheritance is complex. It is thought that rather than a single disease gene there is a complex set of gene variants that lead to an aberrant response to environmental factors. Several allelic variants or single nucleotide polymorphisms (SNPs) have been identified. Classic genome-wide association studies (GWAS) have pointed to additional susceptibility loci and polymorphisms. At least 34 chromosomal loci have been identified. These have been labeled as *psoriasis susceptibility loci* 1 through 34 (*PSORS1* through *PSORS34*).[3]

The genes identified by GWAS can be grouped into four signal transduction and gene regulation pathways involved in epidermal differentiation, in inflammation, and in innate and adaptive immunity (Fig. 64.3).[3] First, skin barrier function pathways have been strongly associated with psoriasis. Several studies have identified the late cornified envelope *(LCE)* gene cluster, which is highly expressed in psoriatic skin, and corneodesmosin *(CDSN)* and coiled-coil, x-helical rod protein 1 (*CCHCR1)*, which are highly polymorphic genes that map to the *PSORS1* locus. Candidate gene studies have also implicated the β-defensin gene cluster *(DEFB)* on chromosome 8 and interferon (IFN) induced with helicase C domain 1 *(IFIH1)*, which are both involved in the protection against microbial agents. The second pathway involved candidate genes found within nuclear factor κB (NF-κB) genes and their related pathways, as well as in a genetic region involved in inflammatory responses and in the modulation of Th immune responses. Association studies point to tumor necrosis factor (TNF)–induced protein 3 *(TNFAIP3)* and TNFAIP3-interacting protein 1 *(TNIP1)*, which regulate the activity of NF-κB and can induce keratinocyte hyperproliferation. In

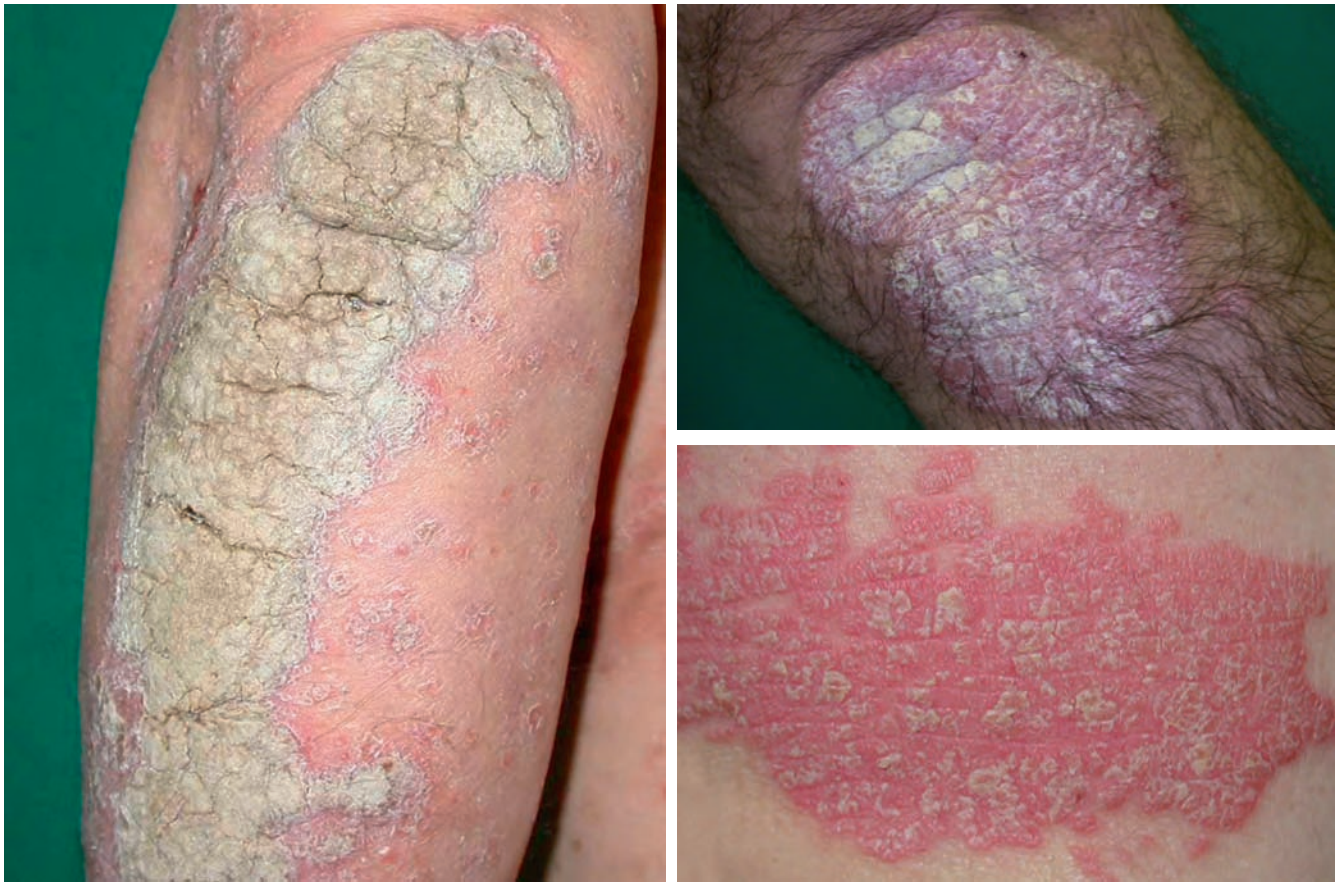


FIG. 64.1 Clinical Features of Plaque Psoriasis. Scaly, erythematous, sharply demarcated plaques in different sizes and shapes are hallmarks of psoriasis.

Caucasian populations, association studies have pointed to TNF receptor–associated factor 3–interacting protein 2 *(TRAF3IP2)*, a gene that encodes a protein that interacts with the *NF-κB/REL* subunit.[3] The third and potentially largest pathway, in terms of numbers of genetic loci implicated in psoriasis, influences T-cell signaling. Gene loci include *SOCS1* (suppressor of cytokine signaling1), *STAT3/5 A/5B* (signal transducer and activator of transcription3/5 A/5B), interleukin *(IL)-12B*, *IL-23A*, and *IL-23R*. The latter gene is involved in the regulation of Th17 and Th1 responses. Specific haplotypes of *CTLA4* (cytotoxic T-lymphocyte antigen), which encodes a protein that downregulates activation of T lymphocytes, have been associated.[3] The fourth pathway is linked to antigen presentation, and involves human leukocyte antigen *(HLA)-C* and *ERAP1*.[3]

As is the case in many immune-mediated diseases, the major genetic determinant of psoriasis, which contains the majority of allelic variants is located within the HLA complex on chromosome 6p21 (see Chapter 5). In the case of psoriasis, this is identified as the *PSORS1* locus. This locus spans the class I region of the major histocompatibility complex (MHC) and encompasses nine genes, including *HLA-C, CDSN*, *CCHCR1*, and *TNFA*, that are highly polymorphic. Among these, the *HLA-Cw6* allele is the susceptibility genetic factor most highly associated with psoriasis, even though more than 100 HLA-C variants have been described. High-density polymorphism analysis has identified several SNPs within the minimal promoter region of HLA-C, which affect HLA-C expression. Among HLA-C haplotypes, *HLA-Cw6* is associated with early-onset (type I), compared with late-onset (type II) psoriasis.[4] In addition to HLA-Cw6, significant associations with HLA-B13, HLA-B17, HLA-B37, HLA-B57, HLA-Cw7, and class II molecules HLA-DR4 and HLA-DR7 in psoriasis have been identified. Exciting evidence has emerged for an interaction between HLA-C and ERAP1 (involved in trimming peptides to enable effective loading and processing onto MHC class I) in psoriasis.

Despite the strong genetic evidence and the obvious immunologic function of HLA-C to regulate both innate and adaptive responses, functional studies addressing the precise mechanism by which *HLA-Cw6* alleles predispose to psoriasis are still missing. HLA-C interacts with natural killer (NK) cell KIR receptors, and thus it is possible that susceptibility to psoriasis reflects differences in both antigen presentation and NK cell regulation.[4]

EFFECTOR CELLS AND IMMUNE MECHANISMS OPERATING IN PSORIASIS

The primary pathogenic mechanism involved in the onset of psoriasis in predisposed individuals appears to follow their exposure to certain trigger factors (Fig. 64.4). These range from nonspecific stimuli, such as skin trauma (termed the *Koebner effect*), to more specific triggers, such as pathogens (i.e., streptococci) or drugs (i.e., lithium, IFN-α). All of these factors

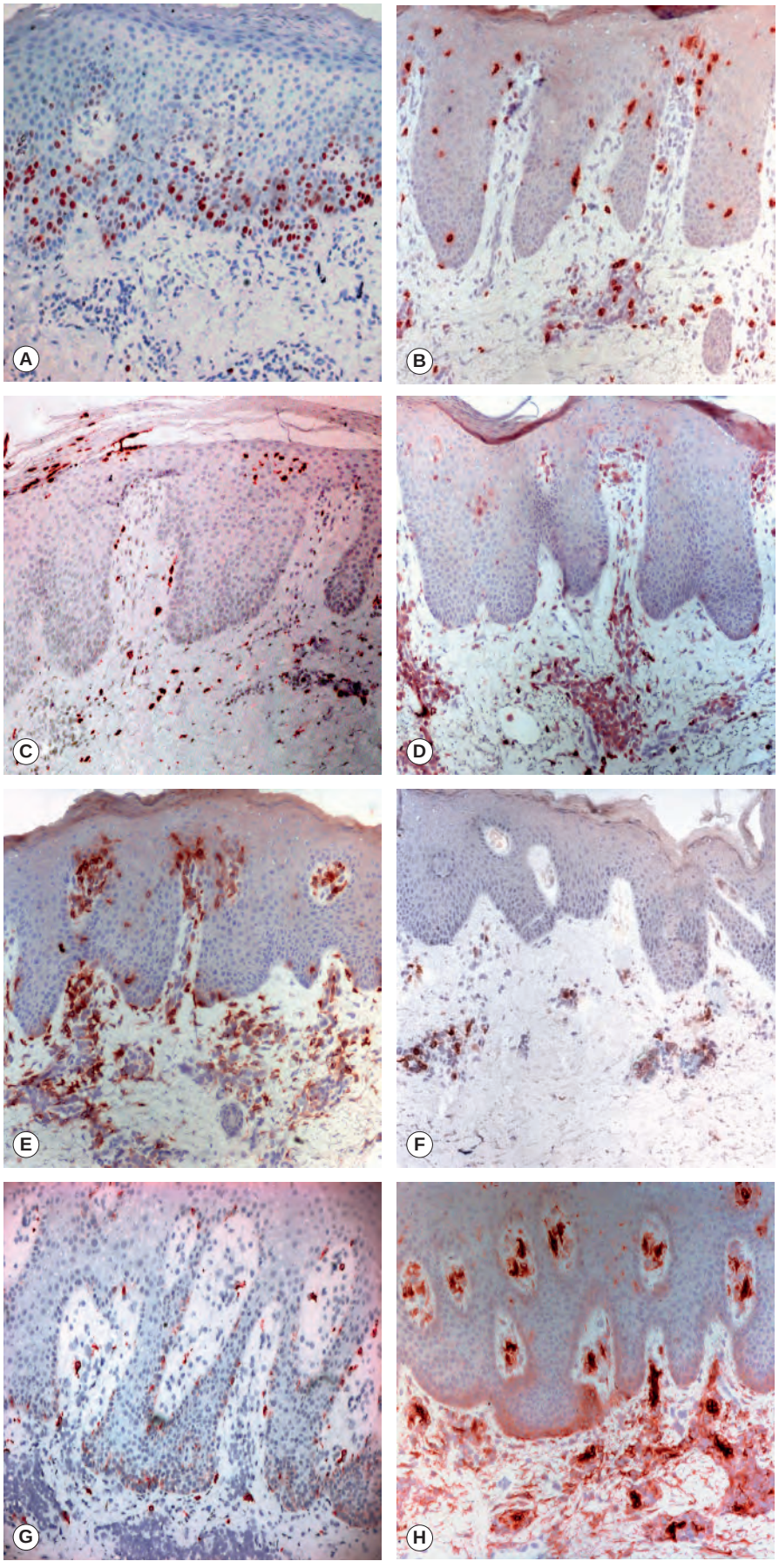


FIG. 64.2 Histologic Components of a Mature Psoriatic Plaque. Psoriatic skin lesions are characterized by a hyperproliferative epidermis showing an increased mitotic rate of the basal keratinocytes (A, Ki67 immunostaining). As a consequence, the epidermis thickens, with elongated rete ridges that form typical fingerlike protrusions into the dermis. The epidermis becomes infiltrated by activated CD8 T lymphocytes and neutrophils (B and C, immunostaining for CD8 and CD15, respectively). Within the dermis, an inflammatory infiltrate mainly composed of $CD3^+$ T cells (D), $CD11c^+$ dendritic cells (E), BDCA-2^+ plasmacytoid dendritic cells (F), c-kit^+ mast cells (G), and neutrophils (C) is observed. Elongated and dilated ICAM-1^+ (H) blood vessels in the dermis represent another histologic hallmark of psoriatic skin.

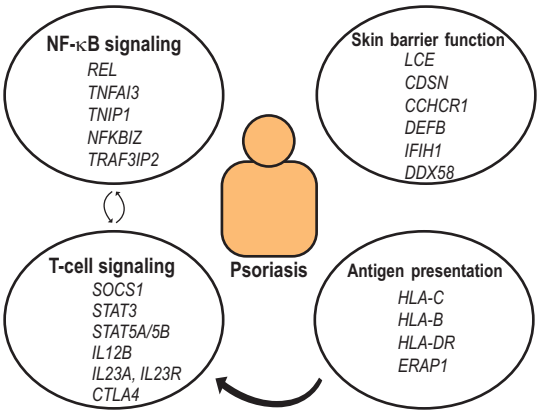


FIG. 64.3 The Psoriasis Genetic Pathway Identified by Genome-Wide Association Studies. A schematic representation of genetic variants within psoriasis-related risk loci that are highly represented in the psoriatic population (allele frequency ≥0.3). They can be grouped into broad immunologic intersected processes, including those involved in skin barrier function, inflammatory pathways (NF-κB pathway) and immune responses (T-cell signaling and antigen presentation). *Arrows* signify the cross-talk between the immune pathways shown. *CDSN*, Corneodesmosin; *DEFB*, β-defensin gene cluster; *HLA*, human leukocyte antigen; *LCE*, late cornified envelope; *NF-κB*, nuclear factor kappa B.

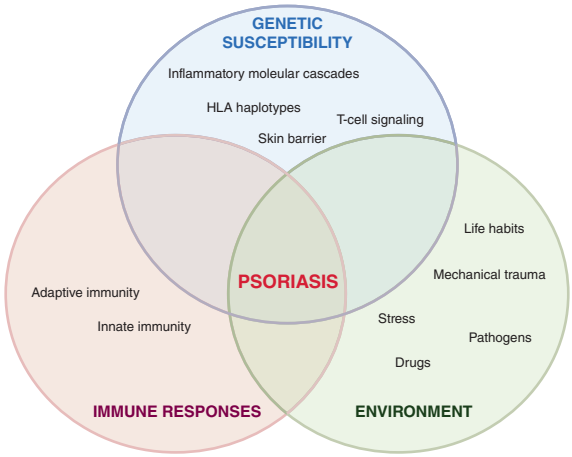


FIG. 64.4 Extrinsic and Intrinsic Factors That Can Contribute to the Development of Psoriasis. The disease occurs in genetically predisposed individuals that carry one or more allelic variants of psoriasis susceptibility genes (i.e., genes involved in skin barrier function, inflammatory and T-cell signaling, and determining specific human leukocyte antigen [*HLA*] haplotypes). Dysregulated innate and adaptive immune responses can follow exposure to certain environmental triggers.

generate a pathogenic cascade that culminates in the expansion of lesional and/or circulating T cells in the psoriatic skin.[5,6]

In the past, much effort has been devoted to understanding the link between the trigger stimuli and the pathogenic T-cell cascade that leads to psoriasis. A number of studies have demonstrated that type-1 IFN represents this link.[5,7] The prototypical type I IFN, IFN-α, is abundantly produced by plasmacytoid dendritic cells (pDCs) during the acute phase of psoriasis (Fig. 64.5, A). In turn, IFN-α indirectly stimulates keratinocyte immune activation and maturation of myeloid DCs (mDCs), with the consequent beginning of an adaptive immune response phase. This results in the establishment of an IL-23/IL-17 and IL-12/IFN-γ inflammatory environment in psoriatic skin, with DC-derived IL-23 and IL-12 promoting Th17 and Th1 cell effector functions, respectively.[6,8,9]

Innate lymphoid cells (ILCs) and innate-like γδ T cells are also critical contributors to plaque development. They release considerable levels of type 17 cytokines.[10] Mast cells and neutrophils represent additional, innate sources of IL-17 in psoriatic skin.[5]

The leukocyte infiltrate present in active psoriatic skin establishes a cytokine milieu that dictates specific and pathogenic gene signatures in resident skin cells (see Fig. 64.5, B). Cytokine-activated keratinocytes overexpress a number of inflammatory mediators that aberrantly amplify and sustain the psoriasiform tissue reactions (Fig. 64.5, C). Intrinsic defects and/or alterations of keratinocytes in their immune response to proinflammatory cytokines are fundamental to the induction of psoriatic processes, as demonstrated in genetically manipulated mouse systems.

Plasmacytoid Dendritic Cells as Inducers of Primary Immune Responses in Psoriasis

pDCs (see Chapter 6) are characterized by a plasma cell morphology and a distinctive surface phenotype ($CD4^+$, $CD45RA^+$, $CD123^+$, $BDCA\text{-}2^+$, $BDCA\text{-}4^+$, $CD62L^+$, cutaneous lymphocyte-associated antigen $[CLA]^+$, and $CD11c^-$). They are considered key effector cells in antiviral defense due of their ability to produce large amounts of type I IFN.[11] Upon viral stimulation, pDCs differentiate into a unique type of mature DC and induce an IFN-α–dependent activation of bystander mDCs with the ability to induce Th1-responses (Chapter 11), thus providing a necessary link between innate and adaptive immunity (see Chapter 3).[11]

Several studises have demonstrated that pDCs infiltrate psoriatic skin and that pDC-derived IFN-α initiates the expansion of autoimmune T cells, leading eventually to the skin lesions of psoriasis.[5,7,12,13] Blocking of type I IFN signaling with neutralizing antibodies to IFN-α/β receptors (IFN-AR1 and 2) or BDCA-2 inhibits the activation and expansion of pathogenic T cells, as well as the development of a psoriatic phenotype. The mechanisms responsible for the IFN-α-induced expansion of T cells in psoriasis depend on the capability of IFN-α to favor cross-presentation of sequestered tissue-specific autoantigens by mDCs, and to enhance the survival of autoreactive Th1-cell bias through the induction of T-bet and IL-12$R\beta_2$ expression.[13] The pathogenic role of IFN-α is also suggested by the observations that its signaling signature (i.e., type I IFN, IFNAR1, IFNAR2, STAT1, IRF1, and IRF7) is present in resident skin cells of psoriatic plaques[12,14] and that psoriasis is exacerbated in patients treated with recombinant IFN-α for unrelated conditions (i.e., viral infections or tumors), or with imiquimod, a Toll-like receptor (TLR) agonist that induces production of IFN-α.

The molecular mechanisms leading pDCs to produce type I IFN involve the activation of TLR7 and TLR9 (see Chapter 3), intracellular receptors that recognize viral/microbial nucleic acids within endosomal compartments. In psoriatic skin, pDCs can be activated to produce massive amounts of type I IFN in response to extracellular self-DNA coupled to the endogenous antimicrobial peptide LL37, which is known to be overexpressed in psoriatic skin.[15] LL37 breaks innate tolerance to self-DNA by

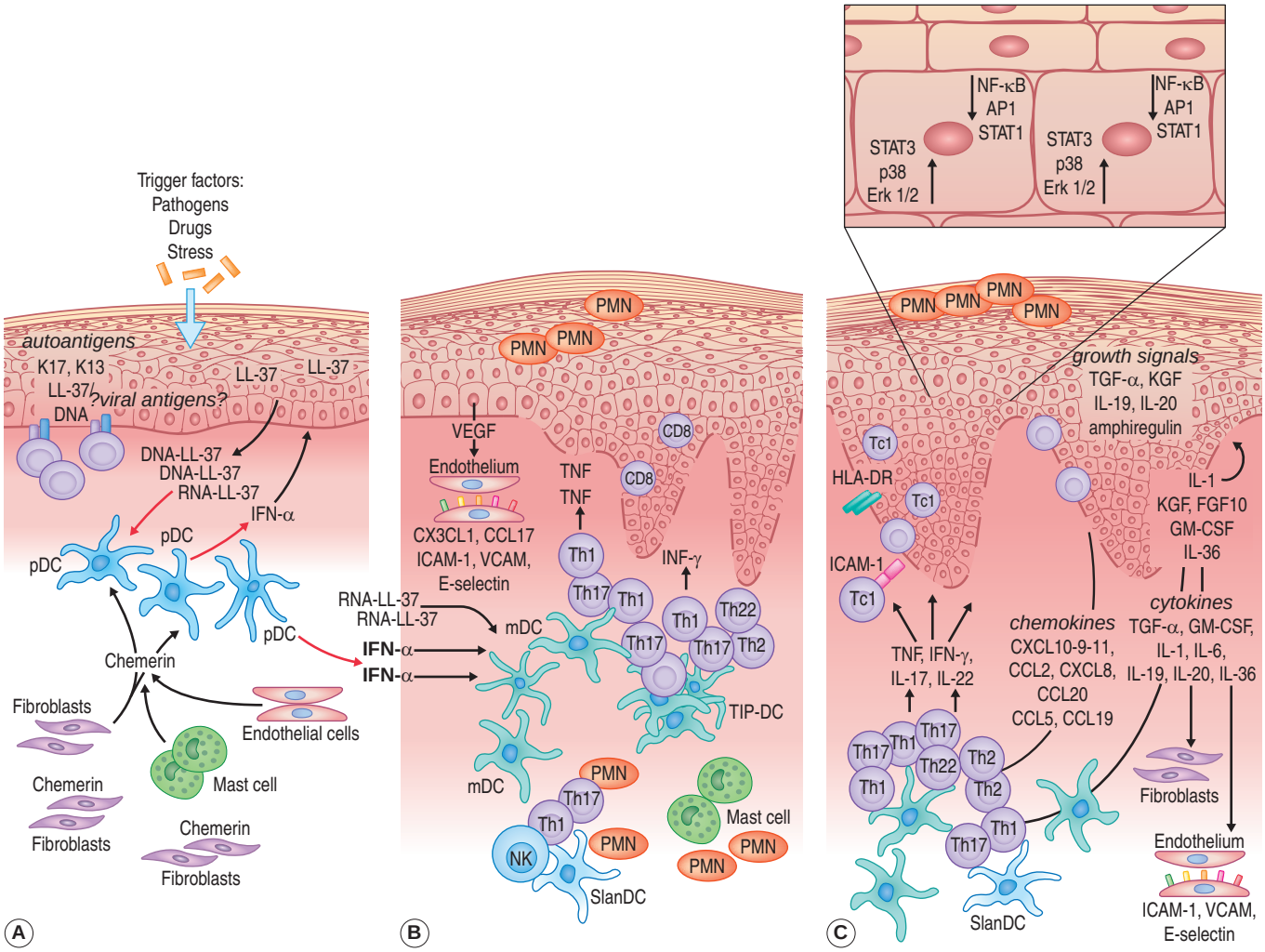


FIG. 64.5 Scheme of Pathogenic Mechanisms Operating in Psoriasis. The psoriatic lesion starts to evolve after keratinocytes are injured (e.g., by physical trauma or bacterial products). Thereafter, a cascade of events, including the formation of complexes formed by keratinocyte-derived DNA and the cathelicidin LL37, leads to the activation of plasmacytoid dendritic cells *(pDCs)* that routinely patrol psoriatic skin (A). In the early phase of disease development, other pDCs are recruited by the chemokine chemerin, which is derived primarily from dermal fibroblasts and, to less extent, mast cells and endothelial cells. These pDCs are induced to release high amounts of interferon *(IFN)-α*. IFN-α locally activates keratinocytes and participates in the activation processes affecting mDCs. Subsequently, DCs migrate into draining lymph nodes and induce the differentiation of naïve T cells into effector cells, such as type 17 T-helper *(Th17)* or type 17 T- cytotoxic *(Th17) (Tc)* cells and type 1 Th1 or Tc1 cells. These effector T cells recirculate into psoriatic skin, proliferate, and produce massive amounts of proinflammatory cytokines, such as IFN-γ and tumor necrosis factor *(TNF)*. Effector cells also include innate immunity cells, such as innate lymphoid cells *(ILC)* and γδ T cells, which routinely patrol the skin (B). TNF is abundantly released by dermal DCs, mainly represented by inflammatory DCs. Skin Langerhans DCs also reinforce immunity in psoriasis by interacting with and potentiating the activity of both neutrophils (PMN) and NK cells, as well as inducing Th1 and Th17 responses. IFN-γ, TNF, and interleukin *(IL)*-17 are responsible for the activation of resident skin cells, in particular keratinocytes, which respond to cytokines with a stereotypical set of genomic responses leading to synthesis of inflammatory mediators (C). Keratinocytes are also targets of T cell–derived IL-22 that, together with IL-17, induces proliferation and de-differentiation of psoriatic keratinocytes in a signal transducer and activator of transcription 3 *(STAT3)*-dependent manner. Keratinocyte-derived chemokines, cytokines, and membrane molecules have a major role in maintaining the recruitment of leukocytes into inflammatory sites. Because of their intrinsic defects, psoriatic keratinocytes aberrantly respond to cytokines and show altered intracellular signaling pathways, including STAT3 and RAS cascade (C). The uncontrolled hyperproliferation and differentiation observed in psoriatic skin could also derive from dysregulated production of tissue growth factors and regulators, such as transforming growth factor *(TGF)*-α, keratinocyte growth factor *(KGF)*, amphiregulin, granulocyte macrophage–colony-stimulating factor *(GM-CSF)*, fibroblast growth factor-10 *(FGF-10)*, *IL*-19, IL-20, and IL-36 produced by keratinocytes and fibroblasts. Psoriatic keratinocytes produce and capture from melanocytes autoantigens (i,e., nucleic acid/LL37 complexes, ADAMTSL5, PLA2G4A) capable of inducing clonal T-cell responses. Finally, the inflammatory cytokine milieu also influences the immune functions of fibroblasts and endothelium, with the latter being critical for leukocyte trafficking and extravasation.

forming aggregated and condensed structures that can trigger a robust IFN-α induction via TLR9, as well as form complexes with RNA and activate pDCs through TLR7. In parallel, LL37/RNA can alert mDCs through their TLR8, driving T-cell activation and production of cytokines found in psoriasis. This finding suggests a fundamental role for LL37 in alerting resident skin pDCs of tissue damage associated with cell death and the release of self-DNA.[15]

pDCs are typically absent in unperturbed skin and peripheral tissues under homeostatic conditions. However, they can be detected in prepsoriatic skin during the acute phase of the disease.[13] At this stage, BDCA-2$^+$ pDCs infiltration correlates with the transient high IFN-α local expression and with massive presence of innate immunity cells (i.e., neutrophils, mast cells, macrophages) in the skin. On the contrary, pDCs are almost absent in long-lasting and chronicized lesions.[12,13] Importantly, pDC recruitment in psoriatic skin is strictly associated with the expression of the chemokine (see Chapter 15) chemerin, which is temporally produced by dermal fibroblasts and active during psoriatic plaque development. pDC migration toward fibroblast-derived chemerin is completely dependent on the expression of ChemR23 receptor on pDC. Compared with other chemokines potentially active on pDC (CXCL10 and CXCL12), chemerin is the main, if not the only, protein responsible for the pDC chemotactic activity released by fibroblasts in psoriatic skin.[12]

Dendritic Cell Driving of T-Cell Responses in Psoriatic Skin

Although pDCs are responsible for triggering psoriasis, mDCs are the main amplifiers of local inflammation. Dermal mDCs are dramatically increased in psoriasis. Targeted immunotherapy reduces their quantity, supporting the concept that mDCs have a key pathogenetic role in psoriasis.[13,16] Dermal mDCs are found at the dermal–epidermal junction, as well as throughout the whole dermis. mDCs are able to capture extracellular antigens for presentation to T cells and also intracellular antigens from adjacent cell types via cross-presentation (see Chapter 6). mDCs within psoriatic lesions are intrinsically stronger stimulators of T-cell proliferation compared with DCs derived from peripheral blood or from the skin of healthy patients. mDCs uniformly express CD11c, and they can be further subdivided on the basis of expression of CD1c (BDCA-1). Steady-state skin has a predominance of CD11c$^+$ CD1c$^+$ resident DCs, whereas CD11c$^+$ CD1c$^-$ mDCs predominate in psoriatic inflammation.[13,16]

A small fraction of CD11c$^+$ CD1c$^+$ DC bears "maturation" markers, such as DC-LAMP, CD83, and endocytic receptor DEC-205/CD205, suggesting that they could function as conventional DCs and present antigens to T cells to trigger acquired immune responses.[13,16] These rare, phenotypically mature cells, often aggregating in dermal clusters, could be required for rapid antigen presentation to local T cells or for ongoing "micro-" immune responses.

During psoriasis development, dermal CD11c$^+$ DC mature and acquire a CD1c$^-$ HLA-DR$^+$ CD45$^+$ CD14$^-$ DC-specific ICAM-3–grabbing nonintegrin (DC-SIGN)$^+$ phenotype.[13] These inflammatory mDCs are CCR7$^+$ and respond to the chemokine CCL19, suggesting that they may migrate to draining lymph nodes for antigen presentation.

CD11c$^+$ CD1c$^-$ inflammatory DC express very high levels of the enzyme inducible nitric oxide synthase (iNOS) and TNF-α. These DC can be considered the human equivalent of TIP-DC (TNF-α and iNOS-producing DC), which have been shown in mice to have effector functions in clearing some bacterial infections.[13]

TNF-α in turn targets resident skin cells, including keratinocytes and endothelial cells. TNF-α induces expression of ICAM-1 on keratinocytes, facilitating the adhesion of circulating leukocytes. Moreover, TNF-α can stimulate keratinocytes and dermal fibroblasts to produce the potent neutrophil chemoattractant CXCL8, as well as the proinflammatory cytokines IL-6 and IL-1, which help to generate and maintain Th17 cells (see Chapters 11 and 14).[14]

Inflammatory DCs also produce other cytokines (*e.g.*, IL-23 and IL-12), which are strictly linked to psoriasis. IL-12 mainly induces IFN-γ production and Th1/Tc1 polarization, whereas IL-23 also stimulates IL-17 and IL-22 release by type 17 and type 22 T cells.[6,8,9] Evidence for a role of both TNF and IL-23 in psoriasis includes the therapeutic success of the TNF-α blockers and, more relevantly, the monoclonal antibodies against p40 or p19 subunits of IL-23.[6,8]

Another population of inflammatory mDCs, defined by the selective expression of the 6-sulfo LacNAc residue on the P-selectin glycoprotein ligand 1 membrane molecule, has been identified in psoriatic skin.[17] 6-Sulfo LacNAc$^+$ DCs (slanDCs) have a well-defined phenotype (CD1c$^-$, CD11c$^+$, CD16$^+$, CD14$^-$) that clearly distinguishes them from classic CD1c$^+$ blood DC (CD1c$^+$, CD11c$^+$) or pDC (BDCA-2, BDCA-4). SlanDCs produce more TNF-α, IL-23, IL-12, IL-1β, and IL-6 and can thus induce Th1/Th17 cells. SlanDCs also reinforce innate immunity in psoriasis by interacting with and potentiating the activity of neutrophils (Chapter 39) and NK cells (Chapter 12).[17]

The function of another DC subset, the Langerhans cell (Chapter 23), is still controversial. Langerhans cells are found only in the epidermal compartment, where they are similar in number and phenotype in lesional and nonlesional skin. However, they fail to migrate in response to proinflammatory stimuli (see Chapter 24). Recent studies have shown that Langerhans cells can produce IL-23 as inflammatory DC and are located in the close proximity to pathogenic skin T cells. Interestingly, Langerhans cells from patients who achieved almost complete remission after anti-TNF-α treatment can release higher levels of IL-23 when compared with healthy volunteers. This observation supports the view that Langerhans cells may be associated with the recurrence of psoriasis.[16]

Activation of T Lymphocytes and Establishment of the Cytokine Milieu Influencing Keratinocyte Proliferation and Immune Functions

Psoriasis lesional skin contains many inflammatory CD3$^+$ T cells in both the papillary dermis and the epidermis, determining the epidermal hyperplasia and inflammation picture (Fig. 64.6). Immunophenotyping of these T cells shows that they are primarily activated memory T cells that express the cutaneous lymphocyte antigen (CLA). They belong to distinct subsets of CD4$^+$ and CD8$^+$, Th1/Tc1, Th17/Tc17, and Th22/Tc22 lymphocyte subsets.

Psoriatic T-cell responses are triggered in response to (auto) antigens primarily derived from keratinocytes. They are captured by DCs located at the dermal–epidermal junction of psoriatic skin. To date, three autoantigens have been identified in keratinocytes and found to be involved in the pathogenesis of psoriasis.[4,14,18] In up to 75% of psoriatic patients, LL37 is recognized

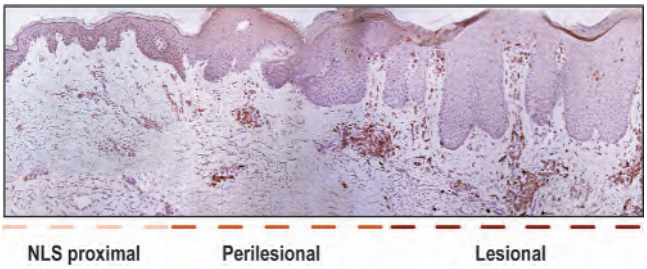


FIG. 64.6 CD3⁺T Lymphocyte Pattern in Psoriasis. Both acute and chronic phases of psoriasis can be found within the same psoriatic plaque, being comprehensive of nonlesional (NLS) proximal to perilesional and lesional areas, with markers of chronic inflammation (i.e., CD3⁺ T cell accumulation in the dermis) predominantly present in lesional skin. Psoriatic skin also shows inflammatory CD3⁺ T cells in the epidermal compartment, which contribute to epidermal hyperplasia and an inflammatory milieu. *NLS,* Nonlesional skin.

as autoantigen by circulating CD4 and CD8 T cells with a cytokine and skin-homing receptor profile typical of psoriatic skin T cells (IFN-γ^{high}, IL-17^{high}, CLA^+, $CCR6^+$, and $CCR10^+$).[15,18]

Phospholipase A2 group IVD has been identified in psoriatic keratinocytes as an important player in the generation of psoriasis autoantigens. The latter include nonprotein neolipids that are recognized by CD1a-restricted T cells, thereby inducing the production of IL-22 and IL-17A. These lipid antigens can be transferred from keratinocytes to neighboring antigen-presenting cells through released exosomes.

Disintegrin and metalloprotease domain containing thrombospondin type 1 motif-like 5 (ADAMTSL5), a protein modulating microfibril functions and identified as autoantigen presented by melanocytes in an HLA-C*06:02-restricted fashion, has been also been localized in keratinocytes throughout the psoriatic epidermis.[4,18]

Keratinocytes can also activate pathogenic T cells by presenting viral or bacterial products. Human papillomavirus-5 DNA and reactive antibodies against virus-related particles have been found in psoriasis. Infections by *Streptococcus* commonly associate with psoriasis, and *Streptococcus*-derived superantigens can be presented to T lymphocytes by MHC class II–bearing keratinocytes. Psoriatic antigens are assumed to be keratinocyte-derived molecules sharing structural homology with streptococcal proteins, which could therefore induce autoreactive T-cell responses against skin components.[14]

Based on the analysis of infiltrating cell types, their secreted products, and genetic signatures present in lesional skin, psoriasis has been considered for many years as a type 1 (i.e., Th1)-mediated reaction, with IFN-γ playing a prominent role. However, other cytokines and T-cell subsets are pivotal in inducing inflammatory responses in psoriasis. These include Th17 and Th22 cells, which produce large amounts of IL-17 and IL-22. These two cytokines, together, mediate most of the epidermal hyperplasia by impairing keratinocyte differentiation and inducing their premature maturation and aberrant cornification.[14,19]

IL-17 activates keratinocytes to produce neutrophil and T cell–recruiting chemokines (i.e., CXCL1/CXCL2/CXCL8 and CCL20, respectively) as well as antimicrobial peptides (i.e., LL37 and S100 family members).[14,19] Thus IL-17 is a central actor in a pathogenic loop linking T cells and keratinocytes.

T cell–derived IFN-γ and TNF-α activate a plethora of inflammatory pathways in resident skin cells, in particular keratinocytes and endothelial cells.[14,19] Each cytokine regulates distinct responses with a certain degree of synergism in regulating gene expression induction/inhibition. Most of the effects induced by IFN-γ are potentiated by TNF. TNF intracellularly activates NF-κB, a transcription factor regulating gene expression frequently in collaboration with the transcription factor STAT1 induced by IFN-γ. TNF induces expression of intercellular adhesion molecule-1 (ICAM-1) on resident skin cells, permitting the adhesion and extravasation of circulating leukocytes.[14] Moreover, TNF stimulates the production of several chemokines active on immune cells, as well as proinflammatory cytokines, in particular IL-6 and IL-1, which sustain Th17 expansion.[14] Importantly, TNF, together with IL-17, induces IL-36γ in psoriasis lesions. This in turn promotes expression of antimicrobial peptides and chemokines recruiting neutrophils and Th17 cells, as well as interferes with terminal differentiation and cornification process of the epidermis.[20]

Transcriptional profiling studies conducted on lesional psoriatic skin showed that the IFN-γ signature predominates, even though IL-17 and TNF also potently induce a vast panel of genes.[14,19] A number of studies demonstrated a central role of IL-22 in psoriasis pathogenesis by activating STAT3-dependent genes involved in differentiation and proliferation processes. However, this cytokine induces a limited panel of genes compared with IL-17, as detected in human lesional psoriatic skin. Among them, the chemokines CXCL1, CXCL2, CXCL8, and CCL20 and the antimicrobial peptides HBD-2, HBD-3, and S100 proteins are induced by IL-22.[14,19]

Although T lymphocytes are required for the development and persistence of immune responses in psoriatic skin, endogenous defects in keratinocytes determining an aberrant response to lymphokines may be concomitantly relevant for psoriasis. An increasing number of allelic variants of genes controlling keratinocyte inflammatory activation, as well as proliferation and differentiation processes in the epidermis, has been associated with psoriasis.[3] Among them, a number of SNPs were found in genes encoding molecules involved in IL-17 or TNF responses, even though functional studies correlating their presence to keratinocyte susceptibility to these cytokines are lacking and controversial. For instance, allelic variants were found in *NFKBIZ*, *TRAF3IP2*, and *TNFAIP3* genes encoding IKBζ, Act-1, and the Act1-dependent A20 protein, respectively, all involved in IL-17 molecular signaling. However, no clear evidence linking these SNPs to enhanced or reduced responses of keratinocytes to IL-17 exist. In fact, Act-1 gene variants overexpressed in human keratinocytes could decrease as well as enhance Act-1–mediated IL-17 signaling, depending on the SNP type. Variants of *NFKBIZ* could also influence keratinocyte response to TNF-α because IKBζ is a transcriptional cofactor of p50 subunit of NF-κB, the main mediator of TNF-α signaling.

Finally, data from mice with engineered epidermal phenotypes also showed that overactivation in keratinocytes of molecular cascades, such as STAT3- or RAS pathways, determines the development of psoriasiform lesions that closely resemble human psoriasis. Similarly, the abrogation of JunB in keratinocytes triggers a skin phenotype with the histologic features of psoriasis, including marked hyperplasia of the epidermis and dense dermal inflammatory cell infiltrates.

Innate Immune Cells Are Fundamental for the Induction of Psoriatic Processes

Several studies have demonstrated the role of innate immunity cells in the development of inflammation in psoriasis patients (see Fig. 64.5). Among them, ILC is a class of immune cells bearing lymphoid morphology but no immune cell lineage markers.[10] ILC in situ mapping in psoriatic skin revealed that lesions contained a prominent population of T-bet$^+$ ILC1 and RORγ$^+$ ILC3 subsets, with nearly absent GATA3$^+$ ILC2. ILC1s are similar to Th1 cells and secrete type 1 cytokines, such as IFN-γ and TNF, whereas ILC3s can be divided into natural cytotoxicity receptor (NCR)$^+$ ILC3s and NCR$^-$ ILC3s17 and produce IL-17 and/or IL-22. These subsets of ILCs were found to reside beneath the dermal–epidermal junction and in close proximity to T lymphocytes, suggesting a potential functional relationship between these cells. This provides new insight into shared mechanisms between psoriasis and innate immunity.[10] Driving of ILC3 responses and IL-17 production are strongly dependent on IL-23, which can be also produced by infiltrating macrophages during the early phase of the disease. It has been proposed that the IL-23/IL-17 axis could be established early in psoriatic skin by innate immunity cells and precede adaptive immune responses.

Together with ILC, the γδ T cells, an innate-like T-cell population involved in surveillance of epithelial surfaces, are also critical contributors to plaque development by releasing considerable levels of IL-17 and IL-22. Similarly, mast cells and neutrophils can represent innate sources of IL-17 in psoriatic skin. Neutrophils also release IL-36, and in turn induce IL-36 in keratinocytes, thereby establishing neutrophil extracellular trap (NET)osis, a defense mechanism operating in psoriasis and based on the formation of cytosolic granule proteins containing autoantigens. NETosis is also fundamental for macrophage priming, Th17 activation, and immune cell recruitment in early psoriasis.[20]

Finally, macrophages can produce high levels of psoriasis-associated cytokines (*e.g.*, IL-23 and TNF). The latter are produced mainly in response to IL-36, which acts on IL-36R–bearing antiinflammatory M2 macrophages, driving them to a proinflammatory M1 phenotype.[20]

ON THE HORIZON

- Continued development of immunomodulatory biologics that target pathogenic molecules. These agents are potentially highly effective in psoriasis and include anti–tumor necrosis factor (TNF) biologics, anti-interleukin (IL)-12/23 inhibitors, anti–IL-17 and IL-17R inhibitors, and anti–IL-23p19 drugs.
- Correlation between drug response variability and the presence of specific genetic variants.
- Validation of these results in adequately powered patient cohorts.
- The discovery of new pharmacogenomic biomarkers, simultaneously present in different genes involved in intersected pathogenic pathways, which could be predictive for the responsiveness of psoriatic patients to biologic drugs.

CONCLUSIONS

Psoriasis results from a complex interplay between environmental and genetic factors. Early upstream events occurring in the disease include activation of innate immunity cells, including pDCs, ILCs, neutrophils, and macrophages, and the generation of effector T cells that migrate into the psoriatic skin lesions and expand there. Cross-talk between keratinocytes and immune cells amplifies inflammation and is responsible for chronicity. Recent research has identified a number of immunologic mechanisms involved in psoriasis progression. This has led to the development of new, pathogenesis-based therapies. Although recent progress is remarkable, much remains unknown, especially regarding how to prevent the condition and how to develop drugs with appropriate risk–benefit and long-term profiles. Future work must take into account these aspects to establish therapeutic and preventive approaches that lead to improved patient outcomes.

REFERENCES

1. Reid C, Griffiths CEM. Psoriasis and treatment: past, present and future aspects. *Acta Derm Venereol*. 2020;100(3) adv00032.
2. Dand N, Mahil SK, Capon F, et al. Psoriasis and genetics. *Acta Derm Venereol*. 2020;100(3) adv00030.
3. Capon F. The genetic basis of psoriasis. *Int J Mol Sci*. 2017;18(12):2526.
4. Prinz JC. Human leukocyte antigen-class I alleles and the autoreactive T cell response in psoriasis pathogenesis. *Front Immunol*. 2018;9:954.
5. Di Meglio P, Villanova F, Nestle FO. Psoriasis. *Cold Spring Harb Perspect Med*. 2014;4(8):a015354.
6. Rendon A, Schakel K. Psoriasis pathogenesis and treatment. *Int J Mol Sci*. 2019;20(6):1475.
7. Zhang LJ. Type1 interferons potential initiating factors linking skin wounds with psoriasis pathogenesis. *Front Immunol*. 2019;10:1440.
8. Girolomoni G, Strohal R. Puig L, et al. The role of IL-23 and the IL-23/TH 17 immune axis in the pathogenesis and treatment of psoriasis. *J Eur Acad Dermatol Venereol*. 2017;31(10):1616–1626.
9. Schon MP, Erpenbeck L. The interleukin-23/interleukin-17 axis links adaptive and innate immunity in psoriasis. *Front Immunol*. 2018;9:1323.
10. Zhou S, Li Q, Wu H, Lu Q. The pathogenic role of innate lymphoid cells in autoimmune-related and inflammatory skin diseases. *Cell Mol Immunol*. 2020;17(4):335–346.
11. Swiecki M, Colonna M. The multifaceted biology of plasmacytoid dendritic cells. *Nat Rev Immunol*. 2015;15(8):471485.
12. Albanesi C, Scarponi C, Bosisio D, et al. Immune functions and recruitment of plasmacytoid dendritic cells in psoriasis. *Autoimmunity*. 2010;43(3):215–219.
13. Nestle FO, Di Meglio P, Qin JZ, Nickoloff BJ. Skin immune sentinels in health and disease. *Nat Rev Immunol*. 2009;9(10):679–691.
14. Albanesi C, Madonna S, Gisondi P, Girolomoni G. The interplay between keratinocytes and immune cells in the pathogenesis of psoriasis. *Front Immunol*. 2018;9:1549.
15. Lande R, Botti E, Jandus C, et al. The antimicrobial peptide LL37 is a T-cell autoantigen in psoriasis. *Nat Commun*. 2014;5:5621.
16. Wang A, Bai Y. Dendritic cells: the driver of psoriasis. *J Dermatol*. 2020;47(2):104–113.
17. Ahmad F, Dobel T, Schmitz M, Schakel K. Current concepts on 6-sulfo LacNAc expressing monocytes (slanMo). *Front Immunol*. 2019;10:948.
18. Fuentes-Duculan J, Bonifacio KM, Hawkes JE, et al. Autoantigens ADAMTSL5 and LL37 are significantly upregulated in active psoriasis and localized with keratinocytes, dendritic cells and other leukocytes. *Exp Dermatol*. 2017;26(11):1075–1082.
19. Chiricozzi A, Romanelli P, Volpe E, et al. Scanning the immunopathogenesis of psoriasis. *Int J Mol Sci*. 2018;19(1):179.
20. Madonna S, Girolomoni G, Dinarello CA, Albanesi C. The significance of IL-36 hyperactivation and IL-36R targeting in psoriasis. *Int J Mol Sci*. 2019;20(13):3318.

65

Myasthenia Gravis

Patricia M. Sikorski, Linda L. Kusner, and Henry J. Kaminski

Myasthenia gravis (MG) is a rare immune disease with a prevalence of 200 to 400 cases per million.[1] Patients have characteristic weakness that is worsened with activity. Muscle weakness is mediated by pathogenic autoantibodies that impair neuromuscular transmission by targeting functionally important proteins of the neuromuscular junction (NMJ), primarily the nicotinic acetylcholine receptor (AChR). However, non-AChR components of the NMJ, such as muscle-specific kinase (MuSK) and low-density lipoprotein receptor-related protein 4 (Lrp4), are also targeted in a minority of patients negative for AChR antibodies. Although the use of experimental animal models of MG has led to well-defined roles of autoantibodies in contributing to immunopathology of the disease, the cellular basis for the development and maintenance of pathogenic antibodies is multifactorial and less well understood. This chapter focuses on the structure and function of myasthenic autoantigens at the NMJ and how pathogenic mechanisms of autoantibodies contribute to the spectrum of disease phenotypes in MG patients.

DISEASE DIAGNOSIS AND CLASSIFICATION

The fundamental manifestation of MG is fluctuating muscle weakness that worsens with repeated muscle use and improves with rest. Heterogeneity in the pattern and severity of weakness exists among patients, but general categories can be used to define patients by clinical characteristics. First, patients may have disease limited to the ocular muscles leading to ptosis or double vision, so-called ocular myasthenia, while the remainder have generalized weakness not limited to the eye muscles. Initial presentation of the disease for about two-thirds of patients involves the ocular muscles, which progresses to generalized MG in a majority of patients.[2,3] The MG Foundation of America Clinical Classification separates patients across the degree of severity from ocular (class 1) to mild, moderate, and severe (classes 2 to 4) and an MG crisis (class 5), which stipulates patients that require respiratory support. The classification further identifies class 2 to 4 patients having predominant weakness of cranial nerve innervated muscle with limited general weakness (subclass b) or the reverse (subclass a). Patients rarely have remarkably isolated weakness (e.g., weakness limited to the neck extensor muscles producing "head drop" or foot drop from ankle dorsiflexor muscles or breathing muscle weakness in absence of more general weakness).

The clinical diagnosis can be confirmed using serological testing. In about 85% to 90% of patients, anti-AChR antibodies are detected in serum, whereas a minority (5% to 10%) of the patients produce antibodies targeting other proteins in the muscle or NMJ, such as MuSK or Lrp4.[4] About half of ocular MG patients produce AChR antibodies with the remainder being seronegative. MuSK antibody patients rarely exhibit purely ocular MG. In patients without serum autoantibodies, electrophysiological testing can confirm diagnosis. In upwards of 80% of patients, a decremental pattern is seen following repetitive nerve stimulations (Fig. 65.1).

Single fiber electromyography evaluates variations in timing of junctional activation and may be detected in expert hands in over 90% of patients.[2] The classic diagnostic evaluation (edrophonium test) involves the administration of cholinesterase inhibitor with an examiner monitoring for unequivocal improvement in muscle strength. In the United States the edrophonium test is not commonly performed because of drug availability and safety concerns; however, the evaluation can be an extremely useful adjunct in diagnosis of seronegative patients.

As the pathophysiological heterogeneity of MG has been refined, categorization has been further refined. AChR antibody-positive patients are subdivided based on disease onset (early onset <45 years [EOMG] or late onset >45 years [LOMG]). The EOMG subtype patients are more likely to have thymic hyperplasia with high levels of AChR antibodies. In contrast, LOMG patients are considered to more often have thymic atrophy, although a recent study of specimens from a surgical study did not support this contention.[5] Thymoma-associated MG patients essentially always have AChR antibodies and tend to have a worse clinical course. As discussed, patients are also categorized by autoantibody status.

CLINICAL PEARLS

- Essentially all patients will have ptosis or double vision at some time during their illness
- Pupillary response, sensation, and reflexes are always normal in myasthenia gravis.
- Significant variation in weakness occurs over a day and over weeks to months

NEUROMUSCULAR TRANSMISSION

The NMJ is a chemical synapse formed between the terminal end of the motor neuron and the muscle fiber that is critical for transmitting signals to the fiber for muscle contraction. Neuromuscular transmission involves release of acetylcholine (ACh) from the nerve ending into the junctional cleft, which then binds to the nicotinic AChR highly concentrated on the postjunctional membrane of the muscle, known as the motor endplate. The binding of ACh to AChR results in opening of ligand-gated ion channels with resultant influx of primarily sodium ions, leading to endplate depolarization that stimulates an action potential generation to initiate muscle contraction. The endplate potential (EPP) is terminated by diffusion of ACh and its rapid hydrolysis by acetylcholinesterase (AChE) anchored to the basal lamina of the synaptic cleft.

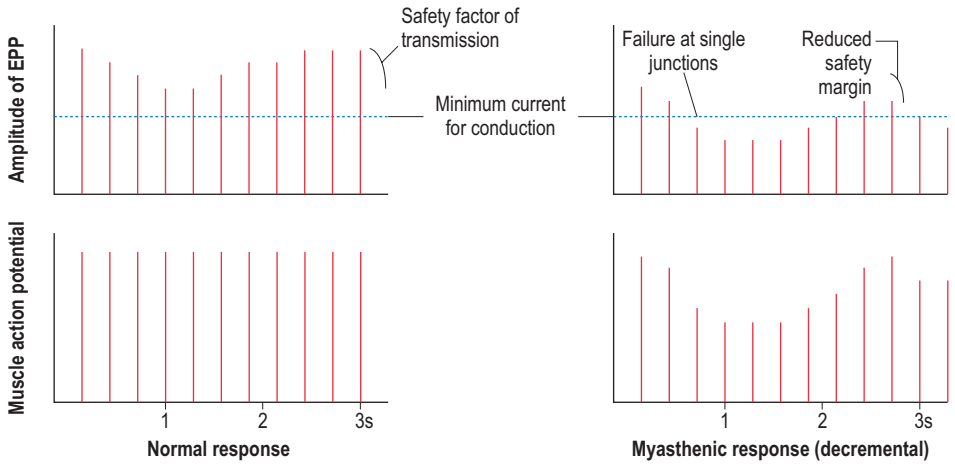


FIG. 65.1 Neuromuscular Transmission in Normal and Myasthenic Subjects. With repetitive stimulation there is a reduction in the efficiency of acetylcholine (ACh) release, with a subsequent recovery in efficiency as the train of stimuli continues. Although the endplate potential *(EPP)* fluctuates at the normal junction, sufficient current is generated to stimulate an action potential of constant magnitude. At the myasthenic junction, however, the amplitude of the EPP in response to a given amount of ACh is reduced. Under conditions of inefficient ACh release (for example, repetitive stimulation), the minimum current for conduction is not generated, resulting in a profile of action potentials that shows a progressive decline or "decrement" with subsequent recovery.

The spontaneous release of ACh from a synaptic vesicle with subsequent binding to AChRs leads to measurable depolarization, known as a miniature endplate potential (MEPP). As a result of motor nerve activation multiple vesicles release ACh to produce summation of MEPPs and an EPP that exceeds the threshold necessary to produce an action potential of the muscle. The ratio between the EPP and threshold required to generate an action potential is known as the safety factor. Postsynaptic folding, the density of AChR, and endplate sodium channels all determine the magnitude of the safety factor. If the EPP fails to reach the threshold for depolarization, an action potential is not generated and the muscle does not contract.

In 1964, electrophysiological studies of a muscle biopsy taken from an MG patient demonstrated that the amplitude of MEPPs was diminished.[6] Studies of the motor endplate ultrastructure in MG muscle determined several postsynaptic abnormalities, including simplification of postsynaptic regions, accumulation of debris in the synaptic space, and a reduction in AChR. The deficiency of AChR decreases the amplitude of MEPP and subsequent EPP, reducing the safety factor for neuromuscular transmission. Any additional stressor, such as repetitive activity, will lead to a greater likelihood of transmission failure among those endplates that have a compromised safety factor.

Differential Muscle Involvement in Myasthenia Gravis

The pattern of muscle weakness markedly varies across patients, but the reasons are poorly understood. Differences in the physiological properties of the NMJ across muscles and their response to autoantibody-mediated injury are likely to play a part in muscle susceptibility, as well as the specific autoantibody profile of an individual MG patient. Extraocular muscles (EOM) are particularly susceptible to MG and several differences between skeletal muscle and EOM may account for their increased susceptibility to injury. Simple histological assessment and gene expression profiling differentiate EOM. The NMJ of EOM are anatomically distinct in that they contain both singly and multiply innervated muscle fibers. These NMJs morphologically exhibit less prominent synaptic folds, and few postsynaptic AChRs, leading to a lower safety factor.[7] In EAMG models, changes in RNA expression profiles related to muscle injury and decreases in mRNA level for several cell-surface complement regulators was observed in EOM.[8] Collectively, these observations suggest that the selective involvement of EOM in MG may be due to subtle differences in properties of NMJs across muscles.

KEY CONCEPTS

Involvement of Anti-Acetylcholine Receptor Antibodies in the Pathogenesis of Myasthenia Gravis (MG)

- Anti-acetylcholine receptor antibodies are found in the serum of 85% to 90% of patients with MG
- IgG and complement are deposited at the postsynaptic junction
- Transfer of serum IgG from patients with MG to mice induces neuromuscular blockade

AUTOANTIGENS IN MYASTHENIA GRAVIS AND THEIR ROLES IN NEUROMUSCULAR TRANSMISSION

Neuromuscular transmission is dependent on highly specialized transduction machinery present at the postsynaptic membrane. The specialized postsynaptic ultrastructure at the motor endplate is highly organized to efficiently and rapidly respond to neurotransmitters released from the nerve terminal, which requires properly clustered AChRs and a complex pattern of synaptic folds on the postsynaptic membrane. The aggregation of AChRs is an intricate process that relies on the agrin-Lrp4-MuSK signaling cascade.

Agrin, a heparan sulfate proteoglycan, is released by motor nerve terminal and binds Lrp4, which then bind MuSK to stimulate homodimerization and transphosphorylation of MuSK. These events stimulate intracellular signaling events that include adaptor proteins Dok7, Tid1, and rapsyn, and result in AChR clustering and stabilization. The autoantibodies against extracellular or transmembrane proteins are pathogenic in MG either directly or indirectly by affecting AChR function and quantity, ultimately resulting in structural alteration or tissue injury (Fig. 65.2).

Acetylcholine Receptor

The nicotinic AChRs are located on top of the postsynaptic folds at the motor endplate at a density of 10,000/μm^2. AChRs are membrane proteins belonging to a superfamily of ligand-gated ion channels consisting of five transmembrane subunits: two α_1 subunits, one β_1 subunit, one δ subunit, and either one γ subunit (embryonic) or ε subunit (adult) to form an ion channel (Fig. 65.3)

Ion channels are gated pores that allow passive flow of ions down electrochemical gradients to induce a cellular response. Upon ligand binding (ACh) to the α subunit, the receptor structure is converted into a channel opening within microseconds to allow the flow of positive ions, leading to endplate depolarization and initiation of action potential. The concentration and arrangement of AChRs at the postsynaptic membrane is critical to ensure successful synaptic transmission. Maintaining the concentration is dependent on signals from the motor neurons to accumulate and stably maintain AChRs and other postsynaptic proteins at restricted synaptic sites.

Muscle-Specific Kinase

The release of agrin, a sulfated proteoglycan secreted by motor neurons, induces the assembly of AChR clusters by interacting with Lrp4 and activation of MuSK. MuSK is selectively expressed in skeletal muscles, co-localizes with AChRs in the postsynaptic membrane, and plays an important role in regulating the high density and clustering of AChRs at the NMJ. Genetic studies have shown that reducing expression of MuSK in mice destabilizes and disassembles postsynaptic membranes, signifying it is a key component in maintaining synaptic function.[9]

MuSK is a member of the RTK class of protein kinases, which consist of an extracellular ligand-binding domain, a single transmembrane spanning region, and a cytoplasmic region that includes a tyrosine kinase domain.[10] MuSK, however, is not a typical tyrosine kinase that relies on ligands that bind directly to the receptor. Rather, it is activated by ligands that do not directly bind it and instead bind a separate accessory component. The extracellular region, or ectodomain, of MuSK contains three immunoglobulin (Ig)-like domains that mediate ligand binding critical for MuSK activation. Biochemical and structural studies determined that the first and second Ig-like domains are important for binding agrin and its co-receptor, Lrp4.[11] MuSK activation triggers homodimerization and autophosphorylation that results in a signaling cascade, resulting in the redistribution of muscle-derived proteins including AChRs, rapsyn, and MuSK at the postsynaptic site and become stably localized in clusters beneath the nerve terminal.

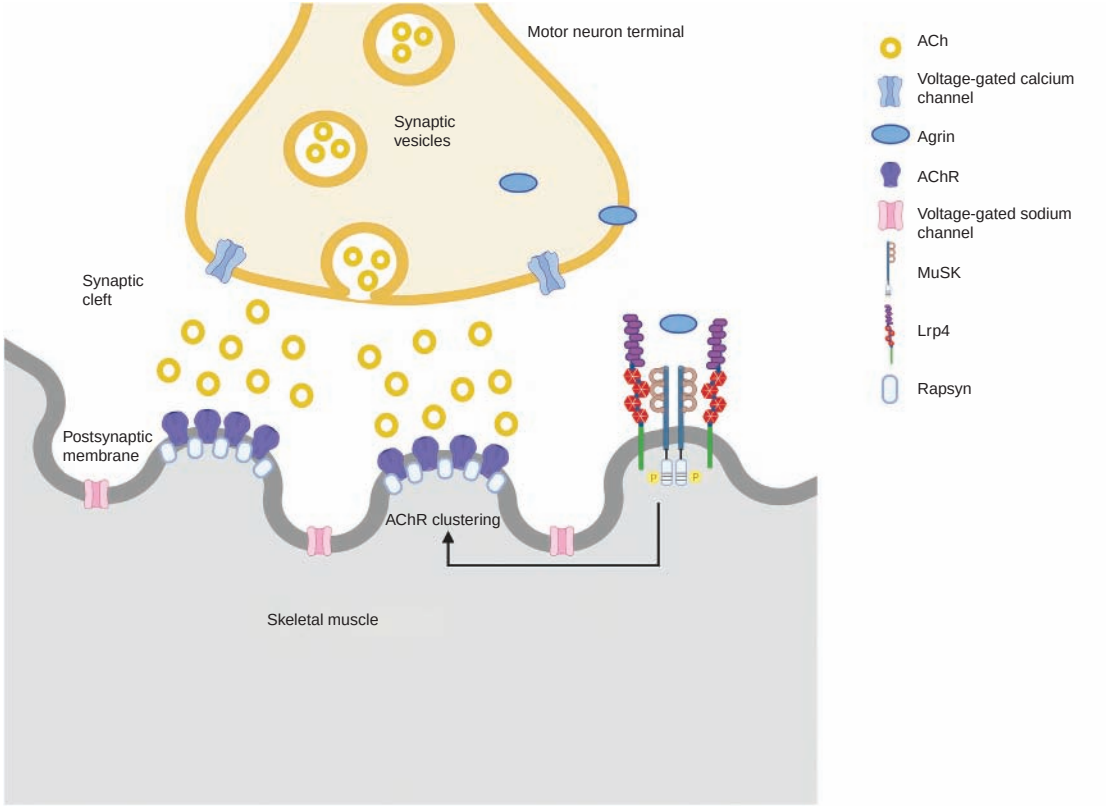


FIG. 65.2 Schematic Representation of the Neuromuscular Junction. Vesicles of acetylcholine *(ACh)* released from the motor neuron into the synaptic cleft in response to impulses conducted down nerve axons. ACh binds AChRs, with the opening of the ion channel and generation of endplate potential. Postsynaptic organization in the skeletal muscle includes AChR clusters at the top of the postsynaptic folds and voltage-gated sodium channels in the troughs. The agrin-Lrp4-MuSK complex is critical for proper AChR clustering. Agrin released from the motor neuron binds MuSK-coreceptor Lrp4 and induces MuSK autophosphorylation. Activation of MuSK stimulates signaling pathways that promote rapsyn-induced clustering of AChR. (Figure generated using BioRender). *AChR*, Acetylcholine receptor; *Lrp4*, low-density lipoprotein receptor-related protein 4; *MuSK*, muscle-specific kinase.

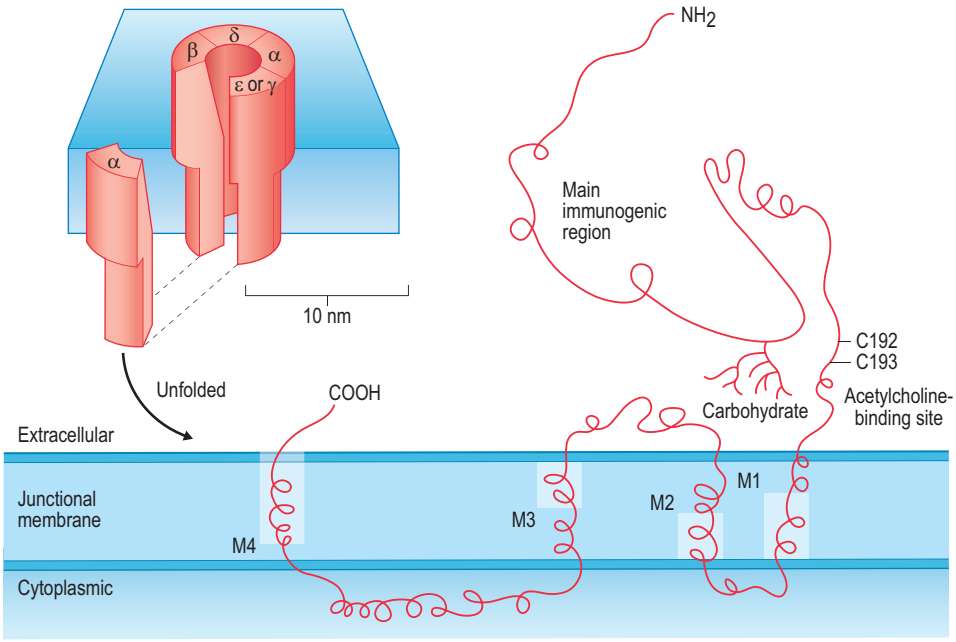


FIG. 65.3 The Acetylcholine Receptor. The subunits of the acetylcholine receptor—α, β, δ, and γ or ε—are arranged like barrel staves around the central ion pore. Each subunit winds through the junctional membrane four times (sites M1, M2, M3, and M4). In the unfolded view of the α subunit, the amino-terminal end of the α subunit is extracellular, where it is accessible to acetylcholine, which binds at the site shown (amino acids 192 and 193). In myasthenia gravis, autoantibodies may bind to various epitopes of all subunits, but a high proportion of autoantibodies bind to the main immunogenic region of the α subunit.

Low-density Lipoprotein Receptor-related Protein 4

Lrp4 is a receptor for neural agrin and subsequently forms a complex with MuSK. Members of the low-density lipoprotein (LDL) receptor-related protein (LRP) family have diverse biological functions, including specific roles in the development and function of the mammalian nervous system. Lrp4 is a single subunit transmembrane protein with a large extracellular domain that contains low-density lipoprotein repeats and is concentrated at the NMJ. Like MuSK- and agrin-deficient mice, genetic screens revealed that mice with mutations in *lrp4* died at birth and did not form NMJ.[12] Thus, Lrp4 plays a role in maintaining and stabilizing the NMJ structure and functions by increasing MuSK activity.

EXPERIMENTAL MODELS TO STUDY MYASTHENIA GRAVIS

Experimental autoimmune MG (EAMG) is induced by immunization of MG autoantigen (AChR, MuSK, or Lrp4). Another model is referred to as passive transfer MG (PTMG) and is the result of injection of pathogenic autoantibodies. The availability of transgenic and knockout mice is an important instrument in identifying the molecular mechanisms underlying the disease. However, several important differences between human and rodent immune responses exist that may impact the applicability to translational work, such as differences in antibody isotype production in response to immunization and species-specific differences in complement activity. Furthermore, there are several aspects of MG pathogenesis that cannot be studied using these models, including the absence of spontaneous disease and lack of thymic involvement and pathology. EAMG has been a critical tool in elucidating the pathogenic mechanisms of autoantibodies in impairing neuromuscular transmission, but also in investigating novel therapeutic approaches.

MECHANISMS OF IMMUNOPATHOGENESIS

MG fulfills the Witebsky postulates that define the criteria of autoimmune disease, which require (1) an autoimmune reaction in the form of autoantibodies or cell-mediated immune reactions, (2) that the autoantigen is known, and (3) an autologous response causes similar disease in an animal model.[13] The immunopathology of MG has been attributed to IgG class autoantibodies specific to the proteins at the postsynaptic membrane of the NMJ and cause organ-specific damage. This is further supported by the short-term improvement of NMJ transmission and muscular strength for patients undergoing plasmapheresis to remove autoantibodies. Anti-AChR antibodies (IgG1 and IgG3 subtypes) are detected in 85% to 90% of MG patients, whereas 5% to 10% of patients are anti-AChR negative and have antibodies against MuSK (IgG4 subtype) and Lrp4 (IgG1 and IgG2 subtypes). Passive transfer of human antibodies to rodents recapitulates the disease in animals, demonstrating that these autoantibodies are responsible for impaired neuromuscular transmission. Furthermore, these autoantibody deposits have been detected at the NMJ in patients and in experimental animal models. These autoantibodies, however, exhibit both structural and functional heterogeneity, which likely contributes to the broad and variable spectrum of muscle group involvement in MG patients.

Role of Pathogenic B cells

B cells play a critical role in the immunopathogenesis of MG as they secrete autoantibodies that target the NMJ, making them an attractive target for therapeutic intervention. The depletion of plasma cells, or antibody-producing cells, with bortezomib reduced anti-AChR titers and weakness in rats.[14] Treatment of MG patients with

rituximab, a treatment targeting CD20+ B cells, has been shown to be effective in refractory MuSK+ MG patients.[15] However, the study of these pathogenic B cells is limited by the exceptionally low frequency of 0.1% to 1% of these cells in peripheral blood. Recent advances in molecular biology are advancing ways to better isolate and characterize these B cells. Novel technologies such as single cell RNA sequencing provide an exciting way forward to better characterize these rare subpopulations of autoreactive B cells and gain further insights in understanding the cellular basis for disease that may lead to novel therapies to target these cells.

The immunological mechanisms contributing to the development and maintenance of autoreactive B cells in MG that produce these autoantibodies are not well understood. Defects in B-cell tolerance checkpoints were found in both AChR and MuSK MG patients.[16] Apoptosis is an important mechanism in central B-cell tolerance that eliminates self-reacting immature B cells from the repertoire of naïve immature B cells in the bone marrow. However, a fraction of B cells is able to escape this stringent selection process. The selection and maturation of these autoreactive B cells in the periphery relies on the strength of the B-cell receptor (BCR) signals it receives as well as survival signals such as B-cell-activating-factor (BAFF). Serum levels of BAFF are elevated in MG patients, but do not correlate with disease severity. Autoreactive B cells from MG patients and EAMG models express an anti-apoptotic marker known as survivin.[17] Furthermore, survivin vaccination in EAMG attenuated disease severity and decreased survivin-positive CD20+ cells, suggesting it plays a role in maintenance of autoreactive B cells. The factors that regulate and promote survivin expression in several autoimmune diseases, including MG, have yet to be elucidated.

Pathogenic Autoantibodies

The differences in clinical phenotypes associated with the subtypes of MG can be attributed to their differences in serum autoantibody targets. Here, we discuss the well-characterized molecular pathogenic mechanisms of autoantibodies in MG and their contribution to disease outcomes.

KEY CONCEPTS

Effects of anti-AChR antibodies in MG pathology:
- Reduced numbers of receptors
- Widening of synaptic cleft
- Injury to synaptic membrane

Mechanisms of damage and reduced neuromuscular transmission:
- Complement-dependent damage to muscle endplate
- Enhanced rate of AChR degradation
- Block of cholinergic binding sites

Acetylcholine Receptor Autoantibodies

The pathogenicity of AChR antibodies was first reported in the 1970s. Passive transfer of either human AChR antibodies or immunization of animals with AChR has been shown to induce EAMG in rodents.[18] These antibodies are polyclonal and heterogeneous, primarily of the IgG1 and IgG3 subtypes, and are directed against the α_1- extracellular domains containing the immunogenic region of overlapping epitopes. Together, these antibodies decrease levels of AChR and reduce postsynaptic folding. However, disease severity does not correlate with serum levels of AChR antibodies and patients with AChR-antibody positive MG exhibit high variability in the distribution of muscle weakness. These observations indicate that additional factors could contribute to disease severity, such as genetic factors, capacity of autoantibodies to activate the complement system, and qualitative differences in antibodies, including variations in IgG isotype distribution, epitope specificities, and binding avidity to AChR.

There are three known effector mechanisms of AChR antibodies that reduce the density of AChR and affect the structural integrity of the postsynaptic region, resulting in impaired neuromuscular transmission (discussed below).

Complement Activation and Destruction of Neuromuscular Junction

The classical pathway (CP), an antibody-dependent activation of the complement system leading to direct lysis of targeted cells, is considered the major contributor to destruction of the postsynaptic membrane in MG. AChR antibodies that bind the NMJ activate the CP and result in membrane attack complex (MAC) formation. The presence of immune complexes and C3 at the motor endplate, the localization of the MAC at the junctional folds, and evidence of injury of the postsynaptic membrane in MG patients demonstrates a role for complement contributing to loss of AChR and NMJ ultrastructure. EAMG-induced mice deficient in complement proteins C3 and C4 were resistant to EAMG. The presence of immune complexes at the NMJ in these EAMG animals did not result in the MAC and reduced AChR. Furthermore, administration of C5 complement inhibitor, an siRNA to silence C5 expression, or anti-C5 antibody was shown to limit the severity of EAMG by ameliorating weakness and reducing C9 deposition at the NMJ.[19] Recently, treatment of MG patients with eculizumab, an antibody that prevents the enzymatic cleavage of C5, has been demonstrated to reduce severity of generalized AChR-antibody-positive MG.[19]

Experimental targeting of CP-specific components such as C1q and C4 have shown to be effective in alleviating EAMG. However, the alternative pathway (AP) is estimated to contribute up to approximately 80% of C3b activation, which rapidly amplifies complement activation and deposition by utilizing CP C3b to generate AP C3 convertases. Recently, an antibody blocking AP convertase formation by targeting Factor B protected rats from EAMG and MuSK positive MG patients were shown to exhibit higher consumption of AP Factor B, but not CP component C4,[20,21] suggesting a possible role for the AP in MG. Both soluble and membrane-bound regulators play an important role in regulating the AP amplification loop on cell surfaces. Studies have shown that membrane bound decay accelerating factor (DAF) is critical for protecting from exacerbated disease in EAMG models.[19] Furthermore, a single nucleotide polymorphism (SNP) in the human *DAF* gene has been associated with severe ocular MG.[22] Thus, these findings indicate that targeting multiple points of complement system activation is an important clinical target for MG treatment.

Acceleration of Acetylcholine Receptor Degradation (Antigenic Modulation)

The main antigenic region (MIR) of AChR consists of conformational-dependent epitopes targeted by a majority of AChR antibodies. AChR antibody binding to MIR is oriented in a way that promotes cross-linking of two AChRs by a single antibody. Early studies indicated that AChR antibodies increased the rate of AChR degradation. The bivalent nature of IgG autoantibodies

promotes cross-linking of AChRs in the membrane, resulting in rapid endocytosis and degradation and further contributing to a reduced number of AChR at the postsynaptic membrane.[23]

Receptor Blockade

AChR antibodies in MG sera are polyclonal and exhibit heterogeneity in antigenic specificity. Not all AChR antibodies have the capacity to cross-link two AChRs or activate complement, and thus may exhibit different functionalities. AChR antibodies were additionally shown to block the alpha-bungarotoxin binding site of the receptor[24] and thus are characterized by their ability to block the ACh-binding sites and interfere with ACh released at the NMJ to impair neuromuscular transmission.

Muscle-Specific Kinase Autoantibodies

About 5% to 10% of MG patients have autoantibodies against MuSK. Either immunization with MuSK or passive transfer of human MuSK antibodies have also been shown to induce pathology in mice. Patients with these autoantibodies have more prominent involvement in facial and bulbar weakness and exhibit more respiratory crises, whereas limb weakness is not common. These antibodies are typically subtype IgG4, and unlike AChR antibodies, do not activate complement or bind Fc receptors. Furthermore, the IgG4 antibodies have structural differences in the Fc region and can exchange Fab arms with other IgG4 molecules, which renders them bispecific (i.e., recognizing two different antigen) and functionally monovalent; hence they do not cross-link antigen.

The pathogenic mechanisms of MuSK antibodies that lead to failure of neuromuscular transmission are distinct from those observed in AChR MG. Passive transfer of human MuSK antibodies to mice revealed structural changes at the postsynaptic membrane, such as diminished AChR and MuSK densities, suggesting that the mechanisms in MuSK MG may involve changes in the function and distribution of key molecules at the NMJ.[25,26] Epitope mapping and functional studies of MuSK antibodies demonstrated they predominately recognize the N-terminal Ig-like domain of MuSK, inhibiting the MuSK-Lrp4 interaction, preventing MuSK dimerization and autophosphorylation, impairing the clustering of the AChR, ultimately resulting in impaired neuromuscular transmission.[27] However, MuSK antibodies are also heterogeneous, containing other IgG subclasses and recognizing other domains on MuSK. The pathogenic mechanisms of these antibodies are not yet well understood. Importantly, the distinct immunopathogenesis of MuSK antibodies indicates that therapies targeting complement would not be suitable treatment for this MG patient subtype.

Lrp4 Autoantibodies

Some patients without AChR and MuSK antibodies have been identified to possess autoantibodies that bind Lrp4. Mice immunized with Lrp4 or passive transfer of Lrp4 antibodies to naïve mice promoted weakness and NMJ defects similar to those of MuSK-EAMG.[28,29] Furthermore, Lrp4 antibodies from mice immunized with Lrp4 were shown to diminish agrin-induced MuSK phosphorylation of C2C12 myotubes and fix complement in vitro, demonstrating potential pathological mechanisms of these antibodies.

Role of Thymus Pathology

The thymus is a primary lymphoid organ that provides a specialized environment for T-cell development, maturation, and negative selection of self-reactive T cells. Most AChR antibody-positive MG patients have thymic pathology, indicating a strong likelihood that the thymic abnormalities are critical for development of MG. Follicular hyperplasia is most commonly observed in 65% to 75% of early-onset patients, whereas 10% of patients have a thymoma. Thymectomy in these patients is associated with clinical improvement and a decrease in AChR antibodies. However, the exact mechanisms by which these thymic pathologies promote autoreactivity in MG is not clear.

Thymuses with follicular hyperplasia in MG have increased numbers of lymphoid germinal centers that is not typical. In AChR antibody positive MG, the degree of follicular hyperplasia often correlates with serum levels of AChR antibodies, suggesting thymic hyperplasia is associated with autoantibody production. Several pathological alternations of the hyperplastic thymuses in MG patients cumulatively provide a rich microenvironment for immune sensitization to AChR. First, both medullary thymic epithelial cells and myoid cells have been shown to express AChR or AChR-like proteins.[30] Second, both thymic effector and regulatory T cells in these MG patients exhibit immunoregulatory defects. Lastly, the presence of increased germinal centers in the thymus is associated with a heterogeneous population of B cells that have undergone clonal expansion and somatic hypermutation.[31] MG patients with thymic hyperplasia have high levels of CXCL13, a chemokine involved in homing B cells to lymphoid tissue, in serum and in the thymus.[32] The expression of BAFF and APRIL, two critical B-cell survival factors, is also abundant in hyperplastic thymuses,[33] which could play a role in differentiation and survival of autoreactive B cells.

About 10% of MG patients have a thymoma. These thymic tumors in MG patients are derived predominately from cortical epithelial cells and exhibit reduced medullary areas. The lack of a functional medulla in thymoma patients may play a role of promoting dysregulated tolerance mechanisms, which includes altered expression of autoimmune regulator AIRE and regulatory T cells (Tregs) and defective expression of human leukocyte antigen (HLA) class II molecules. Furthermore, the epithelial cells are surrounded by mature T cells and express muscle protein epitopes such as AChR, allowing for intrathymic autosensitization to self-antigen. The presence of follicular helper T cells, which play a role in germinal center formation, B-cell maturation, and antibody production, may also play a role in these processes. Together, these studies strongly suggest that thymic pathology in MG patients is associated with a microenvironment associated with dysregulated self-tolerance mechanisms.

KEY CONCEPTS

Pathogenic Roles of Thymus in Myasthenia Gravis

Pathological
- 65% to 75% of MG patients have follicular hyperplasia with germinal centers
- 10% have thymoma

Clinical
- Improvement after thymectomy
- Decrease in AChR antibodies

Immunological
- AChR subunits expressed on myoid cells and thymic epithelial cells
- AChR-reactive T and B cells localized in the thymus
- AChR antibodies secreted by thymic B-cell lineage
- Decreased thymic regulatory T cell (Treg) function
- Overexpression of proinflammatory cytokines and chemokines

Role of T cells

MG is a B-cell-mediated T-cell-dependent disease. B cells require interactions with $CD4^+$ helper T cells to generate high-affinity AChR antibodies. Autoreactive T cells specific for AChR have been detected in both generalized and ocular MG patients. It is unclear how these T cells break tolerance and regulate autoantibody production, but evidence suggests that different subpopulations of CD4 T helper cells and their effector cytokines may be involved in the pathogenesis of MG. Tregs play a role in regulating helper T cells by maintaining immune homeostasis and regulating autoreactive cells. Both circulating and thymic Tregs from MG patients are unable to suppress responder T cells and were associated with reduced levels of Foxp3.[34] Th17 cells have also been implicated, as their frequency and their cytokines are both increased in MG patients with thymoma. Recently, a novel T-cell library assay to screen for rare antigen-reactive T cells demonstrated that $CCR6^+$ memory T cells from MG patients, but not healthy controls, produced INF-γ and IL-17 in response to AChR.[35] The subpopulations of helper T cells prominent in MuSK MG is not yet defined; however, cytokine analyses of T cells from these patients demonstrate that CD4 and CD8 T cells also have increased Th1 and Th17 cytokine responses upon in vitro stimulation.[36]

EAMG models have also been used to investigate and validate the role of T cells and their effector cytokines in the pathogenesis of MG. Studies investigating the role of Th1 effector IFN-γ in EAMG are contradictory in their phenotypes. However, IL-12/IL-23 and IFN-γ double knock-out mice were shown to develop EAMG, indicating Th1 cells may not be necessary to induce an autoimmune condition.[37] These DKO mice exhibited a similar level of IL-17 production and had reduced Treg suppressive activity compared to wild-type mice, suggesting these Th subsets may play an important role. Another study demonstrated that the balance of helper T-cell subsets is dysregulated during the development of disease, with a particular upregulation of Th17 cells.[38] However, based on these studies, how these effector cells contribute toward the development of EAMG is not clear.

Recently, evidence of locally produced complement by activated lymphocytes has demonstrated a novel role for the complement system in regulating the function of T cells. Autocrine signaling through anaphylatoxin receptors C3aR and C5aR promotes Th1 and Th17 effector responses, and absence of this signaling results in polarization of $Foxp3^+$ Tregs in the murine model (reviewed in Le Friec et al.[39]). Deficiency in DAF, a membrane-bound regulator of complement activation, also results in enhanced T-cell responses and leads to exacerbation of several experimental autoimmune diseases, indicating a suppressive role in T-cell function. Interestingly, early-onset MG patients with thymic pathologies demonstrated increased complement activation in the thymus, with increased levels of C1q and C3, increased expression of anaphylatoxin receptors (C3aR, C5aR), and decreased expression of complement regulatory proteins CD46 and CD55.[40] However, the role of complement regulation of T cells has not yet been investigated in the context of MG. Given the evidence of complement activation in the thymus and its known role in regulating effector T-cell responses, this novel line of inquiry could provide additional insight into possible mechanisms of dysregulated T-cell responses in MG.

Epigenetic Factors: Role of miRNAs

Genetic studies and genome-wide association studies have identified several important genetic loci associated with disease and development of MG, including several HLA alleles, *TNIP1*, and *TNF-α* (reviewed in Avidan et al.[41]). Environmental factors, such as infection, microbiota, drugs, and pollutants can also cause epigenetic changes that may contribute to MG. The initiation of MG is thought to be multifactorial, involving genetic and environmental influences.

Emerging evidence of aberrant microRNA expression has been associated with several diseases, including autoimmunity. These regulatory small non-coding RNAs regulate posttranscriptional gene expression by binding to complementary sequences in the 3′ untranslated regions (UTRs) of target mRNAs. Several miRNAs that are known to regulate immunity through proliferation, apoptosis, and differentiation of immune cells have recently been found to be differentially expressed in MG patients.[42,43] The thymus pathology in MG, as described above, involves changes in the miRNA express for $AChR^+$ EOMG, and TAMG. Regulation of these miRNAs is associated with germinal center formation, such as miR-7, miR-24, miR-139, miR-143, miR-145, miR-146, miR-150, miR-452, miR-548, or thymic inflammation, such as miR-125b, miR-146, or miR-29.[44] The changes in the miRNAs due to MG do not reflect the miRNAs necessary for development of the thymus, suggesting the mechanism of dysregulation does not involve upregulation of fetal pathways. Circulating miRNA demonstrate differences based on MG classification. Generalized $AChR^+$ EOMG and LOMG express elevated miR-150-5p and miR-21-5p and is reduced with immunosuppression. However, miR-30e-5p is also elevated in generalized LOMG $AChR^+$ patients' sera. With circulating miRNAs, $MuSK^+$ MG patients demonstrate higher levels of the let-7 family. The expression of elevated miRNA-150-5p in circulation is lowered with thymectomy. Thus, circulating miRNAs may serve as biomarkers for disease progression or therapeutic response. miRNAs in MG may also function to regulate genes involved in immune B-cell survival and cytokine production, which can contribute to MG pathogenesis and thus can also serve as therapeutic targets. Due to the multiple binding sites and regulation of miRNA on various pathway, the role of specific miRNAs in MG remains unknown and requires further investigation.

TREATMENT

Treatment of MG needs to be individualized based on a patient's severity of weakness, autoantibody status, age, gender, and comorbidities. Guidelines for therapeutic approaches have been proposed by groups across the globe. Although there is commonality, there are also differences that highlight variations in care patterns across nations.[45–47] The following sections provide a brief overview of therapies that is not designed to guide therapy. For a comprehensive review of MG treatment, see reference Wang et al.[48]

ACETYLCHOLINESTERASE INHIBITORS

For mild to moderate weakness, AChE inhibitors, most commonly pyridostigmine bromide, are the first line of treatment and for some are the only therapy required.[49] By retarding the hydrolysis of ACh at the NMJ, they increase the chance of

achieving an EPP to activate an action potential at the injured NMJ. Despite the elegant solution to the electrophysiological defect, pyridostigmine usually does not resolve weakness, must be dosed every few hours, and patients often have adverse effects of activation of the muscarinic cholinergic receptors, in particular nausea, bloating, and diarrhea.

CORTICOSTEROIDS

Oral corticosteroids are the most consistently effective medication for MG but also have the poorest adverse-effect profile.[50] Expert opinion and investigations vary considerably in what the ideal dosing should be, but all regimens are provided for months and often years at doses that produce complications.[4,47] A large minority, perhaps a third of patients respond poorly or are intolerant. Treatment resistance is likely related to variations in lymphocyte sensitivity to corticosteroids and unknown factors that define disease severity. Corticosteroids act by binding the glucocorticoid receptor and thereby influence transcriptional pathways that are known to impact apoptosis of lymphocytes and reduce proinflammatory cytokine activity.

THYMECTOMY

As described above, thymic pathology is a hallmark feature of MG for many patients and removal of the thymus in AChR-antibody-positive MG was found to be beneficial in a randomized controlled trial compared to prednisone therapy alone.[51] However, upwards of 30% of patients did not derive significant benefit, indicating that there continues to be a maintenance of pathogenic antibody production. All patients with a thymoma must have it removed to prevent local spread or more malignant transformation, but resection of the tumor and residual thymus does not produce remission. Patients who are seronegative or with MuSK antibodies are not thought to have thymic pathology and thymectomy is generally not performed.

IMMUNOSUPPRESSANTS

Because of the limitations of corticosteroids, a number of immunosuppressants are used in the treatment of MG, usually to reduce the burden of corticosteroid treatment. The importance of steroid sparing is so important that many clinical trials have used reduction of overall corticosteroid dose as a primary outcome measure for efficacy. The first agent to be used was azathioprine,[4] which inhibits purine synthesis leading to inhibition of lymphocyte replication.[52,53] Azathioprine may take more than a year to have a treatment effect and carries an increased risk of neoplasia, particularly lymphoma, with long-term use. Azathioprine also carries a risk of significant hepatotoxicity and myelosuppression. A commonly used agent in the United States is mycophenolate mofetil, which inhibits the de novo pathway for purine synthesis[54] leading to selective inhibition of lymphocyte proliferation. Several months are required to observe a treatment effect. Overall, mycophenolate is well-tolerated, but significant leukopenia and atypical infections may occur. Mycophenolate is a teratogen and should not be used for women intending to become pregnant and may increase the risk of malignancy. Tacrolimus and cyclosporine are calcineurin inhibitors, which produce modulation of T-cell activation. Cyclosporine was first applied to MG treatment in the 1980s, demonstrating a steroid-sparing effect but with significant renal toxicity and increased rates of hypertension. Tremor and paresthesias may also occur and generally resolve after dose adjustment. Tacrolimus has a better safety profile and is widely used in Japan and in Europe.

Other immunosuppressives have been less commonly used. Methotrexate is another inhibitor of pyrimidine and purine syntheses and expected to inhibit cellular proliferation of autoreactive lymphocytes. Despite its frequent use in other autoimmune disease, it is not often used for MG and a randomized controlled trial of methotrexate did not find a steroid-sparing effect. Methotrexate can cause liver injury and significant anemia. Oral and intravenous cyclophosphamide has been used for patients failing other therapies. Cyclophosphamide inhibits lymphocyte replication and has been found to affect the balance of Th2 and Th1 cells, reduce proinflammatory cytokines, and modulate dendritic cell activation. Autologous hematopoietic stem-cell transplantation has been used in combination with cyclophosphamide. Cyclophosphamide can produce severe adverse effects including having carcinogenic and teratogenic potential.[48]

INTRAVENOUS IMMUNOGLOBULIN AND PLASMA EXCHANGE

Intravenous immunoglobulin (IVIG) and plasma exchange are used for significant MG exacerbations, with both therapies designed to reduce levels of pathogenic antibody rapidly but with no effect on antibody production. In rare situations, patients are provided chronic treatment in order to avoid adverse effects of prednisone and immunosuppression, but remission would not be expected with such a treatment plan. IVIG has been considered to have multiple effects on the immune system, but the predominant effect is likely through anti-idiotype recognition and interference with antibody recycling through neonatal Fc receptor (FcRn) pathways. Improvement from IVIG infusion may be seen within days but is often on the order of weeks. Headache is common during and soon after infusion and some patients experience flu-like symptoms and mild rash. Rarely, thrombosis may occur, leading to pulmonary embolism, stroke, and myocardial infarction. Plasma exchange is performed with resins designed to remove proteins of certain molecular weights, particularly circulating antibodies. A total of no more than six exchanges are done with benefit observed at times after one exchange; however, patients may take 2 weeks to improve. Significant complications are related to the large-bore catheters used by some centers to perform exchanges. Plasma exchange and IVIG have similar efficacy, with IVIG having fewer adverse events.

ECULIZUMAB

Eculizumab was approved for use by the US Food and Drug Administration (FDA) for generalized AChR-antibody-positive MG in 2017. The drug is a monoclonal antibody directed against the C5 component of complement. From use in other complement-mediated diseases and MG, eculizumab has been found to have a good safety record, but does carry the potential risk of infection with capsulated bacterium, particularly meningococcus, leading to the requirement for appropriate pre-treatment vaccination. Rarely a patient may be treatment resistant because of a genetic variant of the C5 epitope to which it binds. Eculizumab also suffers from being an extremely expensive drug with a cost of several hundred thousand dollars per year in the United States. Clinical trials have shown benefit for use of eculizumab in AChR+MG patients.[55] Eculizumab

has been the first therapeutic to be approved for treatment of MG through the FDA.

THERAPIES UNDER DEVELOPMENT

In the last decade there has been an explosion of therapeutic development for MG. There are several motivations for this, but the detailed understanding of key aspects of MG pathophysiology, as described above, is the key driver, coupled with significant unmet need. For detailed and more expansive discussion please see reference Wang et al.[48] Below we describe therapies that have been under evaluation.

REDUCING EFFECTOR ACTIVITY: MODERATION OF AUTOANTIBODY LEVELS AND COMPLEMENT INHIBITION

Inhibition of IgG recycling is the subject of two phase 3 trials. IgG molecules in circulation undergo endocytosis by endothelial cells, where the IgG binds the FcRn, and are then returned to the circulation.[56] Monoclonal antibodies against the FcRn disrupt FcRn-IgG interaction, which results in IgG catabolism leading to a reduction of serum IgG levels, including pathogenic antibodies. The validation of complement inhibitor therapy has led to motivation to improve on eculizumab. A small-molecule inhibitor of C5, administered as a daily subcutaneous injection, was found safe with a signal efficacy in a phase 2 study.[57]

MODERATING PROINFLAMMATORY SIGNALS

Focused moderation of proinflammatory signals has been a target of therapy with disappointing results. Etanercept, a TNF-α inhibitor, reduced steroid requirements in a small number of patients; however, some patients worsened with treatment. Because of elevated BAFF, serum levels and its role in promoting B-cell activation, as well as promising preclinical data, belimumab, an antibody directed toward BAFF, was initiated but failed to show an efficacy signal.[58]

Lymphocyte-Targeted Therapy

Nearly all B cells express the cell-surface antigen CD20. Rituximab, a chimeric IgG1 monoclonal antibody that targets CD20, is commonly used in clinical practice, but to date there has not been a clinical trial to confirm its efficacy. A phase 2 trial for patients with AChR antibody-positive MG failed to show reduction of corticosteroid dose. In a prospective, non-randomized study of MuSK antibody-positive patients, rituximab reduced the prednisone dose together with clinical improvement. Antibody-producing cells are mostly CD20 negative and would not be directly affected by rituximab. Bortezomib and ixazomib are proteasome inhibitors used in the treatment of cancers of lymphocytic origin. Each has been assessed in animal models of MG with benefit and applied to other autoimmune diseases but has not moved to clinical testing. CD19-directed antibodies are also being considered for clinical trials in MG. With the central role of lymphocytes in autoantibody production one could expect a significant therapeutic benefit; however, the emergence of the COVID-19 pandemic leads to a significant concern of exacerbation of the severity of the infection and a poor response to vaccination, if used.

As described above, T cells drive the B-cell activity of MG. Abatacept is a drug composed of the extracellular domain of human cytotoxic T-lymphocyte-associated antigen 4 (CTLA-4) and a modified Fc region of human immunoglobulin leading to blockade of the co-stimulatory signal to activate T cells that is mediated by CD80/CD86. Abatacept inhibits T-cell activation, resulting in reduced levels of proinflammatory cytokines that are implicated in MG. Monoclonal antibodies targeting the CD40–CD40L interaction are undergoing clinical trials (NCT02565576). Antigen-presenting cells constitutively express CD40 on their surface. Activated CD4 T cells interact with CD40 through CD40L (CD154), leading to enhanced humoral immune responses. Preclinical studies have been positive.

Induction of Tolerance

Several preclinical studies have been performed to induce tolerance by administration by various routes of nonpathogenic portions of the AChR. A phase 1b study of CV-MG-101 is underway. CV-MG01 is a combination of two synthetic peptides that complement the structure of the main immunogenic region of the AChR. The hope is for the exogenous peptide to induce anti-idiotypic and anti-clonotypic responses against binding sites of antigen receptors on autoreactive lymphocytes and reduce their activation. Stem-cell transplantation approaches have also been used in a limited number of patients to re-induce a state of tolerance as the new immune system repopulates. This is reserved for a small number of highly treatment-resistant patients and has had mixed results.

ON THE HORIZON

Future Directions for the Development of Immunomodulatory Therapies

Development of MG-specific immunotherapies based on immunopathogenic events:

- Molecular pathways driving autoantibody production
- B-cell-lineage inhibitors
- Anti-apoptotic markers on autoreactive B cells
- Aberrant microRNA expression
- Complement inhibitors targeted to the NMJ

FUTURE STUDIES

MG has long been considered to be among the best understood autoimmune disorders, fulfilling strict criteria of autoimmunity. With the discovery that autoantibodies drive its final effector mechanism, antibody removal approaches, such as plasma exchange, IVIG, and FcRn inhibition, have been applied as therapies. Definition of complement destruction as a driver of NMJ injury spanned application of complement inhibitors for MG. Although these treatments can produce dramatic improvement, they also identify patients who remain treatment resistant and do not ultimately moderate autoantibody generation. Existing immunosuppressive therapies act broadly and also do not effectively eliminate production of autoantibodies. Although the presence of autoantibodies is a good biomarker for disease diagnosis, autoantibodies do not predict the response to treatment, disease progression, or disease severity. In order to improve therapeutic development, including reducing the time of clinical trial, improvement in biomarkers, and illuminating the defects leading to induction, as well as maintenance of autoantibody production, additional study will be required. Given the advances in the last decade we can hope that the next decade will bring the field closer to these discoveries.

REFERENCES

1. Robertson DN. Enumerating neurology. *Brain*. 2000;123(4):663–664.
2. Kaminski HJ, Kusner LL. *Myasthenia Gravis and Related Disorders*. New York, NY: Springer International Publishing; 2018.
3. Gilhus NE, Tzartos S, Evoli A, et al. Myasthenia gravis. *Nat Rev Dis Primers*. 2019;5(1):30.
4. Gilhus NE. Myasthenia gravis. *N Engl J Med*. 2016;375(26):2570–2581.
5. Weis CA, Aban IB, Cutter G, et al. Histopathology of thymectomy specimens from the MGTX-trial: entropy analysis as strategy to quantify spatial heterogeneity of lymphoid follicle and fat distribution. *PLoS One*. 2018;13(6):e0197435.
6. Elmqvist D, Hofmann WW, Kugelberg J, Quastel DM. An electrophysiological investigation of neuromuscular transmission in myasthenia gravis. *J Physiol*. 1964;174:417–434.
7. Serra A, Ruff RL, Leigh RJ. Neuromuscular transmission failure in myasthenia gravis: decrement of safety factor and susceptibility of extraocular muscles. *Ann N Y Acad Sci*. 2012;1275:129–135.
8. Kaminski HJ, Li Z, Richmonds C, et al. Complement regulators in extraocular muscle and experimental autoimmune myasthenia gravis. *Exp Neurol*. 2004;189(2):333–342.
9. Hesser BA, Henschel O, Witzemann V. Synapse disassembly and formation of new synapses in postnatal muscle upon conditional inactivation of MuSK. *Mol Cell Neurosci*. 2006;31(3):470–480.
10. Jennings CG, Dyer SM, Burden SJ. Muscle-specific trk-related receptor with a kringle domain defines a distinct class of receptor tyrosine kinases. *Proc Natl Acad Sci U S A*. 1993;90(7):2895–2899.
11. Zhang W, Coldefy AS, Hubbard SR, Burden SJ. Agrin binds to the N-terminal region of Lrp4 protein and stimulates association between Lrp4 and the first immunoglobulin-like domain in muscle-specific kinase (MuSK). *J Biol Chem*. 2011;286(47):40624–40630.
12. Weatherbee SD, Anderson KV, Niswander LA. LDL-receptor-related protein 4 is crucial for formation of the neuromuscular junction. *Development*. 2006;133(24):4993–5000.
13. Rose NR, Bona C. Defining criteria for autoimmune diseases (Witebsky's postulates revisited). *Immunol Today*. 1993;14(9):426–430.
14. Gomez AM, Vrolix K, Martinez-Martinez P, et al. Proteasome inhibition with bortezomib depletes plasma cells and autoantibodies in experimental autoimmune myasthenia gravis. *J Immunol*. 2011;186(4):2503–2513.
15. Topakian R, Zimprich F, Iglseder S, et al. High efficacy of rituximab for myasthenia gravis: a comprehensive nationwide study in Austria. *J Neurol*. 2019;266(3):699–706.
16. Lee JY, Stathopoulos P, Gupta S, et al. Compromised fidelity of B-cell tolerance checkpoints in AChR and MuSK myasthenia gravis. *Ann Clin Transl Neurol*. 2016;3(6):443–454.
17. Kusner LL, Ciesielski MJ, Marx A, et al. Survivin as a potential mediator to support autoreactive cell survival in myasthenia gravis: a human and animal model study. *PLoS One*. 2014;9(7):e102231.
18. Lindstrom JM, Engel AG, Seybold ME, et al. Pathological mechanisms in experimental autoimmune myasthenia gravis. II. Passive transfer of experimental autoimmune myasthenia gravis in rats with anti-acetylcholine receptor antibodies. *J Exp Med*. 1976;144(3):739–753.
19. Albazli K, Kaminski H, Howard JF. Complement inhibitor therapy for myasthenia gravis. *Front Immunol*. 2020;11:917.
20. Subias M, Tortajada A, Gastoldi S, et al. A novel antibody against human factor B that blocks formation of the C3bB proconvertase and inhibits complement activation in disease models. *J Immunol*. 2014;193(11):5567–5575.
21. Tüzün E, Yılmaz V, Parman Y, et al. Increased complement consumption in MuSK-antibody-positive myasthenia gravis patients. *Med Princ Pract*. 2011;20(6):581–583.
22. Heckmann JM, Uwimpuhwe H, Ballo R, et al. A functional SNP in the regulatory region of the decay-accelerating factor gene associates with extraocular muscle pareses in myasthenia gravis. *Genes Immun*. 2010;11(1):1–10.
23. Drachman DB, Angus CW, Adams RN, et al. Myasthenic antibodies cross-link acetylcholine receptors to accelerate degradation. *N Eng J Med*. 1978;298(20):1116–1122.
24. Vernet-der Garabedian B, Morel E, Bach JF. Heterogeneity of antibodies directed against the alpha-bungarotoxin binding site on human acetylcholine receptor and severity of myasthenia gravis. *J Neuroimmunol*. 1986;12(1):65–74.
25. Ghazanfari N, Morsch M, Reddel SW, et al. Muscle-specific kinase (MuSK) autoantibodies suppress the MuSK pathway and ACh receptor retention at the mouse neuromuscular junction. *J Physiol*. 2014;592(13):2881–2897.
26. Morsch M, Reddel SW, Ghazanfari N, et al. Muscle specific kinase autoantibodies cause synaptic failure through progressive wastage of postsynaptic acetylcholine receptors. *Exp Neurol*. 2012;237(2):286–295.
27. Huijbers MG, Zhang W, Klooster R, et al. MuSK IgG4 autoantibodies cause myasthenia gravis by inhibiting binding between MuSK and Lrp4. *Proc Natl Acad Sci U S A*. 2013;110(51):20783–20788.
28. Mori S, Motohashi N, Takashima R, et al. Immunization of mice with LRP4 induces myasthenia similar to MuSK-associated myasthenia gravis. *Exp Neurol*. 2017;297:158–167.
29. Shen C, Lu Y, Zhang B, et al. Antibodies against low-density lipoprotein receptor-related protein 4 induce myasthenia gravis. *J Clin Invest*. 2013;123(12):5190–5202.
30. Schluep M, Willcox N, Vincent A, et al. Acetylcholine receptors in human thymic myoid cells in situ: an immunohistological study. *Ann Neurol*. 1987;22(2):212–222.
31. Sims GP, Shiono H, Willcox N, Stott DI. Somatic hypermutation and selection of B cells in thymic germinal centers responding to acetylcholine receptor in myasthenia gravis. *J Immunol*. 2001;167(4):1935–1944.
32. Meraouna A, Cizeron-Clairac G, Panse RL, et al. The chemokine CXCL13 is a key molecule in autoimmune myasthenia gravis. *Blood*. 2006;108(2):432–440.
33. Thangarajh M, Masterman T, Helgeland L, et al. The thymus is a source of B-cell-survival factors—APRIL and BAFF—in myasthenia gravis. *J Neuroimmunol*. 2006;178(1–2):161–166.
34. Thiruppathi M, Rowin J, Ganesh B, et al. Impaired regulatory function in circulating CD4(+)CD25(high)CD127(low/-) T cells in patients with myasthenia gravis. *Clin Immunol*. 2012;145(3):209–223.
35. Cao Y, Amezquita RA, Kleinstein SH, et al. Autoreactive T cells from patients with myasthenia gravis are characterized by elevated IL-17, IFN-gamma, and GM-CSF and diminished IL-10 production. *J Immunol*. 2016;196(5):2075–2084.
36. Yi JS, Guidon A, Sparks S, et al. Characterization of CD4 and CD8 T cell responses in MuSK myasthenia gravis. *J Autoimmun*. 2014;52:130–138.
37. Wang W, Milani M, Ostlie N, et al. C57BL/6 mice genetically deficient in IL-12/IL-23 and IFN-gamma are susceptible to experimental autoimmune myasthenia gravis, suggesting a pathogenic role of non-Th1 cells. *J Immunol*. 2007;178(11):7072–7080.
38. Mu L, Sun B, Kong Q, et al. Disequilibrium of T helper type 1, 2 and 17 cells and regulatory T cells during the development of experimental autoimmune myasthenia gravis. *Immunology*. 2009;128(1 Suppl):e826–e836.
39. Le Friec G, Kohl J, Kemper C. A complement a day keeps the Fox(p3) away. *Nat Immunol*. 2013;14(2):110–112.
40. Leite MI, Jones M, Ströbel P, et al. Myasthenia gravis thymus: complement vulnerability of epithelial and myoid cells, complement attack on them, and correlations with autoantibody status. *Am J Pathol*. 2007;171(3):893–905.
41. Avidan N, Le Panse R, Berrih-Aknin S, Miller A. Genetic basis of myasthenia gravis—a comprehensive review. *J Autoimmun*. 2014;52:146–153.
42. Cron MA, Maillard S, Truffault F, et al. Causes and consequences of miR-150-5p dysregulation in myasthenia gravis. *Front Immunol*. 2019;10:539.
43. Molin CJ, Sabre L, Weis CA, et al. Thymectomy lowers the myasthenia gravis biomarker miR-150-5p. *Neurol Neuroimmunol Neuroinflamm*. 2018;5(3):e450.
44. Cron MA, Guillochon É, Kusner L, Le Panse R. Role of miRNAs in normal and myasthenia gravis thymus. *Front Immunol*. 2020;11:1074.
45. Sussman J, Farrugia ME, Maddison P, et al. Myasthenia gravis: Association of British Neurologists' management guidelines. *Pract Neurol*. 2015;15(3):199–206.
46. Murai H. Japanese clinical guidelines for myasthenia gravis: putting into practice. *Clin Exp Neuroimmunol*. 2015;2015:21–31.
47. Sanders DB, Wolfe GI, Benatar M, et al. International consensus guidance for management of myasthenia gravis: executive summary. *Neurology*. 2016;87(4):419–425.

48. Wang S, Breskovska I, Gandhy S, et al. Advances in autoimmune myasthenia gravis management. *Expert Rev Neurother*. 2018;18(7):573–588.
49. Maggi L, Mantegazza R. Treatment of myasthenia gravis: focus on pyridostigmine. *Clin Drug Investig*. 2011;31(10):691–701.
50. Hoffmann S, Kohler S, Ziegler A, Meisel A. Glucocorticoids in myasthenia gravis—if, when, how, and how much? *Acta Neurol Scand*. 2014;130(4):211–221.
51. Wolfe GI, Kaminski HJ, Aban IB, et al. Randomized trial of thymectomy in myasthenia gravis. *N Engl J Med*. 2016;375(6):511–522.
52. Pelin M, De Iudicibus S, Londero M, et al. Thiopurine biotransformation and pharmacological effects: contribution of oxidative stress. *Curr Drug Metab*. 2016;17(6):542–549.
53. Shin JY, Wey M, Umutesi HG, et al. Thiopurine prodrugs mediate immunosuppressive effects by interfering with Rac1 protein function. *J Biol Chem*. 2016;291(26):13699–13714.
54. Villarroel MC, Hidalgo M, Jimeno A. Mycophenolate mofetil: an update. *Drugs Today (Barc)*. 2009;45(7):521–532.
55. Howard JF Jr, Utsugisawa K, Benatar M, et al. Safety and efficacy of eculizumab in anti-acetylcholine receptor antibody-positive refractory generalised myasthenia gravis (REGAIN): a phase 3, randomised, double-blind, placebo-controlled, multicentre study. *Lancet Neurol*. 2017;16(12):976–986.
56. Liu X, Ye L, Christianson GJ, et al. NF-kappaB signaling regulates functional expression of the MHC class I-related neonatal Fc receptor for IgG via intronic binding sequences. *J Immunol*. 2007;179(5):2999–3011.
57. Howard Jr. JF, Nowak RJ, Wolfe G, et al. Complement inhibitor Zilucoplan in patients with moderate to severe generalized myasthenia gravis. *JAMA Neurol*. 2020;77(5):582–592.
58. Hewett K, Sanders DB, Grove RA, et al. Randomized study of adjunctive belimumab in participants with generalized myasthenia gravis. *Neurology*. 2018;90(16):e1425–e1434.

66

Multiple Sclerosis

Andrew R. Romeo and Benjamin M. Segal

Multiple sclerosis (MS), a chronic inflammatory demyelinating disorder of the central nervous system (CNS), is the most frequent cause of nontraumatic neurologic disability among young adults in the Western Hemisphere. Although MS is widely considered a disease of North America and Europe, there is increasing evidence that it is more common in other regions of the world than previously appreciated, including Asia and the Middle East. The median age of presentation is between 28 and 31 years, in part responsible for the disproportionately high social and economic toll of the disease. Furthermore, the incidence of MS is increasing for unknown reasons. Fortunately, there have been dramatic advances in the treatment of relapsing forms of MS over the past 20 years, spurred by the introduction of a growing number of disease-modifying therapies (DMTs), and more are actively under development. These treatments significantly decrease the risk of future relapse and lesion formation. Consequently, the implications of a relapsing-remitting (RR) MS diagnosis have changed considerably in the span of a generation. Despite this success, significant challenges remain. There is a dire need for more effective treatments that slow or halt disability accumulation in patients with progressive forms of MS, and for interventions that restore lost neurologic functions across MS subsets.

CLINICAL SUBSETS AND PHENOMONOLOGY

Relapsing-Remitting Multiple Sclerosis

In the majority of cases (85% to 90%), MS presents with a RR course, characterized by discrete episodes of neurologic dysfunction (relapses) separated by clinically quiescent periods (remissions). The frequency of relapses can vary widely between people with MS, as well as during different time periods of an individual's disease course. Currently, no clinical features or biomarkers have been identified that are predictive of relapse rate. The signs and symptoms of relapses are also diverse and unpredictable, because lesions can form at any site in the CNS, spanning the cerebrum, brainstem, cerebellum, optic nerves, and spinal cord. The peripheral nervous system is spared.

MS lesions are visualized in CNS white matter via magnetic resonance imaging (MRI) (Fig. 66.1). Symptomatic lesions generally occur in locations where nerve fibers converge to subserve a common function. Typical presentations of RRMS include optic neuritis with monocular visual deficits (secondary to lesions in the optic nerve); myelitis with weakness and numbness in the extremities, sometimes accompanied by incontinence (due to spinal cord lesions); and brainstem syndromes manifested by imbalance, tremor, double vision, slurred speech, or swallowing difficulties. Serial MRI studies have demonstrated that the majority of MS lesions are actually clinically silent; redundant nerve fiber tracts are abundant, and large areas of cerebral white matter are committed to personality traits and cognitive skills. Consequently, CNS tissue damage may be inflicted surreptitiously during clinical remissions, making MRI a more sensitive indicator of disease activity than the history or neurologic exam.

People with MS often recover function following a clinical relapse, partially or fully, particularly during the early clinical course. Old symptoms can temporarily reemerge when core body temperature is elevated due to infection, strenuous exercise, or environmental conditions. This unmasking of latent deficits, called the Uhthoff phenomenon, is a consequence of the physiologic slowing of axon signal propagation that normally occurs at high core body temperatures. In healthy individuals the degree of slowing has no clinical consequence, but in MS patients it may precipitate the decompensation of white matter tracts already compromised by demyelination and axonal drop out.

Secondary Progressive Multiple Sclerosis

During the course of RRMS, relapses decrease in frequency over time and sometimes disappear completely. In the vast majority of cases they are replaced by an insidious, gradual accumulation of disability, referred to as the secondary progressive (SP) stage. Progressive myelopathy, hemiparesis, and/or gait imbalance are common. Subcortical dementia is increasingly recognized as a feature of the disease. Longitudinal natural history studies conducted before DMTs were widely available found that the majority of RRMS patients transitioned to an SP stage within 15 to 25 years of initial disease presentation. A 2010 epidemiologic study of MS patients in British Columbia found that the median time to secondary progressive multiple sclerosis (SPMS) onset was 21.4 years.[1] Factors that are associated with both a shorter time to, and younger age at, evolution to SPMS include male gender, the presence of motor symptoms at clinical onset, and history of poor recovery from relapses. Previous longitudinal observational and retrospective cohort studies that investigated whether treatment with first-generation DMTs alters the time to reach SPMS yielded conflicting results. However, a prospective study of 517 patients actively treated with second-generation DMTs found that rates of worsening and evolution to SPMS were substantially lower when compared with earlier natural history studies of untreated patients.[2]

The cellular and molecular mechanisms underlying the conversion from an RR to an SP course are poorly understood. Some investigators have questioned the relevance of neuroinflammation during the SP stage and have evoked neurodegeneration (in the form of neuronal death, mitochondrial dysfunction of axons, wallerian degeneration, and gliosis) as

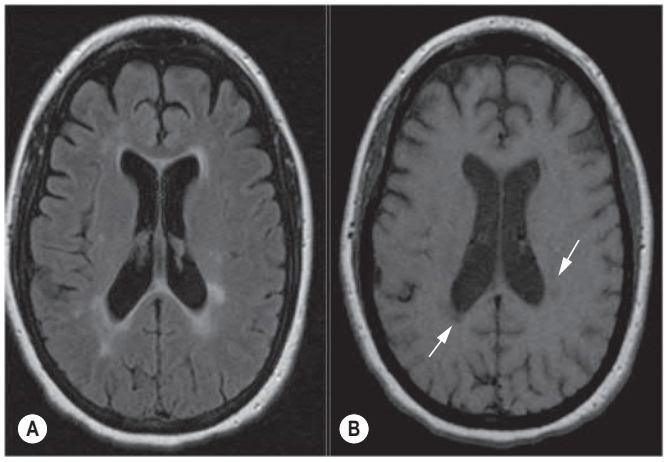


FIG. 66.1 (A) T2-weighted fluid-attenuated inversion recovery (FLAIR) magnetic resonance imaging (MRI) image of the brain of a multiple sclerosis patient showing hyperintense lesions located in the periventricular and subcortical white matter. (B) T1-weighted MRI image showing T1-black holes (*arrows*), indicative of profound axonal loss, and generalized atrophy with enlarged ventricles.

primarily responsible for the clinical deterioration that ensues following an initial period of immune-mediated damage.[3] Conversely, there is evidence of persistent immune dysregulation in SPMS, although it may differ from RRMS with regard to the cytokine networks and leukocyte subsets involved, as well as the distribution of infiltrating cells within the CNS.[4–6] Previously, a number of DMTs that are therapeutically beneficial in RRMS were found to be ineffective in slowing disability accumulation in SPMS.[7] However, potent lymphocyte-targeting therapies (cladribine and siponimod) were recently shown to curtail CNS tissue damage and clinical progression in subsets of individuals with SPMS, particularly those with ongoing or recent inflammatory activity ("active SPMS"), as indicated by superimposed clinical relapses and/or new T2 or enhancing lesion formation.[8,9] The management of SPMS still largely involves alleviation of symptoms, optimization of residual functions, and prevention of complications.

Primary Progressive Multiple Sclerosis

Primary progressive multiple sclerosis (PPMS) is distinguished from SPMS by the absence of an antecedent RR phase. Otherwise, the clinical features of SPMS and PPMS can be indistinguishable. The most common clinical phenotype is spastic paraparesis, followed by cerebellar dysfunction, and hemiplegia. There are striking demographic differences between PPMS and RR/SPMS. PPMS tends to present at an older age (peaking in the fifth and sixth decades) than RRMS (which peaks in the third and fourth decades). Although RRMS occurs two to three times more frequently in females than males, in PPMS the sex ratio is closer to 50:50. The neuropathologic and radiologic features of PPMS overlap extensively with those of SPMS, leading some investigators to conclude that PPMS and SPMS belong to the same disease spectrum. Familial clusters of MS that include members with PPMS and others with RR/SPMS, support that viewpoint. Possibly, analogous to SPMS, acute inflammatory lesions form during early stages of PPMS, prior to overt clinical progression, but happen to arise exclusively in clinically silent areas. Other investigators have argued that PPMS is a distinct disease entity driven initially and primarily by neurodegenerative processes.[3] This stance is supported by the general failure of first-generation immunomodulatory agents to attenuate the course of PPMS. Conversely, treatment with ocrelizumab, a B cell–depleting monoclonal antibody, delayed disability progression in a subset of relatively young subjects with PPMS.[10] The positive trial of siponimod in active SPMS is reminiscent of the positive trial of ocrelizumab in PPMS.[9] The pathogenesis of progressive MS, including PPMS, appears heterogeneous; the relative contributions of inflammation and neurodegeneration to clinical outcomes likely vary between patients.

CLINICAL PEARLS

- Multiple sclerosis (MS) presents with a relapsing-remitting course in the majority of cases (85%–90%).
- The symptoms and signs experienced by MS patients are diverse; lesions can form at any site in the central nervous system, including the optic nerves, cerebrum, brainstem, and spinal cord.
- The rate, severity, and symptoms of relapses are highly variable and unpredictable.
- Most MS lesions form silently; magnetic resonance imaging scans are a more sensitive gauge of disease activity than the history or neurologic exam.
- Acute relapses tend to decrease in frequency over time and are typically replaced by a gradual accumulation of disability. This later phase of disease is referred to as secondary progressive MS (SPMS).
- Less frequently, MS begins with a progressive course, referred to as primary progressive MS (PPMS).

DIAGNOSIS

The McDonald criteria are the most widely used guidelines for diagnosing MS. Originally proposed by the International Panel on Diagnosis of Multiple Sclerosis in 2001, they were revised most recently in 2017 (Table 66.1).[11] An MS relapse is "a clinical episode with patient-reported symptoms and objective findings typical of MS, reflecting a focal or multifocal inflammatory demyelinating event in the CNS, developing acutely or subacutely with a duration of at least 24 hours, with or without recovery, and in the absence of fever or infection." RRMS is, by definition, a dynamic multifocal inflammatory demyelinating disease of the CNS. Therefore the demonstration of lesion dissemination in time and space is essential for diagnosis. Dissemination in space can be demonstrated by objective clinical evidence of involvement of two or more sites in the CNS (based on neurologic exam and/or delayed latencies on evoked potential testing) or by the presence of T2-weighted MRI lesions in at least two of the following areas in the CNS: periventricular, juxtacortical/cortical, infratentorial, and spinal cord. The criterion for dissemination in time can be satisfied by either two or more distinct clinical relapses, one relapse followed by the interval appearance of a new lesion on serial MRI scans, or the simultaneous presence of gadolinium-enhancing (i.e., acute inflammatory) and nonenhancing MRI lesions. The presence of cerebrospinal fluid (CSF)-specific oligoclonal bands may substitute for clinical and/or radiographic demonstration of dissemination in time.

When a patient has had one clinical attack typical of inflammatory demyelination, and the criterion for dissemination in time is not satisfied, the diagnosis is clinically isolated syndrome (CIS). The presence of two or more typical demyelinating lesions on MRI, that do not localize with the presenting

TABLE 66.1 Revised McDonald Criteria for Diagnosis of Relapsing-Remitting Multiple Sclerosis (in a patient presenting with a clinical attack)

Number of Clinical Attacks	Number of Lesions With Objective Clinical Evidence[a]	Additional Data Needed for Diagnosis
≥2	≥2	None
≥2	1, with clear-cut historical evidence of a previous attack involving a distinct neuroanatomic location	None
≥2	1	Dissemination in space (an additional clinical attack implicating a different CNS region or by MRI)
1	≥2	Dissemination in time (an additional clinical attack or a new T2-hyperintense lesion or gadolinium-enhancing lesion on subsequent MRI) *or* CSF-specific oligoclonal bands
1	1	Dissemination in space *and* Dissemination in time or CSF-specific oligoclonal bands

[a]Objective clinical evidence includes an abnormality on neurologic examination, MRI, optical coherence tomography, or evoked potential testing, corresponding to the neuroanatomic location suggested by the clinical attack.
CNS, Central nervous system; *CSF*, cerebrospinal fluid; *MRI*, magnetic resonance imaging.

neurologic deficit(s), substantially increases the risk that CIS will convert to clinically definite MS in the following 5 to 10 years. With increased utilization of MRI, CNS lesions typical of demyelinating plaques may be incidentally discovered in an individual with no history of clinical exacerbations suggestive of CIS or MS. This condition has been termed radiologically isolated syndrome (RIS). In one study of 451 RIS subjects, 34% of individuals experienced an acute neurologic episode consistent with a demyelinating event, or progressive neurologic decline lasting at least 12 months, within a 5-year period from the first brain MRI study.[12] Interestingly, of those who developed symptoms, 9.6% fulfilled criteria for PPMS.

There are no clinical features or biomarkers that are pathognomonic for MS; it is essential to rule out competing diagnoses. The presence of unique oligoclonal bands and/or elevated immunoglobulin G (IgG) index in the CSF (indicative of primary antibody production in the CNS/intrathecal compartment) supports a diagnosis of MS. However, both findings are observed in a wide range of neuroinflammatory conditions, including subacute sclerosing panencephalitis, neurosarcoidosis, Lyme disease, and systemic lupus erythematosus with CNS involvement.

A diagnosis of PPMS requires 1 year of disease progression independent of relapses, plus two of the three following criteria: (i) at least one T2 hyperintense lesion in the periventricular, juxtacortical/cortical, or infratentorial regions; (ii) at least two T2-hyperintense spinal cord lesions, and (iii) the presence of oligoclonal bands on CSF analysis.

RISK FACTORS

Genetic Risk Factors

A monozygotic twin whose co-twin has MS has a 20% to 30% risk of developing the disease, and the corresponding risk for a dizygotic twin is 2% to 5%. The incidence of MS in the general population is approximately 0.1%. The risk of MS drops with increasing genetic distance, suggesting genes play a significant (but incomplete) role in determining MS susceptibility. The human leukocyte antigen (HLA) region of chromosome 6p 21.3 represents the strongest MS susceptibility locus genome wide and has been implicated in all ethnic populations thus far studied. It accounts for up to 10.5% of genetic variance underlying risk.[13] The primary signal maps to the *HLA-DRB1* gene in the class II segment of the locus (in particular the variant *HLA-DRB1*15:01*), implicating a role of CD4 T-cell responses in MS pathogenesis.[14] A class I variant (implicating CD8 T-cell responses), *HLA-A*02*, has been associated with protection from MS.[14] Genome-wide association studies (GWAS) have revealed more than 100 non-HLA susceptibility loci, each of which contributes a small amount to MS risk. Strikingly, most of these map to regions containing genes implicated in immunologic, rather than neuronal or glial, pathways. Genes involved in T helper (Th) cell differentiation are overrepresented, including the interleukin (IL)-2 receptor α chain and the IL-7 α chain, which modulate T-cell proliferation and survival. More than one-third of the MS susceptibility loci overlap with regions previously identified in GWAS of other autoimmune diseases, including celiac disease, type 1 diabetes, rheumatoid arthritis, and/or inflammatory bowel disease. Together these data provide support for a primary immunologic, as opposed to neurodegenerative, etiology of MS. An unresolved question is whether genetic variants affect the clinical course of RRMS, including such features as relapse rate or severity, and time to conversion to the SP stage. Genetic variants have yet to be identified that are predictive of responsiveness to DMT. Epigenetic modifications are a growing area of interest.

Environmental Risk Factors

Twin concordance rates in MS have been used to highlight the importance of heredity in MS susceptibility. Paradoxically, the same data can be used to argue for the importance of environmental influences; more than 70% of monozygotic twins of individuals with MS do not develop the disease. Despite arduous attempts, no convincing evidence has been produced for genetic, epigenetic, or transcriptome differences that explain MS discordance among monozygotic twin pairs. The risk of MS appears to be predicated on a complex interplay between genetic and environmental factors.

Geographic Prevalence Patterns

One of the most convincing illustrations of the impact of the environment on the development of MS is its geographic distribution. The prevalence of MS is highest in the Scandinavian countries, Canada, and Scotland and lowest in the equatorial regions. Kurtzke and colleagues described a latitudinal gradient in MS prevalence across the United States, peaking in the Northern states and gradually declining towards the South.[15] Similar latitudinal gradients have been observed in Europe, Australasia, and Japan. Prepubescent children who migrate assume the risk of their adopted country, whereas adults carry forward the risk of the location where they spent their childhood. An environmental

agent encountered in childhood that acts as a predisposing factor for the development of MS later in life has yet to be identified.

Vitamin D

Vitamin D modulates multiple components of the adaptive and innate immune responses, likely suppressing autoimmune inflammation. The discovery of a protective role of vitamin D in MS may explain, in part, the geographic distribution of MS. Ultraviolet light catalyzes the conversion of vitamin D to its bioactive form, and the prevalence of MS is highest in regions with relatively low annual sunlight exposure. In a large prospective study, the risk of MS decreased with increasing serum levels of 25-hydroxyvitamin D.[16] Therefore interventions that raise the level of 25-hydroxyvitamin D might have a prophylactic protective effect on healthy individuals who are predisposed to develop MS, such as the first-degree relatives of MS patients. Numerous independent prospective studies have found an inverse relationship between dietary or supplemental vitamin D intake in adults and future risk of MS. There is accumulating evidence that low serum 25-hyroxyvitamin D levels and low dietary vitamin D intake during pregnancy increase the risk of MS in the offspring. Prospective and retrospective studies have found an association between higher levels of 25-hydroxyvitamin D and lower risks of relapse and radiologic disease activity in individuals with established MS.[16] Multiple add-on studies examining vitamin D supplementation support a beneficial effect on MS disease activity. A randomized controlled trial of vitamin D supplementation in RRMS is currently underway in 16 academic centers (ClinicalTrials.gov identifier: NCT01490502).

Infection

Another prognostic factor for development of MS is primary infection with Epstein-Barr virus (EBV) as an adult, indicated by emergence of EBV-specific IgM antibodies in the serum.[17] The odds of MS risk are estimated to be more than 10 times higher among EBV-positive than EBV-negative persons. EBV infection could promote MS pathogenesis via "molecular mimicry," in which a microbial epitope shares sequence similarities with a self-peptide (see Chapter 51). T-cell receptors (TCRs) that cross-react with structurally homologous EBV and myelin antigens have been discovered in the 1 peripheral CD4 TCR repertoire of individuals with MS.[18] Peripheral CD4 T cells that express such cross-reactive TCRs could be activated during EBV infection, enabling them to cross the blood–brain barrier (BBB), encounter their cognate myelin antigen within the CNS white matter, and initiate MS lesion formation. An alternative theory involves CNS infiltration by EBV-infected B cells. B cells expressing EBV messenger RNA and proteins have been detected in meningeal follicle-like structures and inflamed cortical lesions in SPMS brain tissue.[19] EBV-driven expansion and activation of meningeal B cells could potentially contribute to the formation of the follicle-like structures, which have been associated with large subpial cortical MS lesions and a more aggressive clinical course.[6]

Although infectious agents such as EBV have been implicated in MS pathogenesis, other pathogens may have a therapeutic effect. Cytomegalovirus (CMV) seropositivity has been associated with reduced risk of MS.[14] It has been proposed that, in addition to high levels of sunlight exposure, endemic helminth infection is one of the factors responsible for the low prevalence of MS in tropical regions. Infection with helminths is protective in animal models of MS.[20] Regulatory B cells, Th2 cells, and/or eosinophils and immunosuppressive cytokines, such as IL-4, IL-10, and transforming growth factor-β (TGF-β), have been implicated in the protection. Several small prospective studies showed that MS patients naturally infected with different species of parasitic worms had a milder disease course and lower MRI inflammatory activity than uninfected patients.[21] Antiparasite treatment was associated with exacerbation of MS. Mechanistic substudies suggest that parasites modulate MS disease activity by boosting the frequency of IL-10 and TGF-β–producing regulatory T and/or B cells.

Obesity

Early adulthood obesity, measured by elevated body mass index (BMI), confers an approximately twofold increased risk of MS.[14] Obesity in girls is associated with an increased risk of pediatric MS or CIS. Higher weight in adolescence and young adulthood is associated with an earlier age at the onset of MS. A mendelian randomization analysis of large GWAS for MS and BMI, respectively, found that one standard deviation increase in genetically-determined BMI conferred a 41% increase in the odds of MS.[22] A number of theories have been proposed for a mechanistic link between obesity and MS risk. First, obesity is known to cause a systemic proinflammatory state, possibly mediated by adipose-derived hormones that create a milieu conducive to the differentiation and/or activation of autoimmune effector cells. Alternatively, there is evidence that genetically elevated BMI decreases 25-hydroxyvitamin D levels. It is unknown if obesity influences MS disease course.

Cigarette Smoking

A substantial body of literature indicates that cigarette smoking increases the risk of MS. A meta-analysis revealed a dose–response relationship between cigarette pack-years and MS risk.[23] Based on data collected from the Swedish National MS Registry, it was estimated that each additional year of smoking after diagnosis accelerates the time to conversion to SPMS by 4.7%; the association between cigarette smoking and risk of conversion from RRMS and SPMS has borne out in meta-analysis.[24] There is growing evidence that passive exposure to smoking may be a risk factor for MS, although data are mixed. Cigarette smoking may interact with HLA and other genetic variants to increase MS risk.[14] It is postulated that cigarette smoking promotes an inflammatory environment in the respiratory system, stimulates autoreactive lymphocytes via cross-reactivity, and promotes breakdown of the blood-brain barrier (BBB). Cigarette smoking may have direct toxic effects on neurons and/or glia.

Sex Hormones

Similar to other autoimmune diseases, MS prevalence is higher among women (2:1 to 3:1), but men with MS often have a worse prognosis.[25] MS relapse rates decline during pregnancy, particularly in the third trimester, and rebound in the first 3 months postpartum before returning to the prepregnancy rate. Parity may impact the risk of MS. These observations have led to the hypothesis that certain female sex hormones may play a protective role in RRMS. Estriol is an estrogen unique to pregnancy. Synthesized by the fetal–placental unit, estriol reaches its highest levels in the last trimester. A randomized, double-blinded, placebo-controlled phase 2 trial of estriol in combination with glatiramer acetate (GA) versus placebo plus GA showed a reduction in the annualized relapse rate at 2 years in the estriol-treated group.[26] Animal model studies indicate that estrogens, including estriol, have antiinflammatory and neuroprotective

effects through engagement of estrogen receptors expressed on leukocytes and CNS-resident cells, respectively.

Testosterone has neuroprotective effects in animal models of MS, and decreased testosterone levels in males with MS were reported to be associated with disability. In a small open-label phase 2 clinical trial, testosterone treatment appeared to arrest gray matter (GM) loss (and even to reverse GM atrophy in the right frontal cortex), as quantified using voxel-based morphology, in 10 male patients with MS.[27]

Microbiome

Commensal microorganisms of the intestines (gut microbiota) influence immune homeostasis, possibly playing a role in autoimmune disease. Germ-free mice (wild-type mice treated with oral antibiotics to eliminate commensal bacteria) are relatively resistant to the induction of experimental autoimmune encephalomyelitis (EAE), the most widely used animal model of MS.[28] Germ-free mice that are colonized with segmented filamentous bacteria or that received fecal transplants from MS patients develop EAE.[29] Collectively, these observations have led to the contention that specific microbiota promote encephalitogenic T-cell responses, via expression of bacterial proteins that cross-react with myelin antigens or production of metabolites that drive effector T-cell polarization. Certain microbial products could cross the BBB and activate glial cells. The majority of publications on the MS microbiome have found no major difference in gut microbiota diversity between MS cases and controls.[30] However, subtle taxonomic differences have been detected, with consistent patterns emerging across studies. In some studies, exposure to immunomodulatory drugs was associated with individual taxonomic differences. Of note, these studies were generally too small to adequately assess potential effect modifiers. Several larger studies are ongoing.

KEY CONCEPTS

Risk Factors

- The risk of multiple sclerosis (MS) is determined by a combination of genetic and environmental factors.
- The majority of MS susceptibility loci map to regions containing genes implicated in immunologic pathways, including human leukocyte antigen (HLA) class II molecules, the interleukin (IL)-2 receptor and the IL-17 receptor.
- Relapse rates decline during the third trimester of pregnancy, in association with high serum levels of estriol.
- Environmental risk factors include low vitamin D levels, exposure to the Epstein-Barr virus as an adult, cigarette smoking, and childhood obesity.

PATHOLOGIC FEATURES OF MULTIPLE SCLEROSIS

White Matter Lesions

The hallmark of MS pathology is the focal demyelinated lesion, or "plaque," present in the white matter of the optic nerves, brain, and spinal cord. Acute lesions are invariably associated with focal breakdown of the BBB and perivascular inflammatory infiltrates. MS infiltrates are dominated by T cells (with a relatively high CD8:CD4 ratio) and myeloid cells (blood-derived monocytes/macrophages and activated microglia). Macrophages/monocytes and activated microglia are spatially associated with disintegrating myelin sheaths and they actively take up myelin debris. Oligodendrocyte apoptosis and loss varies widely between lesions. Frequent sites of lesion formation include the subcortical and periventricular cerebral white matter, middle cerebellar peduncles, and posterior columns of the cervicothoracic spinal cord. In the brain, infiltrates frequently follow the course of pericallosal venules, resulting in "Dawson fingers," which are oblong lesions oriented perpendicular to the long axes of the lateral ventricles (Fig. 66.2).

Classic actively demyelinating plaques are primarily seen during the RR stage of disease and generally drop in frequency with increasing disease duration. Lesions more typical of progressive forms of MS have been termed "chronic active" or smoldering, and "chronic silent" or inactive. Chronic active plaques are distinguished by a rim of activated microglia and deposits of complement at the lesion edge, surrounding a hypocellular and gliotic core. They are slowly expansive as a consequence of active demyelination at the lesion edge. In contrast, chronic silent plaques have a sharp border. Other characteristics of silent plaques include prominent loss of oligodendrocytes and axons, pronounced astrogliosis, and a paucity of macrophages and activated microglia. Immunopathologic changes in the so-called normal-appearing white matter (NAWM), outside of plaques, are pervasive in progressive MS but have also been observed in RRMS. These changes consist of diffuse axonal injury and microglial activation, as well as scattered lymphocytes.

MS is widely classified as a demyelinating disorder. This is due to the fact that a large number of the nerve fiber segments traversing plaques demonstrate myelin loss with relative axonal sparing. However, it is now recognized that axonopathy also occurs and is, in fact, an early and prominent feature of acute MS lesions. Axonal damage results in dysmorphic mitochondria, focal swellings, fragmentation, and frank transections with terminal bulbs at the stumps. Mitochondrial abnormalities and focal swellings have been observed in fully myelinated axons within MS lesions, suggesting that they can occur independent of demyelination.[31] In animal models of MS, axons with

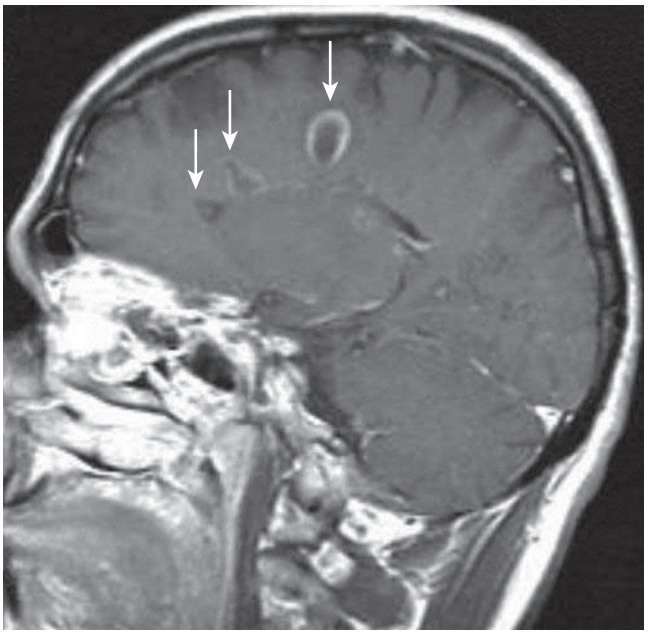


FIG. 66.2 A postgadolinium T1-weighed magnetic resonance image of the brain showing Dawson fingers *(arrows)*.

abnormal mitochondria are restricted to areas of immune infiltration, and progressive axonal changes correlate with the density of infiltrates.[31] Hence it is likely that the axonal damage is directly mediated by direct contact with inflammatory cells. Although demyelination can be reversed to some extent by remyelination, axonal transection is irreversible. Clinic-pathologic investigations have found that permanent motor disability in MS correlates with loss of corticospinal tract axons more so than with degree of demyelination.

Gray Matter Lesions

MS was traditionally considered a white matter disease. It is now established that the gray matter (GM) is affected as well. Three types of cortical lesions have been described: leukocortical (which span the gray and white matter), intracortical, and subpial. All of these lesions show demyelination and oligodendrocyte loss, microglial activation, neuritic transections, neuronal death, and reduced presynaptic terminals. Subpial lesions are the most common. They can cover long distances of the cortical ribbon and usually extend to cortical layer III or IV. Cortical lesions are not visible on conventional MRI scans and require special staining to be appreciated in CNS tissue sections. This explains why they were not recognized as common features of MS until recently. In fact, cortical demyelination and GM atrophy are evident from the earliest stages of disease, even before a clinically definite diagnosis can be made, and continue to advance at an increasing rate throughout the disease course. Extensive cortical demyelination is evident in the forebrain and cerebellum during progressive MS. GM atrophy in individuals with MS correlates strongly with cognitive deficits and clinical disability.[32]

Meningeal Inflammation

White blood cell counts tend to be within normal limits or only slightly elevated in the CSF of most patients with MS. Nonetheless, there is growing recognition that low-grade diffuse meningeal inflammation, as well as focal perivascular meningeal inflammation, are common. Meningeal inflammation is most prominent in progressive forms of MS but is prevalent in early MS as well. The meningeal infiltrates are topographically associated with cortical lesions. Lymphoid follicle-like structures, composed of proliferating B cells, T cells, and follicular dendritic cells, have been observed in up to 40% of autopsied brains from individuals with SPMS.[6] In almost every case the follicles reside in deep sulci and abut an underlying subpial lesion, suggesting that toxic factors are released by the inflammatory cells and diffuse into the brain parenchyma. The presence of lymphoid follicles has been associated with a more severe clinical course, shorter disease duration, and younger age at death.

KEY CONCEPTS

Pathology

- The hallmark of multiple sclerosis (MS) pathology is the focal demyelinated lesion, or "plaque," with perivascular inflammatory infiltration and focal blood–brain barrier breakdown.
- Axonopathy is an early and prominent feature of acute MS lesions.
- Central nervous system damage includes demyelination, oligodendrocyte apoptosis and loss, and axonal swellings and transections.
- Gray and white matter are both affected.
- The pathologic features of MS are heterogeneous and evolve over time.

IMMUNOPATHOGENESIS

Animal Models of Multiple Sclerosis

According to the current dogma, MS is an autoimmune disease mediated by CD4 T cells reactive against myelin antigens. The identification of HLA class II, *IL-2Rα*, and *IL-7Rα* as MS genetic susceptibility loci is consistent with a role of CD4 T cells in MS pathogenesis. An autoimmune etiology is further supported by the animal model, experimental autoimmune encephalomyelitis (EAE). EAE is a multifocal inflammatory demyelinating disease of the CNS that has striking histologic and clinical similarities to MS (Figs. 66.3 and 66.4). It has been induced in a wide variety of mammalian species (including nonhuman primates, but most commonly in rodents) by vaccination against MHC class II-restricted myelin epitopes. EAE can be transferred from myelin-vaccinated mice to syngeneic naïve hosts with purified CD4 T-cell lines or clones. These encephalitogenic myelin-specific CD4 T cells invariably fall within the Th1 or Th17 lineage and produce the proinflammatory cytokines interferon (IFN)-γ and IL-17, respectively, in response to antigenic stimulation (see Chapter 11).[33] Both Th1 and Th17 cells produce granulocyte–macrophage colony-stimulating factor (GM-CSF), a monocyte mobilizing and growth factor that plays a critical role in many models of EAE. Upon activation in the periphery, myelin-reactive CD4 T cells upregulate adhesion molecules and chemokine receptors, thereby acquiring the ability to cross the BBB. After infiltrating the CNS, they are reactivated by local antigen-presenting cells (APCs), such as perivascular macrophages or microglia, which constitutively express surface MHC class II molecules bound to myelin peptides. GM-CSF, as well as other Th1 and/or Th17 cytokines, are subsequently released in situ and initiate an inflammatory cascade, resulting in the production of chemokines, mobilizing factors, and vasoactive substances, upregulation of adhesion molecules on cerebrovascular endothelium, and, consequently, the recruitment of myeloid cells and lymphocytes from the circulation to the nascent plaque. GM-CSF may drive the differentiation of infiltrating monocytes and CNS-resident microglia into $CD11c^+$ dendritic cells, which are among the most potent APC. Adoptive transfer studies with labeled donor T cells have demonstrated that the myelin-specific T cells remain clustered in the perivascular space throughout lesion development. A secondary wave of myeloid cells infiltrates deep into the CNS white matter, associates with nodes of Ranvier, and directly inflicts damage to the myelin sheath and axons (see Figs. 66.3 and 66.4).[31]

Immune Dysregulation in Patients With Multiple Sclerosis

Studies on the frequency of myelin-reactive T cells in MS patients are conflicting; some investigators report a significantly higher incidence of myelin-specific peripheral blood mononuclear cells (PBMCs) in individuals with MS compared with age- and gender-matched healthy controls (HC), whereas others report no significant difference.[5] Many of the earlier studies that found no differences between patients and HC used proliferation as a measure of T-cell reactivity and nonhuman myelin proteins for antigenic stimulation. In contrast, several laboratories have found that untreated RRMS patients have increased frequencies of PBMC that produce IFN-γ or IL-17 in response to ex vivo challenge with human myelin basic protein (MBP), human proteolipid protein (PLP), or their constituent peptides.[4,34] IFN-γ

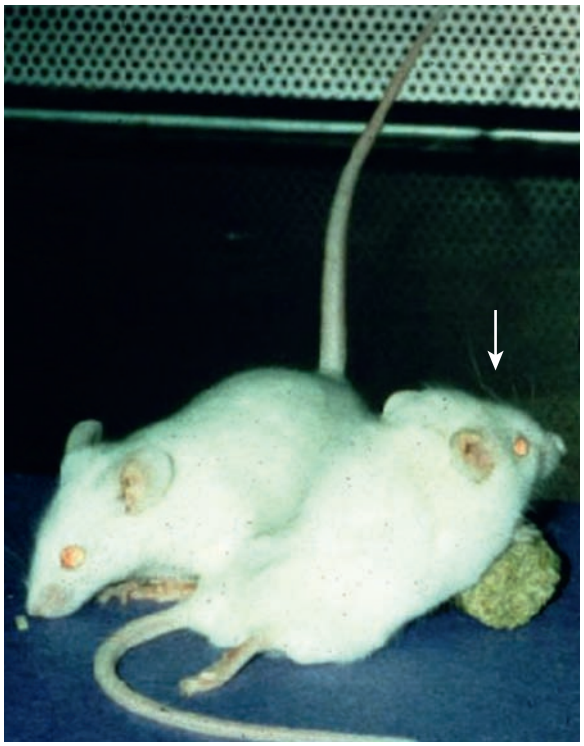


FIG. 66.3 A mouse with experimental autoimmune encephalomyelitis (EAE) *(arrow)* and a healthy litter mate. The mouse with EAE has a limp tail and hind limb weakness.

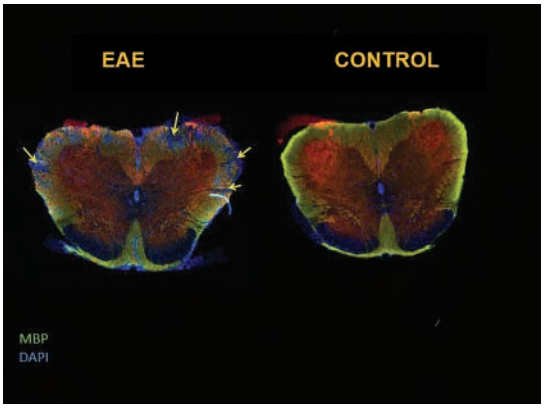


FIG. 66.4 Immunofluorescent histology of spinal cord sections from a mouse with experimental autoimmune encephalomyelitis *(EAE) (left)* and a healthy control *(right)*. White matter tracks were stained with a monoclonal antibody specific for myelin basic protein (*green*). The nuclei of inflammatory cells are stained with 4′,6-diamidino-2-phenylindole *(DAPI)* (*blue*). The arrows point to areas of demyelination. *MBP*, Myelin basic protein.

responses to PLP peptides correlated with level of clinical disability.[34] T cells that coexpress IL-17 and IFN-γ T cells were identified in brain tissue of MS patients, and circulating lymphocytes obtained from MS patients were found to have an increased propensity to differentiate into IL-17/IFN-γ double producers.[35] IL-17/IFN-γ coexpressing T cells, and IL-17–producing T cells that convert into IFN-γ producers (so-called ex-Th17 cells), have been implicated in the pathogenesis of some EAE models.

There is an increasing appreciation that the cytokine dysregulation that occurs during EAE and MS is heterogeneous. Hence clinically identical forms of EAE can be induced with stable murine Th1 or Th17 cells independently.[36] Subsets of MS patients were recently identified that consistently mounted either IFN-γ– or IL-17–skewed responses to human MBP over the course of a year, while others exhibited mixed or oscillating responses.[33] Diversity in the immune pathways that drive MS pathology holds important implications for the development of a more personalized approach to the management of patients in the future. Indeed, Th1- and Th17-mediated forms of EAE show different patterns of responsiveness to the same immunomodulatory drugs.[36]

Perhaps the strongest evidence of an autoimmune basis of inflammatory demyelination in humans comes from clinical trials of immunomodulatory agents. RRMS patients were treated with an altered peptide ligand (APL) of MBP with the intention of tolerizing MBP-reactive T cells or deflecting their differentiation towards an immunosuppressive, regulatory, or an innocuous Th2 phenotype. Unexpectedly, administration of the APL was temporally associated with expansion of circulating MBP-reactive Th1 cells and clinical worsening in a subgroup of patients.[37] In contrast, as will be discussed in detail later, drugs that impede lymphocyte trafficking to the CNS or deplete lymphocytes from the periphery reduce MS relapse rates and the accumulation of MRI lesions.

DISEASE-MODIFYING THERAPIES

Prior to 1993, corticosteroids were the only class of drugs routinely used to treat MS. Corticosteroids accelerate the rate of recovery from acute exacerbations, but there is little evidence they alter ultimate clinical outcomes or prevent subsequent disease activity. The US Food and Drug Administration (FDA) approved IFN-β-1b (Betaseron) in 1993 for the management of RRMS, beginning a new era in MS therapeutics. A host of additional DMT have since been approved, all of which significantly reduce the annualized relapse rate (ranging from 25% to perhaps 80%), and the frequency of gadolinium-enhancing (acutely inflamed) MRI lesions. The introduction of DMT has represented a major advance in the treatment of individuals with RRMS and has had a profound impact by mitigating morbidity and enhancing quality of life.

Recombinant Interferon Beta

IFN-β is a type I IFN with potent antiviral properties and pleiotropic effects on the innate immune system. Recombinant IFN-β was first trialed in MS based on the contemporaneous theory that the disease was caused by an active viral infection of the CNS. Recombinant IFN-β therapy was serendipitously found to significantly reduce annualized MS relapse rates. It is manufactured as five different commercial products, all of which are self-administered by either subcutaneous or intramuscular injection. There are two distinct structural forms. IFN-β-1a is produced in mammalian cells and has the same sequence as the naturally occurring compound, whereas IFN-β-1b is produced in modified *Escherichia coli* and has a Met-1 deletion and a Cys-17 to Ser mutation. IFN-β-1a is glycosylated, and IFN-β-1b is nonglycosylated. Studies of the efficacy of IFN-β in RRMS have yielded remarkably consistent results across different formulations of the drug. In multiple randomized, double-blinded placebo-controlled trials, IFN-β therapy reduced annual relapse rates by 20% to 35%.[38] However, approximately 30% of MS patients do not respond. Recent trials of novel DMT have used IFN-β as an active comparator. The mechanism of action of IFN-β in MS has

not been definitively elucidated. Proposed mechanisms include induction of the immunosuppressive cytokine IL-10, inhibition of pathogenic Th17 cells, and stabilization of the BBB via direct effects on cerebrovascular endothelium.

Glatiramer Acetate

GA is composed of a mixture of polypeptides of different lengths and sequences, synthesized by the randomized polymerization of four amino acids (glutamic acid, lysine, alanine, and tyrosine), and is administered subcutaneously. These are the most prevalent amino acids in MBP, a candidate autoantigen in MS. It was originally thought that GA would act as a competitive antagonist of MBP peptides for binding to MHC class II molecules on APC. Subsequent mechanistic studies did not support that theory. Alternative hypotheses include immune deviation of myelin-reactive T cells from a destructive Th1 to an innocuous Th2 phenotype, increase in frequency and function of $FOXP3^+$ regulatory CD4 T cells or regulatory CD8 T cells, and induction of antiinflammatory type II monocytes. A pivotal trial of GA versus placebo in 1995 demonstrated a mean reduction in the relapse rate of approximately 30%.[38]

Teriflunomide

Teriflunomide is an oral pyrimidine synthesis inhibitor approved for relapsing forms of MS. It is the active metabolite of leflunomide, which has been used for years in the treatment of rheumatoid arthritis. Teriflunomide has broad immunosuppressive effects, including a cytostatic effect on proliferating T and B lymphocytes. In a pivotal phase 3 trial, teriflunomide reduced annualized relapse rate by 31% compared with placebo.[38]

Dimethyl Fumarate and Diroximel Fumarate

Dimethyl fumarate (DMF) is the methyl ester of fumaric acid. Prior to being tested as a DMT in MS, DMF was used as a biocide in furniture and shoes to prevent the growth of molds during storage or transport. A combination of DMF and three other fumaric acid esters was marketed in Germany as a treatment for psoriasis. In two phase 3 trials, oral DMF was demonstrated to decrease the annualized relapse rate of adults with RRMS by approximately 34% to 50% compared with placebo.[39] DMF therapy is often associated with lymphopenia, but therapeutic efficacy does not appear to be inversely related to lymphocyte counts. In the postmarketing setting, several cases of progressive multifocal leukoencephalopathy (PML), a rare viral infection of the brain caused by the JC virus, were reported in patients taking DMF who had persistent lymphopenia. The mechanism by which DMF suppresses MS disease activity remains unclear.

The most common side effects of DMF are flushing and gastrointestinal upset. Diroximel fumarate, a novel oral fumarate that causes less gastrointestinal adverse effects than DMF, was recently FDA approved for the treatment of relapsing forms of MS. DRF undergoes rapid cleavage in the gut to monomethylfumarate, the same pharmacologically active metabolite as DMF.

Fingolimod and Siponimod

During homeostasis, sphingosine-1-phosphate (S1P)-1-signaling drives lymphocyte egress from lymph nodes into the bloodstream. Fingolimod is an orally administered S1P1 receptor modulator that effectively traps myelin-specific T cells in lymph nodes so that they cannot reenter the circulation and gain access to the CNS. In a 24-month randomized, double-blinded placebo-controlled trial, fingolimod was shown to reduce annualized relapse rates by 50%.[38] In a phase 3 randomized, double-blind, placebo-controlled trial, fingolimod did not slow disease progression in patients with PPMS. Fingolimod does not distinguish between pathogenic and protective lymphocytes, and a number of opportunistic infections have emerged as complications of the treatment. Patients taking fingolimod are susceptible to herpes virus infections, particularly shingles. Consequently, patients are screened for immunity to varicella zoster before commencing treatment. In the postmarketing setting, PML and cryptococcal meningitis have been reported in fingolimod-treated patients with no, or only distant, prior exposure to immunosuppressive drugs. Other potential side effects of fingolimod include macular edema, bradycardia, and atrioventricular block, which may be secondary to cross-binding of the drug to the S1P1, S1P2, and S1P3 receptors in the cardiovascular system and on retinal endothelial cells.

Siponimod is a selective S1P receptor modulator (S1P1 and S1P5); this selectivity confers low risk of bradycardia and atrioventricular block when the drug is appropriately titrated, while still reducing the risk of relapse and new lesion accumulation in patients with RRMS. In a phase 3 double-blind, placebo-controlled trial in SPMS, siponimod was found to significantly reduce risk of confirmed disability progression at 3 and 6 months.[9] Secondary analyses suggested that patients who experienced relapses in the 2 years prior to enrollment were preferentially responsive to siponimod. Siponimod is labeled for use in relapsing forms of MS, including active SPMS.

Cladribine

Cladribine is a synthetic purine analogue that preferentially depletes peripheral B and T lymphocytes.[38] This preferential effect on B and T cells owes to relatively higher levels of deoxycytidine kinase, an enzyme necessary for conversion of the prodrug to its active form, in these cells. Intravenous cladribine is used for treatment of hairy cell leukemia. Oral cladribine is approved for treatment of RRMS and active SPMS; dosing is weight based and consists of repeated short courses. This dosing strategy intentionally induces lymphopenia and then allows for immune reconstitution. In a phase 3 placebo-controlled trial, oral cladribine reduced annualized relapse rate by approximately 58%; nearly 80% of patients remained relapse free at 96 weeks. Potential adverse effects include severe lymphopenia, infection (specifically herpes zoster), teratogenicity, and malignancy.

Natalizumab

Natalizumab is a humanized monoclonal antibody against the cell adhesion molecule α_4-integrin, which is widely expressed on lymphocytes and monocytes. Natalizumab is believed to mediate its ameliorative effects in RRMS by blocking interactions between the very late antigen (VLA)-4 (a heterodimer composed of the α_4-and β_1-integrin chains) on leukocytes with its cognate ligand, vascular cell adhesion molecule (VCAM)-1, on cerebrovascular endothelial cells. VLA-4/VCAM-1 interactions are required for the passage of lymphocytes and monocytes past the BBB. In a phase 3 placebo-controlled trial, natalizumab reduced the relapse rate at 1 year by 68% and the number of gadolinium-enhancing MRI lesions at both 1 and 2 years by more than 90%.[39] The most serious complication of natalizumab treatment is PML; there have been more than 700 reported cases. Factors that increase the risk of natalizumab-associated PML include JC virus seropositivity, duration of treatment more than 2 years, and prior exposure to immunosuppressive drugs (e.g., cyclophosphamide, azathioprine, or mycophenolate mofetil).

Alemtuzumab

Alemtuzumab is a humanized anti-CD52 monoclonal antibody that globally depletes circulating T and B lymphocytes via antibody-dependent cell-mediated cytotoxicity, complement-dependent cytolysis, and induction of apoptosis. CD52 is primarily expressed on the cell surface of T and B lymphocytes, but it is also expressed at lower levels on macrophages, eosinophils, and natural killer (NK) cells. Hematopoietic stem cells do not express CD52, allowing the subsequent repopulation of circulating lymphocyte pools following alemtuzumab treatment, which occurs at variable rates between lymphocyte subsets. Reconstituted circulating T-cell pools have been found to be enriched in $CD4^+CD25^{high}CD127^{low}FOXP3^+$ regulatory T cells that might be responsible, in part, for the long-lasting therapeutic effect of alemtuzumab, which has been observed in some individuals. In a 2-year, rater-masked, randomized controlled phase 3 trial, previously untreated RRMS patients were randomly allocated to receive intravenous alemtuzumab or subcutaneous IFN-β-1a.[38] Alemtuzumab was given once per day for 5 days at baseline and once per day for 3 days at 12 months. Relapse rates were reduced 54.9% in the alemtuzumab group compared with the IFN-β-1a group. In an independent study, alemtuzumab reduced relapse rates and the risk of sustained accumulation of disability in RRMS patients refractory to first-line DMT.[38] Although alemtuzumab suppressed new MRI lesion formation and superimposed relapses in patients with SPMS, it did not prevent clinical progression or progressive cerebral atrophy.[7] Extension studies of alemtuzumab indicate that a substantial portion of patients treated with only two cycles of alemtuzumab have durable treatment effect out to 5 years.

The principal adverse effect of alemtuzumab in MS is antibody-mediated autoimmune disease, most commonly Graves disease. Idiopathic thrombocytopenic purpura, Goodpasture syndrome, and antiglomerular basement membrane disease have been reported less frequently. The underlying mechanism is not fully understood but may be a consequence of early recovery of immature B cells in the absence of regulatory T cells. Another possibility is that homeostatic T-cell proliferation following lymphocyte ablation leads to the generation of chronically activated oligoclonal CD4 and CD8 T cells capable of producing proinflammatory cytokines. Alemtuzumab therapy has been associated with opportunistic infections, including listeriosis.

B Cell–Depleting Monoclonal Antibodies

A role of B cells in the pathogenesis of MS has been suspected since the discovery of unique oligoclonal bands in the CSF of the majority of individuals with MS. Nonetheless, B cells are not a prominent component of the perivascular infiltrates in MS lesions. This apparent paradox has been resolved, at least in part, by the discovery of lymphoid follicle-like structures, in the meninges of some individuals with MS (as previously discussed). The most direct evidence for a role of B cells as effector cells in MS pathogenesis comes from trials of monoclonal antibodies directed against CD20, a B-cell surface molecule. CD20 is expressed on pre-B cells and mature B cells and not on antibody-secreting plasma cells. Rituximab, a chimeric monoclonal anti-CD20 antibody, induces a rapid depletion of circulating B cells lasting approximately 6 to 9 months. In a phase 2 clinical trial of rituximab in RRMS, subjects in the active treatment arm experienced a significant reduction in clinical relapses and in the formation of new or enhancing MRI lesions when compared with subjects in the placebo arm. Ocrelizumab is a next-generation fully human recombinant anti-CD20 monoclonal antibody that binds to a different epitope from rituximab, and with higher affinity. In two large phase 3 studies, ocrelizumab reduced the annualized relapse rate in subjects with RRMS by nearly 50% compared with IFN-β-1a over a 2-year period.[38] In addition, ocrelizumab delayed confirmed disability progression by approximately 40% and the total number of gadolinium-enhancing lesions by more than 90% compared with IFNβ-1a.

Interestingly, some patients with progressive MS might also benefit from the depletion of $CD20^+$ cells, particularly if there is evidence of ongoing neuroinflammatory activity at the time of treatment initiation. Rituximab significantly delayed the time to confirmed disease progression in a subset of PPMS patients who had gadolinium-enhancing lesions on baseline MRI of the brain. The benefit was most dramatic in subjects aged younger than 51 years. In a placebo-controlled trial of ocrelizumab in individuals with PPMS between 18 and 55 years of age, ocrelizumab was associated with lower rates of clinical and MRI progression.[10] The FDA granted ocrelizumab "breakthrough therapy designation" for PPMS.

The mechanism by which B cell–depleting therapy ameliorates MS does not appear to be reduction in antibody titers. Rituximab depletes B cells from the CSF of RRMS patients with little effect on CSF IgG levels. B cells are APC, and there is some evidence that they are important for sustaining myelin-specific Th17 responses in MS. The frequency of GM-CSF–producing B cells was found to be elevated in RRMS.[40] Hence ocrelizumab and rituximab may eliminate an important cellular source of GM-CSF, a monocyte/macrophage mobilizing cytokine that has been strongly implicated in the pathogenesis of EAE. Disruption of meningeal follicle-like structures is yet another putative mechanism of action that might be particularly relevant to progressive forms of MS.

THERAPEUTIC PRINCIPLES

- Corticosteroids accelerate the rate of recovery from multiple sclerosis (MS) relapses, but there is little evidence that they impact the ultimate degree of recovery or the future clinical course.
- There are a growing number of US Food and Drug Administration (FDA)-approved disease-modifying therapies (DMTs) that reduce annualized relapse rates in patients with relapsing-remitting MS by approximately 20% to 80%
- Newer-generation DMTs either deplete lymphocytes, inhibit their expansion, or block their migration to the central nervous system.
- DMTs differ in efficacy and safety profiles. Opportunistic infection, in particular progressive multifocal leukoencephalopathy, has been observed in association with several DMT.
- The choice of DMT must be customized for on an individual basis, taking into account disease activity and risk tolerance.

FUTURE DIRECTIONS

Dramatic advances have been made in the management of RRMS, but much remains to be done. The DMTs currently in use have global effects on lymphocytes as opposed to specifically targeting autoreactive Th1 and Th17 cells. Consequently, protective as well as pathogenic immune responses may be undermined, increasing the risk of infection. Furthermore, DMTs are only modestly

effective in progressive forms of disease. Critical goals of the MS research community currently are to better define the cellular and molecular mechanisms that link neuroinflammation to end-organ injury (namely demyelination and axonopathy) and to elucidate the pathogenic pathways that underlie clinical progression. It is also imperative to develop a deeper understanding of endogenous obstacles to repair pathways in the CNS. This knowledge could ultimately lead to the discovery of laboratory-based surrogate biomarkers of both acute and chronic disease activity, as well as drugs that block, or even slow, disability accumulation in progressive MS and promote remyelination and axonal regeneration across clinical subtypes.

ON THE HORIZON

- The introduction of disease-modifying therapies has represented a major advance in the treatment of individuals with relapsing-remitting multiple sclerosis (MS). However, these drugs have global effects on lymphocytes, thereby increasing the risk of infection. A future goal will be to develop drugs that specifically target pathogenic leukocytes and/or autoreactive lymphocytes.
- Disease-modifying therapy is currently focused on the modulation of lymphocytes. Drugs that target innate immune cells may enhance MS therapeutics in the future.
- There is still a dire need for drugs that slow, or even halt, disability accumulation in progressive forms of MS and that are effective in older patients without evidence of recent neuroinflammatory activity.
- An increased understanding of the mechanisms by which inflammatory cells inflict central nervous system damage, and the endogenous obstacles to repair pathways in MS, may inform the development of neuroprotective and neuroregenerative agents.

REFERENCES

1. Koch M, Kingwell E, Rieckmann P, Tremlett H, UBC MS. Clinic Neurologists. The natural history of secondary progressive multiple sclerosis. *J Neurol Neurosurg Psychiatry*. 2010;81(9):1039–1043.
2. University of California San Francisco MS-EPIC Team Cree B, Gourraud PA, Oksenberg JR, et al. Long-term evolution of multiple sclerosis disability in the treatment era. *Ann Neurol*. 2016;80(4):499–510.
3. Mahad DH, Trapp BD, Lassmann H. Pathological mechanisms in progressive multiple sclerosis. *Lancet Neurol*. 2015;14(2):183–193.
4. Huber AK, Wang L, Han P, et al. Dysregulation of the IL-23/IL-17 axis and myeloid factors in secondary progressive MS. *Neurology*. 2014;83(17):1500–1507.
5. Segal BM. Stage-specific immune dysregulation in multiple sclerosis. *J Interferon Cytokine Res*. 2014;34(8):633–640.
6. Magliozzi R, Howell O, Vora A, et al. Meningeal B-cell follicles in secondary progressive multiple sclerosis associate with early onset of disease and severe cortical pathology. *Brain*. 2007;130(Pt 4):1089–1104.
7. Coles AJ, Cox A, Le Page E, et al. The window of therapeutic opportunity in multiple sclerosis: evidence from monoclonal antibody therapy. *J Neurol*. 2006;253(1):98–108.
8. Yildiz O, Mao Z, Adams A, et al. Disease activity in progressive multiple sclerosis can be effectively reduced by cladribine. *Mult Scler Relat Disord*. 2018;24:20–27.
9. Kappos L. Siponimod versus placebo in secondary progressive multiple sclerosis (EXPAND): a double-blind, randomised, phase 3 study. *Lancet*. 2018;391(10127):1263–1273.
10. Montalban X, Hauser SL, Kappos L, et al. Ocrelizumab versus placebo in primary progressive multiple sclerosis. *N Engl J Med*. 2017;376(3):209–220.
11. Thompson AJ, Banwell BL, Barkhof F, et al. Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. *Lancet Neurol*. 2018;17(2):162–173.
12. Okuda DT, Siva A, Kantarci O, et al. Radiologically isolated syndrome: 5-year risk for an initial clinical event. *PLoS One*. 2014;9(3):e90509.
13. Hollenbach JA, Oksenberg JR. The immunogenetics of multiple sclerosis: a comprehensive review. *J Autoimmun*. 2015;64:13–25.
14. Olsson T, Barcellos LF, Alfredsson L. Interactions between genetic, lifestyle and environmental risk factors for multiple sclerosis. *Nat Rev Neurol*. 2017;13(1):25–36.
15. Kurtzke JF. An epidemiologic approach to multiple sclerosis. *Arch Neurol*. 1966;14(2):213–222.
16. Sintzel MB, Rametta M, Reder AT. Vitamin D and multiple sclerosis: a comprehensive review. *Neurol Ther*. 2018;7(1):59–85.
17. Ascherio A, Munger KL. Epstein-Barr virus infection and multiple sclerosis: a review. *J Neuroimmune Pharmacol*. 2010;5(3):271–277.
18. Lang HL, Jacobsen H, Ikemizu S, et al. A functional and structural basis for TCR cross-reactivity in multiple sclerosis. *Nat Immunol*. 2002;3(10):940–943.
19. Magliozzi R, Serafini B, Rosicarelli B, et al. B-cell enrichment and Epstein-Barr virus infection in inflammatory cortical lesions in secondary progressive multiple sclerosis. *J Neuropathol Exp Neurol*. 2013;72(1):29–41.
20. Wu Z, Nagano I, Asano K, et al. Infection of non-encapsulated species of *Trichinella* ameliorates experimental autoimmune encephalomyelitis involving suppression of Th17 and Th1 response. *Parasitol Res*. 2010;107(5):1173–1188.
21. Correale J. Helminth/parasite treatment of multiple sclerosis. *Curr Treat Options Neurol*. 2014;16(6):296.
22. Mokry LE, Ross S, Timpson NJ, et al. Obesity and multiple sclerosis: a mendelian randomization study. *PLoS Med*. 2016;13(6):e1002053.
23. Jalal P, Masoud B, Manoochehr K, Elham H. Effect of smoking on multiple sclerosis: a meta-analysis. *Journal of Public Health*. 2017;39(2):312–320.
24. Degelman ML, Herman KM. Smoking and multiple sclerosis: a systematic review and meta-analysis using the Bradford Hill criteria for causation. *Mult Scler Relat Disord*. 2017;17:207–216.
25. Ysrraelit MC, Correale J. Impact of sex hormones on immune function and multiple sclerosis development. *Immunology*. 2019;156(1):9–22.
26. Voskuhl RR, Wang H, Wu TC, et al. Estriol combined with glatiramer acetate for women with relapsing-remitting multiple sclerosis: a randomised, placebo-controlled, phase 2 trial. *Lancet Neurol*. 2016;15(1):35–46.
27. Kurth F, Luders E, Sicotte NL, et al. Neuroprotective effects of testosterone treatment in men with multiple sclerosis. *Neuroimage Clin*. 2014;4:454–460.
28. Berer K, Mues M, Koutrolos M, et al. Commensal microbiota and myelin autoantigen cooperate to trigger autoimmune demyelination. *Nature*. 2011;479(7374):538–541.
29. Lee YK, Menezes JS, Umesaki Y, et al. Proinflammatory T-cell responses to gut microbiota promote experimental autoimmune encephalomyelitis. *Proc Natl Acad Sci U S A*. 2011;108(Suppl 1):4615–4622.
30. Mirza A, Forbes JD, Zhu F, et al. The multiple sclerosis gut microbiota: a systematic review. *Mult Scler Relat Disord*. 2020;37:101427.
31. Nikic I, Merkler D, Sorbara C, et al. A reversible form of axon damage in experimental autoimmune encephalomyelitis and multiple sclerosis. *Nat Med*. 2011;17(4):495–499.
32. Preziosa P, Rocca MA, Pagani E, et al. Structural MRI correlates of cognitive impairment in patients with multiple sclerosis: a multicenter study. *Hum Brain Mapp*. 2016;37(4):1627–1644.
33. Carbajal KS, Mironova Y, Ulrich-Lewis JT, et al. Th cell diversity in experimental autoimmune encephalomyelitis and multiple sclerosis. *J Immunol*. 2015;195(6):2552–2559.
34. Moldovan IR, Rudick RA, Cotleur AC, et al. Interferon gamma responses to myelin peptides in multiple sclerosis correlate with a new clinical measure of disease progression. *J Neuroimmunol*. 2003;141(1–2):132–140.
35. Kebir H, Ifergan I, Alvarez JI, et al. Preferential recruitment of interferon-gamma-expressing TH17 cells in multiple sclerosis. *Ann Neurol*. 2009;66(3):390–402.

36. Kroenke MA, Carlson TJ, Andjelkovic AV, et al. IL-12- and IL-23-modulated T cells induce distinct types of EAE based on histology, CNS chemokine profile, and response to cytokine inhibition. *J Exp Med*. 2008;205(7):1535–1541.
37. Bielekova B, Goodwin B, Richert N, et al. Encephalitogenic potential of the myelin basic protein peptide (amino acids 83–99) in multiple sclerosis: results of a phase II clinical trial with an altered peptide ligand. *Nat Med*. 2000;6(10):1167–1175.
38. Tintore M, Vidal-Jordana A, Sastre-Garriga J. Treatment of multiple sclerosis—success from bench to bedside. *Nat Rev Neurol*. 2019;15(1):53–58.
39. Polman CH, O'Connor PW, Havrdova E, et al. A randomized, placebo-controlled trial of natalizumab for relapsing multiple sclerosis. *N Engl J Med*. 2006;354(9):899–910.
40. Li R, Rezk A, Miyazaki Y, et al. Proinflammatory GM-CSF-producing B cells in multiple sclerosis and B cell depletion therapy. *Sci Transl Med*. 2015;7(310): 310ra166.

67

Autoimmune Peripheral Neuropathies

Marinos C. Dalakas

Autoimmune peripheral neuropathies (APNs) occur when immunological tolerance to peripheral nerve components (myelin, Schwann cell, axon, and motor or ganglionic neurons) is lost. In some neuropathies, there are pathogenic antibodies or cytotoxic T lymphocytes against peripheral nerve antigens, but in others the immune-mediated mechanisms are undefined or suspected when the neuropathy coexists with another systemic autoimmune disease or viral infection.

This chapter reviews the most common autoimmune neuropathies (Table 67.1), their clinical features, diagnostic criteria, the prevailing autoimmune phenomena, and effective therapies.

ACUTE INFLAMMATORY POLYNEUROPATHY: GUILLAIN-BARRÉ SYNDROME(S)

Guillain-Barré syndrome (GBS) is characterized by acute (within 1 week) or subacute (within 4 weeks) ascending motor weakness, mild or moderate sensory abnormalities, occasional cranial nerve involvement, and muscle or radicular pain.[1–4] Tendon reflexes are reduced but can be normal, especially in axonal GBS. It is a disease of all ages that occurs sporadically with an incidence of 0.8 to 1.9 (median 1.1) per 100,000.[4] In typical cases, the disease peaks by the fourth week, a sign conventionally used to separate GBS from chronic inflammatory demyelinating polyneuropathy (CIDP), in which the disease begins slowly and usually reaches a nadir after at least 2 months. There are, however, patients with CIDP (perhaps up to 16%) with subacute onset and monophasic course falling between the two timeframes and still others with an even more acute onset, reaching a nadir within 6 to 8 weeks resembling GBS.[1–4] Distinguishing GBS from acute-onset CIDP can be at times challenging and becomes clear in retrospect, although proposed criteria may help separate the two early in the disease course.[1–4]

After reaching the clinical peak, there is a recovery period that varies from weeks to more than a year according to disease subtype or severity at onset; rapid initial progression over less than 7 days, quadriparesis, need for mechanical ventilation, age greater than 60 years, and preceding diarrheal illness are signs associated with incomplete recovery or worse outcome. GBS is not one but several syndromes, reflecting the varying degree of involvement of the motor or sensory nerve fibers and the myelin sheath or the axon. The GBS subtypes, or GBS variants,[1–4] include:

- *Acute inflammatory demyelinating polyneuropathy (AIDP)*, which accounts for the majority (probably 80%) of patients. In classic cases, the weakness starts from the legs and spreads up to the arms, intercostal and diaphragmatic muscles, and facial or bulbar muscles. At times, the weakness may be limited to one or two limbs or to cranial nerves. Patients need to be monitored for impending respiratory failure: hence the need for early admission to intensive care units (ICUs). Autonomic dysfunction occurs in up to 65% of patients and may cause cardiac arrhythmias or hemodynamic changes.[1–4]
- *Acute motor axonal neuropathy* (*AMAN*), which exhibits primary axonal damage caused by massive acute demyelination and inflammation, as occurs in experimental allergic neuritis (EAN) when animals are immunized with a high dose of myelin antigen,[1–4] or by a primary axonal event mediated by macrophages. These patients have a fulminant course with paralysis and electrical inexcitability of motor nerves as early as 3 to 5 days after onset.[1–4] In contrast to AIDP, involvement of cranial nerves is infrequent, and reflexes are normal or increased, especially early in the disease. AMAN is common in Asia and Central South America, accounting for 30% to 65% of all GBS cases in these regions.[1–4] Recovery is variable; some patients recover within days as a result of resolution of conduction block, but others have slow or poor recovery because of excessive axonal degeneration.[1–4] Infection with *Campylobacter jejuni* appears to trigger many of these cases.[1–4] A number of patients also have high levels of antiganglioside GM1 antibodies.[1–4]
- *Acute motor–sensory axonal neuropathy (AMSAN)* is similar to AMAN, but with concurrent involvement of the sensory axons, and has a pathomechanism similar to AMAN, including frequent antibodies against GM1 and GD1a gangliosides.
- *Miller Fisher syndrome (MFS)* is characterized by acute onset of ophthalmoplegia, gait ataxia, normal sensation, and areflexia.[1–6] In some patients pharyngeal, facial, trunk, and respiratory muscles are involved; rarely it can present as ocular nerve palsies.[1–5] MFS is also distinct because of the presence of a unique immunoglobulin G (IgG) antibody against GQ1b ganglioside.[1–5]
- *Sensory ataxic GBS* results from the involvement of roots and ganglionic neurons. Some of these patients have antibodies to GD1b ganglioside, probably forming a continuum with MFS because they share autoantibodies with the same sialic groups.[2,3,5]
- *Acute pandysautonomic neuropathy* affects ganglionic neurons and causes pure autonomic dysfunction.[1–4]

Diagnosis

The diagnosis, often suspected on clinical grounds, is confirmed with elevated cerebrospinal fluid (CSF) protein and electrophysiological studies consistent with active demyelination or nerve inexcitability.

CSF protein may be normal in the early phase of the disease, but it can be as high as 1000 mg/dL by the sixth week. The elevated CSF protein is probably related to root inflammation but, as the blood–nerve barrier becomes impaired, serum albumin and IgG

TABLE 67.1 Common Autoimmune Neuropathies

- Guillain-Barré syndrome(s) (GBS)
- Chronic inflammatory demyelinating polyneuropathy (CIDP) and its variants
- Polyneuropathy associated with IgM monoclonal gammopathy
- Multifocal motor neuropathy with conduction block
- Polyneuropathy, organomegaly, endocrinopathy, myeloma, and skin changes (POEMS) syndrome
- Cryoglobulinemic polyneuropathy
- Paraneoplastic neuropathies associated with anti-Hu or CRMP-5 antibodies
- Autoimmune autonomic neuropathies
- Vasculitic neuropathies
- Possibly autoimmune small fiber sensory neuropathy with neuropathic pain
- Infectious neuropathies (human immunodeficiency virus [HIV], cytomegalovirus [CMV], Epstein-Barr virus [EBV], and herpes virus infections; Lyme disease; leprosy; Chagas disease; diphtheria; others)

may enter freely into CSF, contributing further to protein elevation. CSF cell count is normal (or slightly increased <50 cells per microliter); there is, however, lymphocytosis when GBS occurs in conjunction with viral infections, such as human immunodeficiency virus (HIV), cytomegalovirus (CMV), Epstein-Barr virus (EBV), or Lyme disease. When CSF protein is very high, papilledema can develop because of impaired reabsorption of CSF and raised intracranial pressure. Oligoclonal IgG bands can be seen. Nerve conduction studies can be normal early in the disease but remain helpful to distinguish AIDP from AMAN or AMSAN. They also have prognostic value because features of demyelination suggest higher chances of needing mechanical ventilation, whereas severe axonal loss from the outset predicts poor outcome.[1–4] Differential diagnosis of GBS should include other forms of acute flaccid paralysis, such as brainstem stroke; brainstem encephalitis; acute motor neuron involvement caused by poliomyelitis or West Nile virus infection; acute myelopathy; disorders of neuromuscular transmissions, such as myasthenia gravis or botulism; acute inflammatory or necrotizing myopathies; periodic paralysis; and rare causes of acute neuropathy, such as porphyria, toxins, and vasculitis. Also important to consider is the critical illness polyneuropathy.

Antecedent Events

Two-thirds of GBS patients give a history of a flu-like illness or acute dysenteric episodes that precede the development of GBS by 1 to 3 weeks.[1–4,6–8] Among the implicated viruses are CMV, EBV, herpesvirus, hepatitis A, HIV, Zika virus, and coronavirus (COVID-19). Among bacteria, infection with *Mycoplasma pneumoniae* and, most importantly, *C. jejuni* may be present in greater than 25% of the patients and in some parts of the world up to 50%.[1–4] *Campylobacter* is of special interest because it contains glycoconjugates that share epitopes with the peripheral myelin, as discussed later. Two vaccines—one against rabies and the other against the swine flu A/New Jersey influenza strain that caused an outbreak of GBS in 1976[1–4]—have been associated with GBS. Rabies vaccine that contains brain material is followed by GBS in about 1 in 1000 cases. Despite anecdotal reports, there is no convincing evidence that the incidence of GBS is increased in association with other vaccines. Recent viral epidemics or pandemics have been triggering factors for GBS. Zika virus, an arbovirus in the family of Flaviviridae, which emerged in South America countries, is estimated to cause GBS in 0.24 per 1000 Zika infections.[6] Most of Zika-associated GBS have AMAN with rapid disease evolution—within 4 to 6 days—and antiglycolipid antibodies in 31% of the patients.[6] COVID-19 can also trigger GBS based on many well documented series.[7,8] The most characteristic finding of COVID-19–triggered GBS were preceding, or concomitant, signs of anosmia and ageusia and frequent presence of cranial neuropathies including ophthalmoplegias.[7,8] Antibodies to GD1b gangliosides have been seen in two patients and the possibility of molecular mimicry was raised because the COVID-19 spike protein also contains ganglioside moieties.[7,8] Surgery can precede the development of GBS in some patients;[1–4] surgical stress, release of nerve autoantigens, or infections have been implicated. Three drugs—gold, perhexiline, and suramin at high doses—have been causally associated with acute demyelinating neuropathy.[2,3] GBS has occurred in patients who suffer from neoplasms, especially lymphoma, melanoma, and Hodgkin disease and rarely as a complication of immune checkpoint inhibitors.[1–4,9] Interestingly, GBS is rarely seen as part of another connective tissue disorder.

Immunopathology of Guillain-Barré Syndrome

GBS is an inflammatory demyelinating polyneuropathy in which an immune attack, triggered by various antecedent events, is directed against peripheral nerve antigens including myelin, axon, or nodes of Ranvier. Both cellular and humoral immune components have been implicated.[1–4]

Cellular Factors

Endoneurial inflammatory infiltrates throughout the nerves, roots, or plexuses[1–4] and segmental demyelination in areas associated with the lymphoid infiltrates, especially macrophages, are prominent in typical GBS. Macrophages, which are the most prominent cells in contact with nerve fibers, break through the basement membrane of healthy Schwann cells and make direct contact with the outermost myelin lamellae, leading to lysis of the myelin sheath (macrophage-mediated demyelination). Cytokines and chemokines released by the activated T cells or complement activation may increase capillary permeability and facilitate transmigration of additional macrophages or T cells. When the demyelination is extensive or chronic, it is followed by axonal degeneration.[1–4] The effectiveness of remyelination and degree of axonal regeneration dictate the prospects of clinical recovery.

The T-cell mediated process in GBS is mostly derived by analogy to the animal model of EAN, which resembles GBS in pathology and clinical course.[1–4] Animals sensitized to whole human nerve or various myelin proteins, such as P0, P2, and the neutral glycolipid galactocerebroside, develop segmental demyelination with macrophages and T-cell infiltrates. In EAN, T cells are sensitized against myelin and can passively transfer the disease to healthy animals. Interleukin-2 (IL-2) and soluble IL-2 receptors are increased during the acute phase, suggesting T-cell activation. Furthermore, lymphocytes from GBS patients exert myelinotoxic activity when applied to cultures of myelinated axons.

Humoral Factors and Antiganglioside Antibodies

There is much stronger evidence that circulating serum factors are responsible for GBS. On clinical grounds, this is supported

by the beneficial effect of plasmapheresis, presumably by removing putative antibodies. On laboratory grounds, it is supported by the variety of autoantibodies detected in patients' sera. Serum from the acute phase of GBS can demyelinate rodent dorsal root ganglionic extracts in a complement-dependent manner. Furthermore, GBS serum injected into rat sciatic nerves causes demyelination and conduction block. Complement-fixing IgM antibodies against a human peripheral nerve myelin glycolipid that contains carbohydrate epitopes, as well as high-titer antibodies against various sulfated or acidic glycosphingolipids, are present in several patients with GBS.[1–5]

Gangliosides are present in all tissues but are especially abundant in the nervous system. Their lipid portion lies in the cell membrane, and their signature sugar residues are exposed at the extracellular surface bearing one or more sialic acid molecules, such as one sialic acid ganglioside (GM1), two (GD1a), three (GT1a), or four (GQ1b).[1–5] Although they do not form a common "GBS antigen," different gangliosides are involved in different GBS subtypes. They are of pathogenic relevance because immunization of rabbits with GM1 and GD1b induces acute neuropathy with histological features of AMAN.[1–5] Their pathogenicity was also confirmed by an inadvertent experiment in humans who had received ganglioside injections for various maladies and developed AMAN accompanied by anti-GM1 antibodies.[10] Additionally, antibodies to GQ1b or GD1a cause conduction block at the motor nerve terminals in a mouse phrenic nerve preparation.[1–6] Similar effects were noted with anti- GalNAc-Gd1a antibodies from a patient with AMAN.

IgG antibodies reacting with GM1, GD1a, GalNAc-GD1a, and GM1b are found in 80% of cases with axonal GBS (AMAN and AMSAN), but in the most common GBS subtype AIDP, ganglioside-specific antibodies are uncommon. Among gangliosides, the one that clearly correlates with a specific clinical syndrome is GQ1b, associated with the MFS-variant where IgG anti-GQ1b antibodies are present in greater than 90% of these patients.[1–5] In contrast, IgM-anti-GQ1b antibodies are found in chronic IgM paraproteinemic polyneuropathies,[5] as discussed later. Anti-GQ1b IgG antibodies are also found in postinfectious ophthalmoplegias and in GBS cases with ophthalmoplegia.[1–5] Anti-GQ1b antibody binds the paranodal regions of oculomotor nerves III, IV, VI, and may block impulse propagation at the nodes of Ranvier, resulting in conduction block. Many patients with GQ1b antibodies also have antibodies to GD1a. The recent finding of GD1b antibodies in two COVID-19–triggered GBS patients with ophthalmoplegia is of great interest as this pandemic is still evolving.[7,8]

The reasons for different clinical syndromes in connection with specific gangliosides remains unclear, but distribution, accessibility, and density or configuration of gangliosides at different sites may be critical factors. For example, there is more GM1 in ventral than dorsal roots: hence the predominantly motor neuropathy seen with anti-GM1 antibodies. There is also more GQ1b and possibly GD1b in oculomotor nerves, which may explain their involvement in MFS and COVID-19–triggered GBS.

Molecular Mimicry: Relationship Between *Campylobacter jejuni* and Gangliosides

Antecedent infection with *C. jejuni* has been commonly associated with AMAN. The strain of *C. jejuni*-associated AMAN (Penner D:19 serogroup) is, however, different from those causing common enteritis and is more likely to have the genes for enzymes that synthesize sialic acid in the bacterial wall, mimicking ganglioside GM1, GD1a, or GQ1b.[1–5] These patients have a higher incidence of anti-GM1 antibodies, suggesting cross-reactivity between epitopes in the lipooligosaccharide in the bacterial wall and nerve ganglioside.[1–5] Furthermore, injection of lipooligosaccharides extracted from *C. jejuni* into rabbits induces acute neuropathy with anti-GM1 antibodies identical to those found in AMAN.[1–5] Additionally, immunization of mice with these lipooligosaccharides generates a monoclonal antibody that reacts with GM1, binds to human peripheral nerve, and blocks muscle action potentials in muscle–spinal cord cocultures exactly as the anti-GM1 IgG extracted from GBS patients. Carbohydrate mimicry between the bacterial lipooligosaccharide and human GM1 is therefore an important cause of AMAN. Because *C. jejuni* is a common cause of a diarrheal illness worldwide, and diarrhea is an antecedent event in up to 50% of GBS patients, *Campylobacter* is a triggering factor for GBS in certain parts of the world. Isolation of *Campylobacter* from stools early in acute GBS varies from 44% to 88% of patients, and IgG or IgM *Campylobacter*-specific antibody titers are seen in a higher percentage (36%) of patients with GBS than in controls (10%).

Molecular mimicry may not be limited to *C. jejuni* because GM1 and GQ1b epitopes are also found in the bacteria wall of *Haemophilus influenzae*, which can also trigger GBS. Similarly, *Cytomegalovirus*-triggered GBS has been associated with IgM anti-GM2 antibodies. Molecular mimicry is also a factor in *M. pneumoniae*, which precedes GBS in 5% of cases and stimulates antibodies against human galactocerebroside, the main glycolipid in peripheral nerves.[1–5] Molecular mimicry may also play a role in Zika-associated GBS as antiglycolipid antibodies were found in 31% of these patients.[6] It may also be the case in some COVID-19–triggered GBS because GD1b antibodies were detected in 2 of 10 tested cases, and COVID-19 spike protein contains gangliosides and various sphingoglycolipds.[7]

Molecular mimicry between epitopes of disease-triggering viral proteins and myelin components may result in sensitization of cross-reactive T cells that stimulate B cells to produce specific antibodies against myelin components or recruit macrophages as effector cells. A combination of cellular and humoral factors probably participates in this process.[1–4] Circulating cytokines triggered by the initiating infections could also upregulate intercellular adhesion molecule (ICAM)-1 expression on the endothelial cells facilitating the entrance of activated T cells or antibodies to the endoneurial parenchyma. Of relevance, ICAM-1 is increased in patients with GBS.[1–5] A scheme summarizing the immunopathogenic mechanism is shown in Fig. 67.1.

Antibodies to paranodal antigens neurofascin and contactin, seen in CIDP subsets, as discussed later, have also been detected in a small number of GBS patients. These antibodies may potentially cause conduction block or paranodal axonal degeneration and could explain the rapid reversibility or slow recovery seen in several patients with AMAN.[1–5]

CHRONIC INFLAMMATORY DEMYELINATING POLYNEUROPATHY

CIDP is the most common form of chronic APN, with prevalence as high as 9/100,000.[1–3,11,12] It is also the most gratifying

KEY CONCEPTS

Autoimmunity in Guillain-Barré Syndrome

Cellular Factors

- The peripheral myelin or the Schwann cells are targets.
- Activated macrophages are the dominant endoneurial cells and lift the outermost myelin lamellae, lysing the superficial myelin sheaths.
- Peripheral blood lymphocytes exert myelinotoxic activity in vitro.
- Levels of interleukin-2 (IL-2) and soluble IL-2 receptors are increased during the acute phase of the disease and decline during recovery.

Humoral Factors

- Serum exerts a complement-dependent demyelination in vitro.
- Intraneural injections of serum from patients with acute Guillain-Barré syndrome (GBS) cause demyelination and conduction block.
- IgG, IgM, and membranolytic attack complex are detected immunocytochemically on the patients' nerves.
- High IgG antibody titers against peripheral nerve acidic glycolipids (GM1, GQ1b) are detected in the sera of patients with acute motor axonal neuropathy (AMAN) and Miller Fisher syndrome (MFS). The GQ1b ganglioside is a specific antigen for MFS. Emerging data suggest that GD1b ganglioside antibodies are seen in some COVID-19–triggered GBS
- There is a high incidence of antibodies to *Campylobacter jejuni*, and GM1, with molecular mimicry between *Campylobacter* and nerve gangliosides.
- Injection of lipooligosaccharides extracted from *C. jejuni* causes AMAN and elicits GM1 antibodies in rabbits.
- Antiganglioside antibodies extracted from patients with GBS block muscle action potentials in vitro.

chronic APN because it is treatable in the majority of cases. CIDP can be considered the chronic counterpart of GBS because of the various clinical, electrophysiological, histological, and laboratory similarities. CIDP differs from GBS predominantly by its tempo, mode of evolution, prognosis, and responsiveness to steroids. First described as a "steroid-responding relapsing polyneuropathy," CIDP shares with GBS a variety of common autoimmune features.

Clinical Features and Disease Variants

The typical CIDP is characterized by progressive, symmetrical, proximal, and distal muscle weakness, paresthesias, sensory dysfunction, and impaired balance that evolve slowly over at least 2 months.[1–3,11,12] Tendon reflexes are absent or reduced. Cranial nerves rarely may be affected. The course can be monophasic with stepwise progression, but also relapsing with even spontaneous remissions necessitating the need to periodically evaluate the usefulness of continuing immunotherapies. Because the demyelination is multifocal, affecting roots, plexuses, and proximal nerve trunks, the clinicopathological picture may be variable, accounting for the different manifestation of symptoms and signs.[1–3,11,12] CIDP variants include the asymmetrical, unifocal or multifocal, motor–sensory form (Lewis-Sumner syndrome); pure motor; sensory or sensory ataxia; and the distal variant.

Diagnosis

The CSF protein is elevated, up to sixfold, without pleocytosis (except if an infection coexists). Nerve biopsy shows demyelination and remyelination, occasional epineurial or

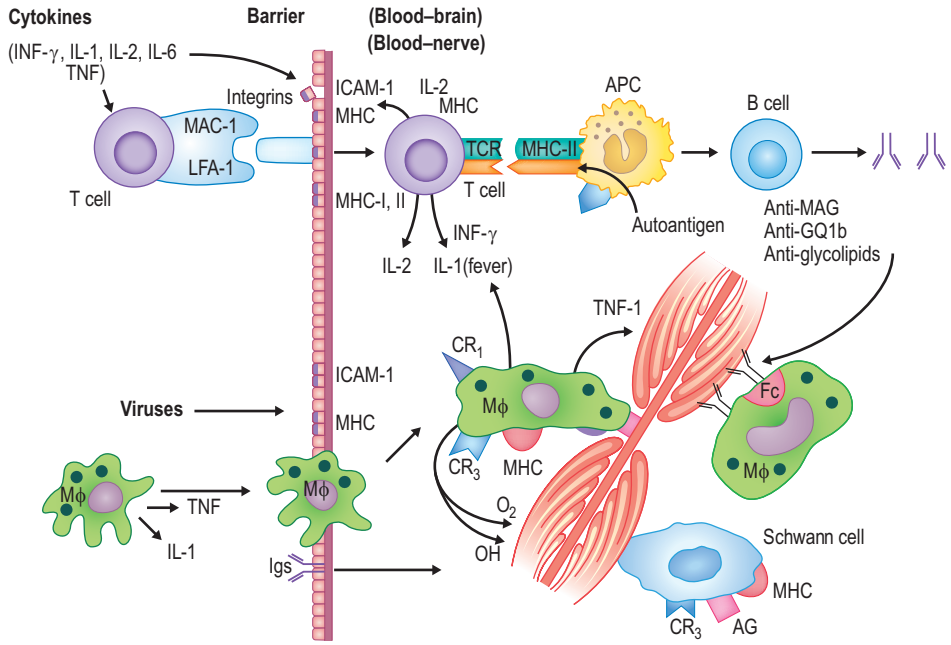


FIG. 67.1 Sequence of Events in the Mechanisms of Immune-Mediated Demyelinating Polyneuropathy. Cytokines lead to increased expression of major histocompatibility complex class I *(MHC-I)* and intercellular adhesion molecules (ICAM-1), allowing the sensitized T cells and macrophages to exit the endothelial cell wall and traffic to the peripheral nerve. There they recognize myelin antigen and induce a macrophage-mediated demyelination. The antigen-presenting cells (*APCs*; probably Schwann cells or macrophages), in concert with MHC class II *(MHC-II)* expression, interact with CD4 T cells and lead to clonal expansion of B cells, producing antibodies against various peripheral nerve antigens. *AG*, antigen; *APC*, antigen-presenting cell; *CR*, complement receptor; *GQ1b*, gangliocide GQ1b antibody; *Igs*, immunoglobulins; *IL-1,2,6*, Interleukin-1,2,6; *INF*, interferon; *LFA-1*, lymphocyte function associated antigen-1; *MAG*, myelin-associated glycoprotein; *MHC*, major histocompatibility complex; *OH*, hydroxyl radical; *TCR*, T cell Receptor; *TNF*, Tumor necrosis factor.

endoneurial T cells, and several macrophages scattered or in small perivascular clusters in the endoneurium (Fig. 67.2).[1–3,12,13] Electrophysiological testing is fundamental for the diagnosis by demonstrating demyelination in motor and sensory fibers. including slow conduction velocity, prolonged distal motor or sensory latencies, prolonged F-wave latencies, and conduction block with dispersion of the compound muscle action potentials. An associated axonal loss is common in the majority of cases. A variety of diagnostic criteria have been proposed to capture the most pertinent of the aforementioned features; the revised European Federation of Neurological Societies/Peripheral Nerve Society (EFNS/PNS) guidelines seem most appropriate, offering 81% sensitivity and 96% specificity in capturing patients more likely to respond to immunotherapies.[1–3,12] Routine CSF testing and nerve biopsy are not mandatory for diagnosis, but can be helpful when the electrophysiological results are not clear or need to exclude other causes.[3,12] Concomitant diabetes is an important factor because it is more frequent in CIDP and is needed to dissect if any neuropathy component is diabetes related. Other neuropathy causes that should be excluded are paraneoplastic, IgM paraproteinemias (IgG or IgA monoclonal gammopathy of undetermined significance [MGUS] can be seen in CIDP), myelomas, vasculitis, alcoholism, neurotoxic drugs, or family history.

KEY CONCEPTS

Autoimmunity in Chronic Inflammatory Demyelinating Polyneuropathy

- Activated macrophages are the predominant endoneurial cell, displacing the Schwann cell cytoplasm, disrupting myelin, and lysing superficial myelin lamellae.
- Complement-fixing IgG and IgM antibodies are deposited on the myelin sheath.
- IgG antibodies to acidic glycolipids LM1, GM1, or GD1b and against the 28-kDa P0 myelin proteins are detected in the sera of some patients.
- There is upregulation of DR and B-7 costimulatory molecules in Schwann cells and macrophages.
- Serum IgG can induce conduction block when injected into rat nerves.
- Up to 25% of patients with *chronic inflammatory demyelinating polyneuropathy* (CIDP) harbor specific antibodies against antigens in the nodes of Ranvier; in 10% of them these antibodies have been identified as directed against neurofascin-155 and contactin-associated protein (CASPR), causing conduction block.

Immunopathogenesis

Activated T cells, macrophages, complement, and autoantibodies seem to work in concert with each other to induce an immune attack against peripheral nerve antigens (see Fig. 67.1).[2,3,11,12] No triggering factors have been identified, however.

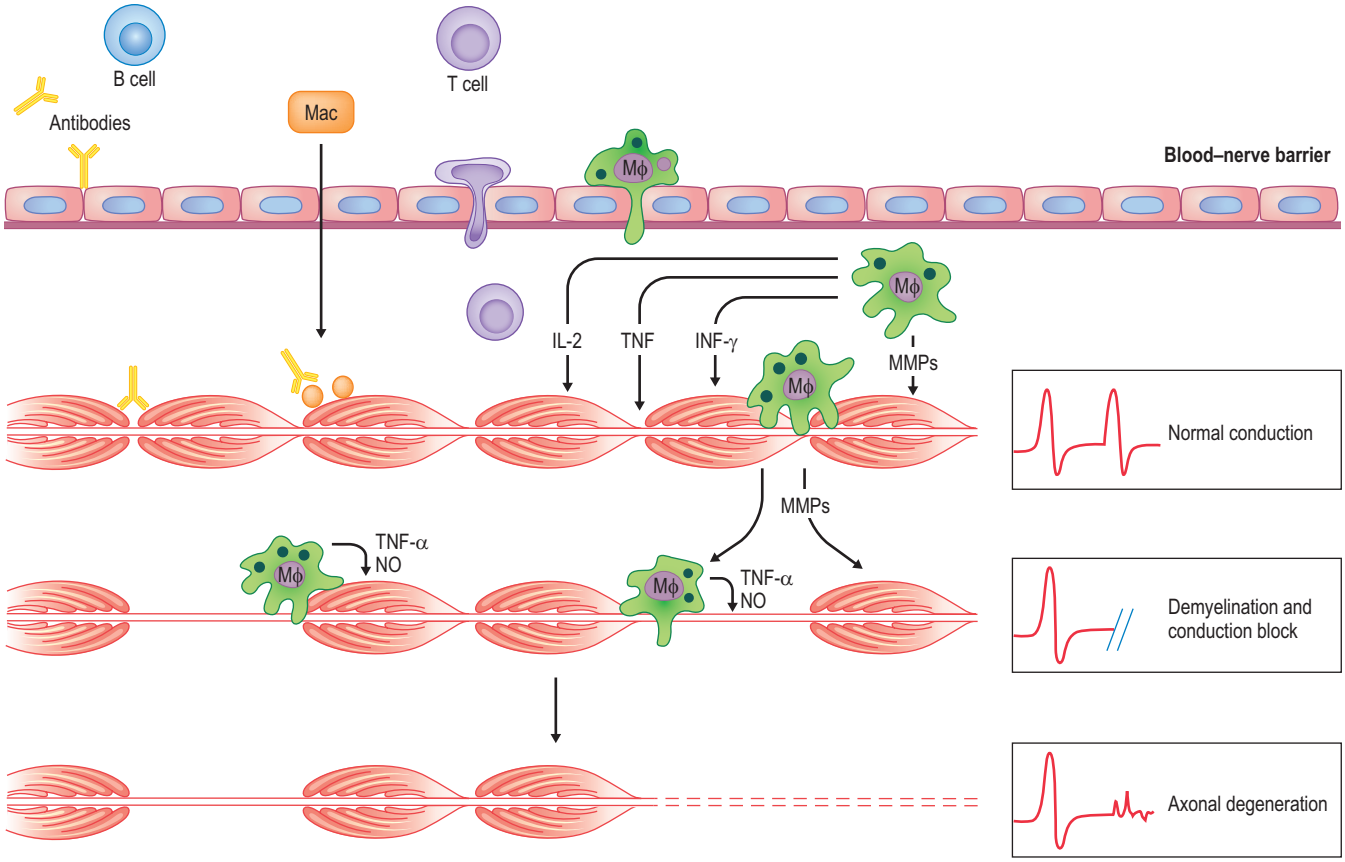


FIG. 67.2 Diagrammatic Scheme of the Main Cellular and Humoral Factors Implicated in the Demyelinating Process of Chronic Inflammatory Demyelinating Polyneuropathy (CIDP) Leading to Axonal Loss. Activated macrophages (Mϕ) and T cells cross the endothelial cell wall of the blood–nerve barrier and reach the myelinated fibers. Activated, TNF-α-positive Mϕ invade the myelin sheath, causing Mϕ-mediated segmental demyelination. Axonal loss secondary to demyelination, probably enhanced by TNF-α and metalloproteinases, may become prominent in the chronic phases of the disease. Other cytokines, T cells sensitized to unidentified antigens, and putative antibodies may participate. *IFN*, Interferon; *IL*, interleukin; *MMP*, metalloproteinase; *Mϕ*, activated macrophage; *NO*, nitrous oxide; *TNF*, tumor necrosis factor.

The predominant endoneurial mononuclear cells are macrophages that constitute the final effector cells associated with demyelination; macrophages express activation markers probably induced by cytokines released by autoreactive T cells in situ or the circulation. These cells penetrate the basement membrane of the Schwann cell, displace the cytoplasm, and split the myelin lamellae.[1–3,12,13] Macrophages and Schwann cells may serve as antigen-presenting cells because they express human leukocyte antigen (HLA)-DR and costimulatory molecules B7-1 (CD80) and B7-2 (CD86), whereas their counterreceptors cytotoxic T lymphocyte antigen-4 (CTLA-4) and CD28 are expressed on rare endoneurial CD4 T cells.[2,3,13] B7-2–deficient mice also develop CIDP.[13] Soluble adhesion molecules, cytokines, and metalloproteinases, detected in serum and CSF, facilitate lymphoid-cell transmigration across the blood–nerve barrier. Although T cells are not prominent, the few endoneurial CD8 and CD4 cells have monoclonal or oligoclonal restrictions in their T-cell receptor repertoire, implying an antigen-driven T-cell response.[2,3,12]

Although pathogenic antigens remain elusive, humoral factors seem to play a role as supported by the beneficial effect of plasmapheresis. That antibodies may be implicated dates back to 40 years ago when complement-fixing IgG and IgM were first found deposited on the patient's myelin sheath.[14] Antibodies to glycolipids LM1, GM1, or GD1b also have been seen in some patients, more frequently than in controls.[2,3,11] Overwhelming evidence the last 5 years shows that molecules associated with saltatory conduction at the nodes of Ranvier are more meaningful targets[1–4,11,15–18] because functional blockade in these regions can best account for the rapid improvement noticeable within days after plasmapheresis or intravenous immunoglobulin (IVIG). Antigenic targets at the nodes of Ranvier are seen in approximately 10% of CIDP patients and include pathogenic antibodies against neurofascin-186, moesin, and gliomedin (at the node); and neurofascin-155 (NF155), contacting CASPR 1 (CNTN1), and connexins (at the paranode).[1–4,11,15–17] These antibodies are of the IgG4 subclass causing disadhesion of the myelin lamellae without fixing complement. Of interest, anti-NF155– and CNTN1-positive patients have distinct clinical phenotypes with more severe disease, axonal involvement, tremors, sensory ataxia, and suboptimal response to IVIG;[1–4,11,15–17] they respond, however, to rituximab that effectively depletes IgG4 from short-lived plasma cells.

Overall, the immunopathogenetic scheme proposed for GBS (see Fig. 67.1) is also applicable to CIDP. Molecular mimicry can be implicated in rare CIDP cases associated with melanoma because the carbohydrate myelin epitopes GM2, GM3, and GD3 are also expressed on melanoma cells, and antibodies against melanoma cells react with myelin glycoproteins.[18] The axonal involvement accompanying or following demyelination in CIDP and GBS is depicted in Fig. 67.2 and the nodal antigens in Fig. 67.3.[2,3,11,15,16]

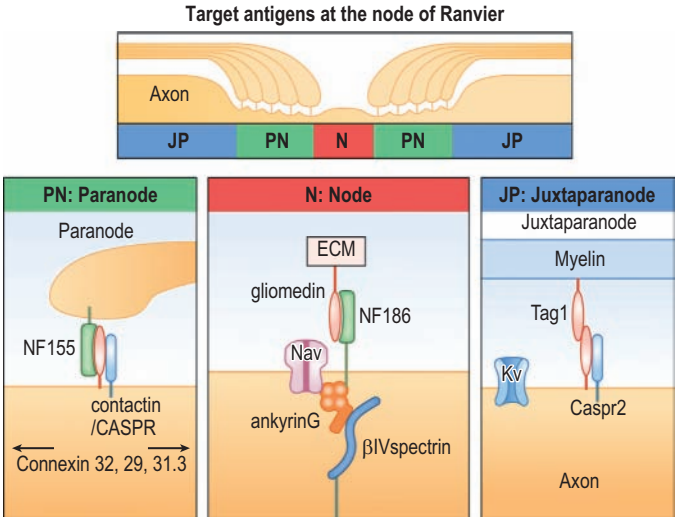


FIG. 67.3 Main proteins in the nodal (*N*), paranodal (*PN*), and juxtaparanodal (*JP*) regions in the nodes of Ranvier implicated as antigens in acquired demyelinating neuropathies (Guillain-Barré syndrome [GBS] and predominantly chronic inflammatory demyelinating polyneuropathy [CIDP]). The two main antigenic targets in CIDP are neurofascin-155 and CASPR2 located in the paranodal regions; antibodies against these proteins are detected in the serum of 10% of patients with CIDP, resulting in conduction block and paranodal axonal changes that probably account for resistance to intravenous immunoglobulin (IVIG).[4,13]

MULTIFOCAL MOTOR NEUROPATHY WITH CONDUCTION BLOCK

Multifocal motor neuropathy (MMN) is a distinct disease that, although rare with prevalence of 0.6/100,000, should be recognized early because it is treatable. It affects males more than females and presents with progressive weakness, atrophy, and areflexia. MMN often begins in the hands and is prominent in distal muscle groups supplied by many individual peripheral nerves (multifocal).[1–3,19] It differs from vasculitic neuropathy because it has slow onset, is painless, and affects only the motor nerve fibers. It also differs from the motor variant of CIDP because it is multifocal, distal, and asymmetrical. MMN is sometimes misdiagnosed as lower motor neuron disease, especially when patients experience cramps and fasciculations, and treatment is delayed. In contrast to lower motor neuron disease, MMN progresses very slowly, the weakness is within distributions of peripheral nerves and not multisegmental, while cranial musculature is often spared. CSF protein is normal, in contrast to the motor CIDP variant. MMN is electrophysiologically distinct because it exhibits multifocal conduction blocks only in motor nerves; sensory conduction remains normal across the nerve segments with motor block.

Up to 50% of MMN patients have high IgM antibodies to GM1 ganglioside, which activate complement in association with disease severity[1–3,19] but their pathogenic role is uncertain. GM1 antibodies can also be seen in other autoimmune neuropathies and up to 25% of amyotropic lateral sclerosis (ALS) patients. The reason for the selective motor involvement is unclear. Differences in the antigenic specificities of myelin components between motor and sensory fibers are suspected because the ceramide composition of gangliosides differs between sensory and motor fibers. MMN responds remarkably well to IVIG.

POLYNEUROPATHIES ASSOCIATED WITH MONOCLONAL GAMMOPATHIES OF UNDETERMINED SIGNIFICANCE

A distinct subset of acquired polyneuropathies has been associated with a circulating paraprotein (see Chapter 79). Although neuropathy occurs in a setting of myeloma, plasmacytoma, or Waldenström macroglobulinemia, the majority of patients with paraproteinemic neuropathies do not have a lymphoproliferative

disease, and the MGUS. Up to 1% of normal people greater than 50 years of age may have MGUS but the incidence increases to 1.7% above 70, reaching up to 6% above the age of 90. Monoclonal gammopathies, however, are 10 times more frequent in patients with polyneuropathy than in an age-matched control population, and almost 10% of patients with acquired polyneuropathy have MGUS.[2,3,20] If these gammopathies are categorized into subclasses, the incidence of polyneuropathy among patients with IgM monoclonal proteins is as high as 50%,[2,3,20] implying that almost 50% of patients with IgM MGUS may have or will develop polyneuropathy. At present, polyneuropathies with MGUS comprise 10% of patients with acquired neuropathy.[2,3,20] Patients with demyelinating polyneuropathy associated with IgG or IgA MGUS are indistinguishable from CIDP and the MGUS is coincidental and causally unrelated to neuropathy. In contrast, polyneuropathy associated with IgM MGUS is a distinct clinicopathological entity, and the IgM is considered pathogenic, often directed against myelin glycoproteins or glycolipids.[2,3,20–23]

Some patients with paraprotein may have an associated amyloidosis derived from the variable region of the Ig light chain, primarily λ. When amyloidosis is present, the neuropathy is painful and often accompanied by autonomic symptoms of orthostatic hypotension, impotence, and impaired gastric motility.

ANTIBODIES TO MYELIN-ASSOCIATED GLYCOPROTEIN (MAG) IN PATIENTS WITH IgM MGUS (ANTI-MAG NEUROPATHY)

Most of these patients present with a sensory, large-fiber, demyelinating polyneuropathy that manifests as sensory ataxia.[1–3,20,21] Others have a sensorimotor polyneuropathy with mixed demyelinating and axonal features. CSF protein is often elevated. Nerve conduction studies typically demonstrate distal demyelination with prolonged distal motor and sensory latencies. Sural nerve biopsy shows a diminished number of myelinated axons. On electron microscopy, there is splitting of the outer myelin lamellae, linked to the presence of IgM deposits in the same area of the split myelin sheath.[1–3,20,21]

Approximately 50% of these patients react with MAG, a 100 kDa glycoprotein of the central and peripheral nerve myelin, as well as other glycoproteins or glycolipids that share antigenic determinants with MAG.[2,3,20–23] The anti-MAG IgM paraproteins co-react with an acidic glycolipid in the ganglioside fraction of peripheral nerve, identified as a sulfoglucuronyl glycosphingolipid (SGPG).[2,3,20–23] In contrast to MAG, which is mostly present in the central nervous system, SGPG is only in the peripheral nerves. In some patients with IgM MGUS the IgM also reacts with various gangliosides, most commonly those that contain either a disialosyl moiety, such as GD1b, GQ1b, GT1b, GalNac-GM1b, and GalNAc-GD1a, or two gangliosides that share epitopes with GM2, or a combination of GM2 and GM1, GM1 and GD1b.[2,3,20–23] More than half of the IgM paraproteins recognize MAG and SGPG, and 75% of the rest recognize ganglioside antigens, indicating that acidic glycolipids are the most common antigenic epitopes.[2,3,20–23] The glycolipids implicated in immune-mediated neuropathies are depicted in Fig. 67.4.

Anti-MAG antibodies are detected readily in patients' sera with ELISA or preferably with Western blot. Because anti-MAG–reacting sera always recognize the SGPG glycolipid, the assay is often performed by using SGPG as antigen instead of purified human MAG. It is preferable, however, to use MAG as the target antigen so as not to miss low-affinity antibodies because IgM binds to MAG with 10 to 100 times higher affinity compared with SGPG.

The following factors suggest that MAG antibodies are related to the cause of the neuropathy.[2,3,20–23]

1. IgM and complement are deposited on the patient's myelinated fibers suggesting that activated complement may be needed in the induction of demyelination.
2. IgM recognizes neural cell adhesion molecules and colocalizes with MAG on the areas of the split myelin lamellae, suggesting involvement in myelin disadhesion. Skin biopsies from these patients have also confirmed the presence of IgM, complement C3d, and MAG deposition on the dermal myelinated fibers with concurrent loss of nerve fibers.[24]
3. Serum from these patients injected into feline peripheral nerve causes complement-dependent demyelination and conduction block;[25] the injected IgM binds to the outer layer of the myelin sheath.
4. Systemic transfusion of anti-MAG IgM paraproteins produces segmental demyelination in chickens,[26] with the IgM splitting the myelin lamellae, similar to human neuropathy.
5. Immunization of cats with purified SGPG causes ataxic neuropathy, similar to human disease, with inflammation of the dorsal root ganglionic neurons.[27]

KEY CONCEPTS

Autoimmunity in Polyneuropathy With IgM Monoclonal Gammopathy

- In more than 50% of patients IgM is an antibody against two antigens, myelin-associated glycoprotein (MAG) and sulfoglucuronyl glycosphingolipid (SGPG).
- In many patients with non-MAG–reacting monoclonal IgM, the IgM recognizes *(i)* gangliosides containing disialosyl moieties, including GM1, GM2, GD1b, GD1a, and LM1; *(ii)* sulfatides; and *(iii)* rarely, chondroitin sulfate.
- Overall, in at least 75% of patients the IgM recognizes gangliosides that appear to be the primary antigenic targets.
- IgM is deposited on the homologous myelin sheath and fixes complement.
- IgM, when deposited on the myelin sheath, results in disadhesion and separation of the myelin lamellae and disruption of normal myelin function.
- Intraneural injection of anti-MAG–reacting IgM or passive transfusion into experimental animals causes segmental demyelination, whereas complement-fixing IgM is immunolocalized to the myelin sheath, causing myelin separation.
- Immunization of cats with purified SGPG causes an ataxic neuropathy, similar to the one seen in humans, with involvement of the dorsal root ganglia.

POLYNEUROPATHY, ORGANOMEGALY, ENDOCRINOPATHY, MYELOMA, AND SKIN CHANGES (POEMS SYNDROME)

A subset of patients with malignant IgG or IgA monoclonal proteins have polyneuropathy with osteosclerotic myeloma. Most of them have POEMS syndrome (*p*olyneuropathy, *o*rganomegaly, *e*ndocrinopathy, *M* protein, and *s*kin changes).[1–3,28] Not included in the acronym are sclerotic bone lesions, giant lymph node hyperplasia (Castleman disease), papilledema, pleural effusion,

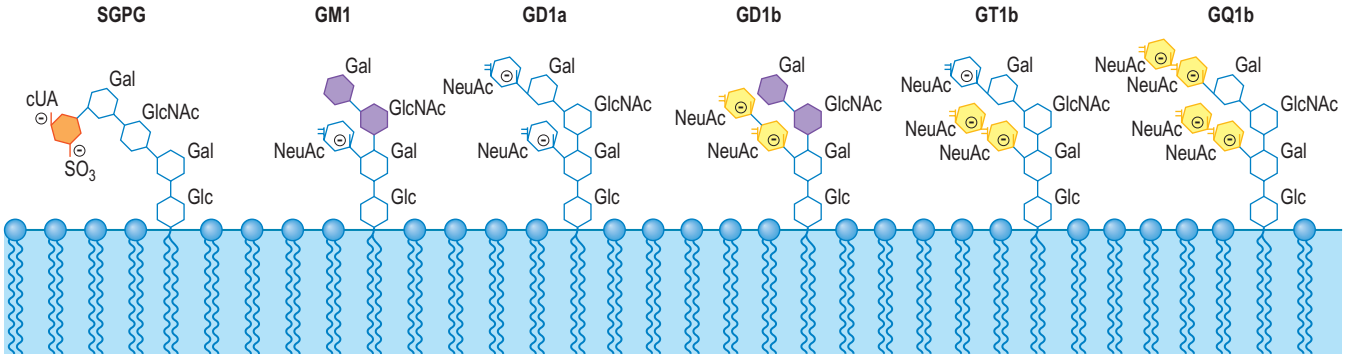


FIG. 67.4 Glycolipids Implicated as Antigens in Immune-Mediated Neuropathies. Sulfate-3-glucuronyl paragloboside (SGPG) is the glycolipid sharing a carbohydrate epitope with myelin-associated glycoprotein (MAG), and the terminal sulfated glucuronic acid is a key part of the epitope. GM1 is the ganglioside implicated in motor nerve disorders, and in most cases the terminal Gal (β1–3) GalNAc epitope, which is shared with GD1b, is involved. The disialosyl moiety implicated in sensory neuropathies consists of NeuAca2—8NeuAc—and is present in GD1b and GT1b gangliosides, as well as the simpler GD2 and GD3 gangliosides (not shown). GQ1b ganglioside, which is the target antigen in Miller Fisher syndrome (MFS), has two disialosyl moieties. Although GD1a ganglioside has two sialic acid residues, they are not linked to each other, so antibodies to GD1a do not cross-react with anti-GD1b antibodies. The color-coded sugar moieties represent key aspects of the various epitopes, but carbohydrate sequences recognized by the antibodies may include additional sugar residues. *Gal*, Galactose; *GalNAc*, *N*-acetylgalactosamine; *Glc*, glucose; *G1cNAc*, *N*-acetylglucosamine; *GlcUA*, glucuronic acid; *NeuAc*, *N*-acetylneuraminic acid (sialic acid).

edema, ascites, and thrombocytosis.[28] More than 50% of patients with osteosclerotic myeloma of the IgA or IgG type have sensorimotor polyneuropathy with mixed demyelinating and axonal features and high CSF protein. The neuropathy tends to be associated with leg edema, hyperpigmentation, sclerodermatous thickening, and hypertrichosis with dark hair. Endocrinopathy most often includes gonadal failure, amenorrhea, impotence, gynecomastia, hypothyroidism, diabetes, or elevated prolactin levels. The IgG class is slightly more common than the IgA, with λ light chain present in the majority of patients.[28] Bone lesions can be sclerotic, solitary, or multiple, sparing the skull and the extremities. Pathological changes in lymph nodes resemble those of Castleman disease, which can also be associated with polyneuropathy.[28] There is imbalance of proinflammatory cytokines with increased IL-1β, IL-6, and tumor necrosis factor-α. Vascular endothelial growth factor (VEGF) may play a major role because it induces rapid increase in vascular permeability and angiogenesis.[28]

In some patients, the neuropathy responds to steroids, tamoxifen, or alkylating agents. In others, it responds to removal or irradiation of the solitary sclerotic lesion, suggesting that the tumor may secrete neurotoxic factors. IVIG and plasmapheresis are ineffective. Autologous stem cell transplantation may result in significant improvement with reduction of VEGF, improved nerve conduction velocity, and increased survival.[29]

CRYOGLOBULINEMIC NEUROPATHY

Cryoglobulins are proteins that precipitate in the cold and redissolve when heated. There are three types of cryoglobulin: type I, which is monoclonal, often of the IgM and IgG class; type II, which is mixed polyclonal, with one monoclonal (often monoclonal IgM with polyclonal IgG); and type III, which is polyclonal (often IgM and IgG). Polyneuropathy occurs most often with mixed cryoglobulinemias and presents with distal sensorimotor involvement or as mononeuropathy multiplex. Nerve biopsy shows perivascular inflammatory cuffing with axonal degeneration. Patients also have purpura, polyarthralgias, cutaneous vasculitis, Raynaud phenomenon, renal involvement, and increased incidence (up to 90%) of hepatitis C infection.

PARANEOPLASTIC PERIPHERAL NEUROPATHIES WITH ANTI-HU OR CRMP-5 ANTIBODIES

Peripheral neuropathy in patients with cancer is either related to the systemic effects of the tumor or, more often, to various chemotherapeutic agents that typically presents as painful dysesthesias. The most distinct immune-related neuropathy is the *paraneoplastic sensory neuronopathy* (PSN), often associated with small cell lung cancer, breast cancer, lymphomas, or thymomas.[30] It might be the presenting symptom of the underlying neoplasm typically characterized by gait ataxia with choreoathetotic movements related to loss of proprioception in the feet and hands, and distal paresthesias. Strength is normal. Some patients may have autonomic dysfunction. CSF protein is increased and electrophysiology shows axonal sensory neuropathy. PSN is a sensory neuronopathy caused by a variable degree of inflammation in the dorsal root ganglionic neurons. Typically, the patients have specific IgG anti-Hu autoantibodies directed against a closely spaced group of proteins with a molecular weight of 35 to 40 kDa.[30] The antibodies are in higher titers in the CSF, suggesting intrathecal synthesis. The Hu protein may also be expressed in the tumors. Because low-titer anti-Hu antibodies are seen in 20% of patients with small cell lung cancer even without neurological symptoms, PSN may represent an autoimmune reaction against antigens shared by both the tumor cells and the dorsal root ganglionic neurons. Anti-Hu antibodies are helpful markers in suspecting occult small cell lung cancer in patients presenting with sensory ataxic neuropathy. Some patients may also have other paraneoplastic antibodies, such as collapsing response mediator protein (CRMP-5), often referred to as anti-CV2.[30] Immune neuropathies are also seen with immune-checkpoint inhibitors, most often GBS or CIDP, responding to immunotherapies.[9]

AUTOIMMUNE AUTONOMIC NEUROPATHIES

Autoimmune autonomic neuropathy (AAN) is highlighted by high-titer antibodies against the ganglionic nicotinic acetylcholine receptors (Gn-AChRs).[31] Patients present with a subacute (within 4 weeks) or chronic (within months) onset of neurogenic orthostatic hypotension, with systolic blood pressure reduction of at least 30 mm Hg or mean blood pressure reduction of at least 20 mm Hg within 3 minutes of head tilting. Onset may be preceded by viral infections. Patients may also exhibit parasympathetic/enteric symptoms including sicca (dry eyes, dry mouth); abnormal pupillary responses; gastrointestinal symptoms (postprandial nausea and vomiting leading to weight loss); and neurogenic bladder. Some symptoms can be passively transferred to mice injected with the patient's IgG while rabbits immunized with a fragment of Gn-AChRs exhibit autonomic failure similar to human disease, suggesting that the antibodies may be pathogenic.[30] Because Gn-AChRs have been found in small cell lung carcinoma cell lines, cancer may be a potential initiator of ganglionic autoimmunity.

A subset of patients with these symptoms do not have Gn-AChRs and an autonomic autoimmune ganglionopathy is circumstantially suspected when symptoms occur in a setting of an underlying autoimmune neurological or rheumatological disease. It has been generally difficult to clinically distinguish whether some of these patients have an organic autonomic nervous system autoimmunity or a functional disorder.

POSSIBLY AUTOIMMUNE SMALL FIBER SENSORY NEUROPATHIES, NEUROPATHIC PAIN, AND NEURAL ANTIBODIES

Small fiber sensory neuropathy (SFN) is now one of the commonest neuropathies. Patients present with diffuse pains, intolerance to light touch, allodynia or hyperalgesia, and some with autonomic features. They typically have normal neurological examination including sensation, reflexes, strength, balance, and nerve conduction studies.[32] Reduced distal intraepidermal nerve fiber density is seen on skin biopsy. The condition represents involvement of unmyelinated C-fibers, thinly myelinated A-δ somatosensory fibers, and possibly autonomic neurons, leading to diffuse painful dysesthesias. In several patients these symptoms overlap with other painful disorders such as fibromyalgia and erythromelalgia. When all causes are excluded, they are categorized as idiopathic. A functional component may at times be prominent. There is recent emphasis on autoimmunity because almost 20% of the patients may have a systemic autoimmune or rheumatic disease such as Sjögren syndrome, celiac disease, rheumatoid arthritis, or nonspecific immunological abnormalities such as extractable nuclear antigen antibodies.[32,33] Some patients anecdotally respond to IVIG.[32] Although no underlying autoimmunity has been identified, two nonspecific autoantibodies, one against *trisulfated heparin disaccharide (TS-HDS)*, a disaccharide component of the glycosylation moieties of heparin and heparan sulfate, and another against *fibroblast growth factor-3 (FGFR3)*, a secreted cell surface receptor, have been more frequently detected in such patients than controls.[34]

MONONEUROPATHY MULTIPLEX AND LOCALIZED, ISOLATED VASCULITIS OF THE PERIPHERAL NERVES

Polyneuropathy is a common manifestation of systemic vasculitis. It occurs in patients with polyarteritis nodosa; connective tissue diseases; hypersensitivity vasculitis; Churg-Strauss syndrome; temporal arteritis; and viral infections, especially retroviruses and hepatitis. It classically presents as mononeuritis multiplex affecting several individual nerves with painful weakness and paresthesias caused by ischemia and infarcts due to inflammation of endoneurial blood vessels. There is, however, a distinct vasculitic entity localized only to the peripheral nerve, known as *isolated peripheral nerve vasculitis* (PNV) with presentation similar to vasculitic neuropathy but without any systemic organ involvement, negative serology, and slower onset and progression. PNV involves the small and medium-sized arteries of the epineurium and perineurium and causes ischemic changes within the peripheral nerve. The diagnosis is confirmed with nerve biopsy of the sural or superficial peroneal nerve often combined with muscle biopsy for higher diagnostic yield. PNV has a better prognosis compared to systemic vasculitides and is a treatable form of neuropathy.

NEUROPATHY WITH VIRUSES

Neuropathy can be seen in a setting of infectious, viral, or bacterial processes such as Lyme disease, CMV, hepatitis, herpes, human immunodeficiency virus, or COVID-19 as a manifestation of triggered autoimmunity rather than direct infection of the nerves: most commonly included are GBS, CIDP, acute ganglioneuritis, Bell palsy, mononeuritis multiplex, or even small fiber sensory neuropathy seen early in the infection, or rarely as the presenting manifestation of the infection. The best studied cases are due to HIV where immunocytochemical studies have shown HIV in rare endoneurial macrophages, but not within Schwann cells or axons. There is strong expression of HLA class I and II molecules on Schwann cells, endothelial cells, and macrophages, but sparse CD8 and CD4 T cells (Fig. 67.5).[31]

A rare neuropathy seen in later-stage HIV infection is a lumbosacral polyradiculoneuropathy related to CMV. It affects roots and sensory ganglia and presents with lower-extremity muscle weakness, sacral and distal paresthesias, areflexia, muscular atrophy, and sphincteric dysfunction resembling cauda equina syndrome. CMV inclusions are within Schwann cells or endothelial cells (Fig. 67.6).[31] Presently, the commonest neuropathy in HIV patients is a sensory axonal neuropathy manifested with distal painful dysesthesias and areflexia most often related to antiretroviral drugs.

TREATMENT

APNs are clinically important because they are potentially treatable with various immunosuppressive, immunomodulating, or chemotherapeutic agents. The author's approach to the treatment is as follows.

Guillain-Barré Syndrome

Supportive Care

The dramatic reduction in the mortality of GBS is mainly attributed to early ICU care, improved respiratory support, and control of autonomic cardiac dysregulation. A patient with GBS is best monitored in an ICU, even if respiratory compromise is not evident at the time of admission; when vital capacity drops or bulbar weakness is severe, intubation is necessary.

Plasmapheresis

In double-blind controlled studies, plasmapheresis has been effective if performed within the first week from onset. Five or six

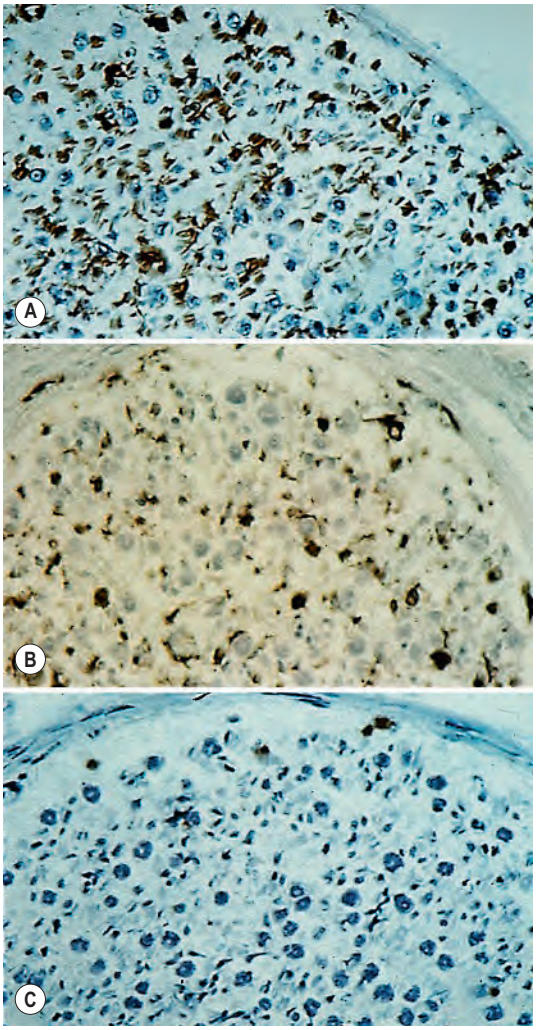


FIG. 67.5 Serial sections of a nerve biopsy from a patient with human immunodeficiency virus (HIV)–chronic inflammatory demyelinating polyneuropathy stained for (A) human leukocyte antigen (*HLA*)-*DR*, (B) macrophages, and (C) CD8 T cells shows that the majority of the endoneurial cells are macrophages. Only rare CD8 cells are noted.

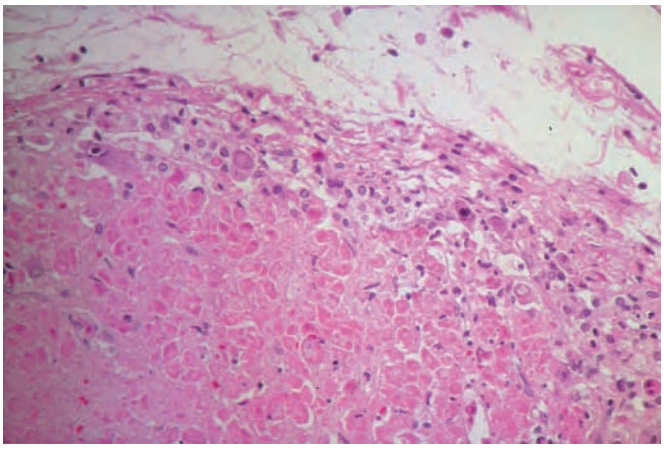


FIG. 67.6 Cross-section of a root from a patient with human immunodeficiency virus (HIV)–associated Guillain-Barré syndrome (GBS) shows cytomegalovirus inclusions within the Schwann cell.

exchanges, one every other day, are sufficient. Early relapses can occur in 20% of patients, who may require a second series.[1–4] In mild cases two exchanges are sufficient and in moderate cases four are optimal.

Intravenous Immunoglobulin

Based on two controlled studies,[1,4] IVIG (see Chapter 82), given at 2 g/kg over 2 to 5 days, is as effective as plasmapheresis, with no added benefit when the two procedures were combined. The decision as to which treatment to choose is governed by circumstances, availability of treatment modality, experience, age of the patient, and other associated conditions. Early relapses can also occur with IVIG, as often as with plasmapheresis. IVIG has become the therapeutic choice worldwide because it is easy to administer and more readily available for early therapy initiation. In spite of anecdotal reports that a second IVIg course may be beneficial, recent data from a controlled study suggest that a second IVIg infusion is not helpful or safe in severely affected GBS patients unresponsive to the first course. Steroids are ineffective and may even increase the incidence of future relapses. Combining IVIG with IV methylprednisolone has shown no added benefit.

Chronic Inflammatory Demyelinating Polyneuropathy

Prednisone

CIDP was originally described as a steroid-responsive polyneuropathy with its efficacy later proven in a small controlled study.[1–3,11,12] Whenever used, a high-dose regimen of 80 to 100 mg prednisone daily is preferred, followed by tapering to every-other-day dosing. Azathioprine, cyclosporine, methotrexate, or mycophenolate are ineffective as steroid-sparing agents.[1–3,11]

Intravenous Immunoglobulin

In several controlled studies,[1–3,11,12] IVIG has been effective and approved as first-line therapy. Monthly maintenance therapy is needed, preventing relapses in up to 60% of patients. A recent controlled trial showed that subcutaneous immunoglobulin (SCIG) is as effective, but whether it is preferable by the patients has not been documented[35] especially since a number of patients relapsed after transitioning from IVIG to SCIG.[35]

Plasmapheresis

It is also effective in controlled studies.[1–3,11,12] After a series of six plasma exchanges, maintenance therapy, with one exchange at least every 8 weeks, may be required. IVIG has now replaced plasmapheresis in most centers.

Polyneuropathy with Paraproteinemias

Patients with benign IgG or IgA demyelinating polyneuropathies respond in a manner similar to CIDP. Patients with malignant paraproteinemias should be treated with chemotherapy, as needed for the underlying disease. When the neuropathy is axonal, treatments are generally disappointing.

For IgM anti-MAG demyelinating polyneuropathies, treatments with prednisone plus chlorambucil, plasmapheresis, and IVIG[2,3,20] has shown a variably marginal or minimal benefit. Rituximab is, however, the most promising therapy,[20] providing

efficacy in almost 40% of the patients in a small double-blind study, as confirmed later with a larger study,[19] even though both did not reach significance. Additional, uncontrolled series with many patients have confirmed that rituximab is effective in 30% to 40% of these patients.[20]

MULTIFOCAL MOTOR NEUROPATHY

MMN responds very well only to IVIG, which is the treatment of choice based on controlled trials. In difficult cases, cyclophosphamide or rituximab may be promising, but no controlled studies have been carried out.[1–4]

Paraneoplastic Neuropathy

Anecdotally, some of these patients have responded to plasma exchange or IVIG, but overall this neuropathy has not been consistently responsive to available therapies.

Vasculitic Neuropathies

For isolated peripheral nerve vasculitis, a combination of prednisone 1.5 mg/kg/day with cyclophosphamide 2 mg/kg/day orally, or 1 g/m^2 intravenously monthly for 6 months, is the treatment of choice. Plasmapheresis has been tried in cryoglobulinemic neuropathies, with variable results.

HIV Neuropathies

GBS and CIDP in a setting of HIV are treated with the same therapies as in HIV-negative patients, with IVIG being the preferable choice. Ganciclovir may be helpful in CMV-related polyradiculoneuropathy.

Possibly Autoimmune Small-Fiber Sensory Neuropathies

This is a difficult group of patients to provide effective therapies because the autoimmunity is unclear. In some patients steroids and IVIG can be anecdotally effective. Tricyclic antidepressants, carbamazepine, gabapentin, Lyrica (pregabalin), topiramate, and Cymbalta (duloxetine) are the preferable therapies. Topical capsaicin in various combinations may provide some pain relief.

ON THE HORIZON

- Identification of common causes that break tolerance, either in connection with the gut microbiome or exogenous agents, and expansion of the current limited data on molecular mimicry between nerve glycoproteins and viral or bacterial antigens.
- Development of neuroimaging capabilities to image peripheral nerves and dorsal roots in vivo and quantify inflammation, demyelination, and axonal degeneration, providing an accessible and reliable tool for diagnosis and monitoring response to therapies, in a model similar to magnetic resonance imaging (MRI) in multiple sclerosis. Nerve ultrasound seems now promising.
- Performance of proteome studies and systematic autoantigenomic approaches to identify biomarkers of nerve autoimmunity that may lead to specific therapies.
- Trials with target-specific therapies for acute and chronic demyelinating neuropathies, *applied in combination with the existing therapies*, targeting early in the disease process complement activation, regulatory T-cell (Treg) functions, key proinflammatory cytokines, and B cells.
- Identification of agents and trophic factors that could prevent axonal degeneration from the outset, promote remyelination, prevent axonal loss, or trigger axonal regeneration and reverse, early in the disease process, a seemingly "permanent" clinical deficit.

ACKNOWLEDGMENTS

The author thanks the various Neurology Fellows who provided excellent care to his patients; and Drs. Amjad Ilyas and Richard Quarles for help with immunochemistry of MAG and SGPG and gangliosides.

REFERENCES

1. Kieseier BC, Mathey EK, Sommer C, Hartung HP. Immune-mediated neuropathies. *Nat Rev Dis Primers*. 2018;4:31.
2. Dalakas MC. Pathophysiology of autoimmune polyneuropathies. *Presse Med*. 2013;42(6 Pt 2):e181–e192.
3. Dalakas MC. Pathogenesis of autoimmune neuropathies. *Biochim Biophys Acta*. 2015;1852:658–666.
4. Willison HJ, Jacobs BC, van Doorn PA. Guillain-Barré syndrome. *Lancet*. 2016; pii: S0140-6736(16).
5. Dalakas MC, Quarles RH. Autoimmune ataxic neuropathies (sensory ganglionopathies): are glycolipids the responsible autoantigens? *Ann Neurol*. 1996;39:419–422.
6. Cao-Lormeau V-M, Blake A, Mons S, et al. Guillain-Barre syndrome outbreak associated with Zika virus infection in French Polynesia: a case controlled study. *Lancet*. 2016;387(10027):1531–1539.
7. Dalakas MC. Guillain-Barré syndrome: the first documented COVID-19–triggered autoimmune neurologic disease: more to come with myositis in the offing. *Neurol Neuroimmunol Neuroinflamm*. 2020;7:e781. https://doi.org/10.1212/NXI.0000000000000781.
8. Costello F, Dalakas MC. Cranial neuropathies and COVID-19: neurotropism and autoimmunity. *Neurology*. 2020;95:1–2. https://doi.org/10.1212/WNL.0000000000009921.
9. Dalakas MC. Neurological complications of immune check-point inhibitors: what happens when you "take the brakes-off" the immune system. *Ther Adv Neurol Disord*. 2018;11:1–9.
10. Illa I, Ortiz N, Gallard E, et al. Acute axonal Guillain–Barré syndrome with IgG antibodies against motor axons following parenteral gangliosides. *Ann Neurol*. 1995;38:218–224.
11. Dalakas MC. Advances in the diagnosis, pathogenesis and treatment of CIDP. *Nat Rev Neurol*. 2011;7(9):507–517.
12. Bunschoten C, Jacobs BC, Van den Bergh PYK, et al. Progress in diagnosis and treatment of chronic inflammatory demyelinating polyradiculoneuropathy. *Lancet Neurol*. 2019;18:784–794.
13. Murata K, Dalakas MC. Expression of the co-stimulatory molecule BB-1, the ligands CTLA-4 and CD28 and their mRNAs in chronic inflammatory demyelinating polyneuropathy. *Brain*. 2000;123:1660–1666.
14. Dalakas M, Engel WK. Immunoglobulin deposits in chronic relapsing polyneuropathies. *Arch Neurol*. 1980;37:637–640.
15. Stathopoulos P, Alexopoulos H, Dalakas MC. Autoimmune antigenic targets at the node of Ranvier in demyelinating disorders. *Nat Rev Neurol*. 2015;11:143–156.
16. Querol L, Devaux J, Rojas-Garcia R, Illa I. Autoantibodies in chronic inflammatory neuropathies: diagnostic and therapeutic implications. *Nat Rev Neurol*. Sep. 2017;13(9):533–547.
17. Dalakas MC, Gooch C. Close to the node but still far: what antibodies tell us about CIDP and its therapies. *Neurology*. 2016;86:796–797.
18. Weiss MD, Luciano CA, Semino-Mora C, et al. Molecular mimicry in chronic inflammatory demyelinating polyneuropathy and melanoma. *Neurology*. 1998;51:1738–1741.
19. Beadon K, Guimarães-Costa R, Léger JM. Multifocal motor neuropathy. *Curr Opin Neurol*. 2018;31(5):559–564.
20. Dalakas MC. Advances in the diagnosis, immunopathogenesis and therapies of IgM-anti-MAG antibody-mediated neuropathies. *Therap Adv Neurol Disord*. 2018;11:1–12.
21. Latov N, Hays A, Sherman WH. Peripheral neuropathy and anti-MAG antibodies. *Crit Rev Neurobiol*. 1988;3:301–332.
22. Ilyas AA, Quarles RH, McIntosh TD, et al. IgM in a human neuropathy related to paraproteinemia binds to a carbohydrate determinant in the myelin-associated glycoprotein and to a ganglioside. *Proc Natl Acad Sci U S A*. 1984;81:12251229.

23. Ilyas AA, Quarles RH, Dalakas MC, et al. Polyneuropathy with monoclonal gammopathy: glycolipids are frequently antigens for IgM paraproteins. *Proc Natl Acad Sci U S A*. 1985;82:6697–6700.
24. Lombardi R, Erne B, Lauria G, et al. IgM deposits on skin nerves in anti-myelin-associated glycoprotein neuropathy. *Ann Neurol*. 2005;57:180–187.
25. Willison HJ, Trapp BD, Bacher JD, et al. Demyelination induced by intraneural injection of human antimyelin associated glycoprotein antibodies. *Muscle Nerve*. 1988;11:1169–1176.
26. Tatum AH. Experimental paraprotein neuropathy; demyelination by passive transfer of human IgM anti-MAG. *Ann Neurol*. 1993;33:502–506.
27. Ilyas AA, Gu Y, Dalakas MC, et al. Induction of experimental ataxic sensory neuronopathy in cats by immunization with purified SGPG. *J Neuroimmunol*. 2008;193:87–93.
28. Dispenzieri A, Kyle RA, Lacy MQ, et al. POEMS syndrome: definitions and long-term outcome. *Blood*. 2003;101:2496–2506.
29. D'Souza A, Lacy M, Gertz M, et al. Long-term outcomes after autologous stem cell transplantation for patients with POEMS syndrome (osteosclerotic myeloma): a single-center experience. *Blood*. 2012;120:56–62.
30. Golden EP, Vernino S. Autoimmune autonomic neuropathies and ganglionopathies: epidemiology, pathophysiology, and therapeutic advances. *Clin Auton Res*. 2019;29:277–288.
31. Dalakas MC, Pezeshkpour GH. Neuromuscular diseases associated with human immunodeficiency virus infection. *Ann Neurol*. 1988;23:38–48.
32. Liu X, Treister R, Lang M, Oaklander AL. IVIg for apparently autoimmune small-fiber polyneuropathy: first analysis of efficacy and safety. *Ther Adv Neurol Disord*. 2018;11. https://doi.org/10.1177/1756285617744484. 1756285617744484.
33. Kosmidis ML, Koutsogeorgopoulou L, Alexopoulos A, et al. Reduction of Intraepidermal Nerve Fiber Density (IENFD) in the skin biopsies of patients with fibromyalgia: a controlled study. *J Neurol Sci*. 2014; 347(1–2):143–147.
34. Tholance Y, Moritz CP, Rosier C, the anti-FGFR3 antibody Study Group. et al. Clinical characterisation of sensory neuropathy with anti-FGFR3 autoantibodies. *J Neurol Neurosurg Psychiatry*. 2020;91(1):49–57. https://doi.org/10.1136/jnnp-2019-321849. Epub 2019 Nov 5. PMID. 31690697.
35. Dalakas MC. Subcutaneous IgG for chronic inflammatory demyelinating plolyneuropathy. *Lancet Neurol*. 2018;17(1):20–21.

68

Immunologic Renal Diseases

Tilo Freiwald, Meryl Waldman, and Behdad Afzali

The immune system is a major pathological component of many human diseases. Kidney diseases are no exception, with compelling clinical, pathological, and experimental data implicating immune-mediated injury, particularly in diseases of the glomeruli.[1] Physiological immunity requires complex interactions between both innate and adaptive arms of the immune system and tissue-derived signals. Humoral factors, including complement, coagulation and inflammatory factors and their regulators, interplay with cellular processes, to alter cellular behavior, microvascular biology and recruit mediators of tissue healing and repair. Underlying this constellation are genetic risk factors that modulate these responses to predispose, or to protect against, nephritogenic responses and loss of kidney tissue.

Despite these insights, primary events directing/perpetuating immunological reactivity against kidney tissues remain poorly understood. The dearth of knowledge about the molecular mechanisms of immunological processes inciting and propagating kidney injury translates to a narrow armamentarium of poorly efficacious and nonspecific immunosuppressive therapies. These remain broad-spectrum and toxic, causing susceptibility to infection, cancer, metabolic and cardiovascular diseases. There are three reasons for the deficiencies in understanding of the immunological bases of human nephrology, namely over-reliance on animal models that partially resemble human disease, the unknown duration of disease prior to presentation, and categorization of kidney diseases according to structural changes rather than molecular level events occurring within cells and tissues. The advent of multi-omics technologies and machine learning approaches promises to address at least some of these deficiencies by integrating nucleic acid, protein structural information, and others, cross-sections in an unbiased manner across populations at different stages of disease.[2] Nephrologists are, accordingly, optimistic about the near future, anticipating significant progress.

In this chapter we focus mainly on glomerular diseases, for which some of the pathophysiological mechanisms are understood. Immune-mediated kidney injury can be incited by several mechanisms, ranging from normal immune responses causing incidental kidney injury (*e.g.*, postinfectious glomerulonephritis, serum sickness) to loss of tolerance to self-components, causing autoimmunity. Autoimmune responses can be directed against antigens shared between kidneys and extra-renal tissues (*e.g.*, antineutrophil cytoplasmic autoantibodies [ANCAs]) or directed specifically against kidney components (*e.g.*, anti-glomerular basement membrane [GBM] disease). Dysregulated immunity generating nephritogenic immune complexes (such as, in lupus nephritis) can also occur.[1] In some cases, we have uncovered candidate antigens inciting nephrogenic autoimmunity, such as phospholipase A2 receptor (PLAR2), antibodies against which cause some subsets of primary membranous nephropathy (MN).

In all these cases, current best practice to evaluate patients with immune-mediated kidney disease involves attention to urinalysis, the type and quantity of proteinuria, renal function, presence/absence of common immune biomarkers and ultrastructural examination of kidney tissues by biopsy. A renal ultrasound scan to identify structural lesions and to determine chronicity of injury (chronically damaged kidneys are often small, acutely injured kidneys are often normal size or slightly enlarged and may be echogenic) is often performed.

HEMATURIA

Red blood cells (RBCs) can enter the urine either via the upper or lower urinary tract. Upper urinary tract RBCs from glomeruli typically appear as dysmorphic *acanthocytes* and should prompt referral to nephrology for further evaluation. Lower urinary tract RBCs from the calices, ureters, bladder or urethra typically appear normal and are usually investigated by a urologist to rule out an occult malignancy.

Cellular casts in the urine are formed by aggregation around tubular proteins, such as Tamm-Horsfall protein. Erythrocyte and/or leukocyte casts typically indicate glomerular or tubular inflammation and, therefore, glomerulonephritis or interstitial nephritis, respectively (Fig. 68.1).

PROTEINURIA

Loss of the size-selective and/or charge-selective properties of the glomerular capillary wall or disruptions of glomerular epithelial cells, called podocytes, allows plasma proteins, especially albumin, to leak into the filtrate causing glomerular proteinuria. Tubulointerstitial nephropathy, on the other hand, impairs reabsorption of normally filtered low-molecular-weight proteins, resulting in low-grade proteinuria (rarely >2 g/day) with generally low albumin content. Thus, heavy proteinuria of greater than 2 g/day is often glomerular in origin or occurs because of frank blood in the urine.

Filtration of abnormal plasma proteins, termed paraproteins, can also cause proteinuria by overwhelming the reabsorptive capacity of renal tubules. This "overflow" proteinuria may be undetected by albumin-sensitive dipstick screening and may require urine immunofixation electrophoresis. Paraproteinemia may also be undetected or underestimated by plasma electrophoresis; measuring serum free light chains directly is a more sensitive method (Fig. 68.2).

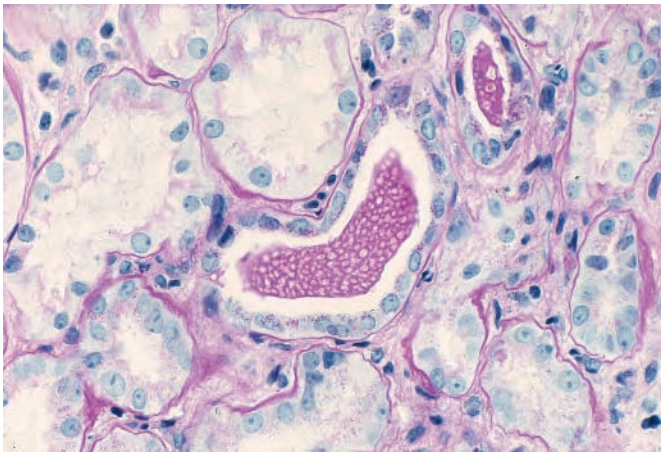


FIG. 68.1 Red Blood Cell Cast. Cast present *in situ* within the lumen of a distal renal tubule (periodic acid–Schiff [PAS] stain).

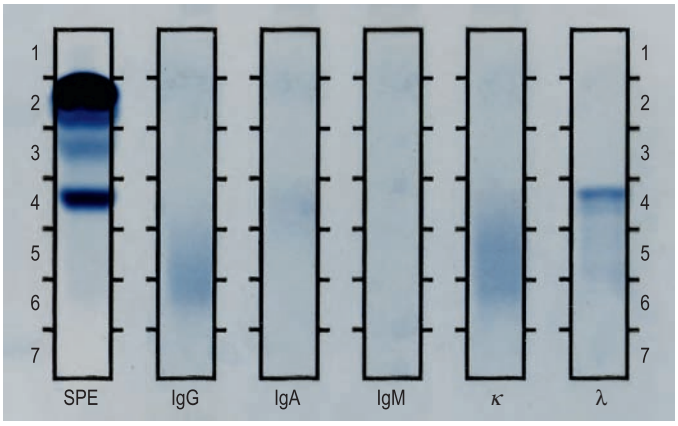


FIG. 68.2 Immunofixation Electrophoresis of Urinary Protein. Proteins in a concentrated urine protein are separated in six replicate lanes by standard protein electrophoresis. Separated proteins are identified by overlaying specific antisera to immunoglobulin *(Ig)*G, IgA, IgM, κ, and λ. In this example, a monoclonal paraprotein composed of λ light chain is identified as an intense narrow band in the far right lane. *SPE,* Serum protein electrophoresis.

NEPHROTIC SYNDROME

Nephrotic syndrome is clinically characterized by heavy proteinuria of greater than 3.5 g/day, hypoalbuminemia, edema, hyperlipidemia, and lipiduria. Conditions that typically have diffuse glomerular disease, including MN and systemic lupus erythematosus (SLE), are more likely than others with common focal disease, such as IgA nephropathy, to cause nephrotic syndrome. Diseases causing significant glomerular necrosis, such as ANCA-associated vasculitis, do not characteristically cause nephrotic syndrome since necrotic glomeruli lose filtration capacity.

ACUTE NEPHRITIC SYNDROME

Acute nephritic syndrome is glomerulonephritis characterized by glomerular hematuria, sub-nephrotic proteinuria, fluid retention and hypertension. A variant called rapidly progressive glomerulonephritis (RPGN) is defined by ≥50% loss of glomerular filtration rate over 3 months, and generally greater than 50% of glomeruli show severe injury, manifest as cellular crescents on biopsy.

TABLE 68.1 Indications for Renal Biopsy

1. Active "nephritic" urine sediment
   - Dysmorphic erythrocytes: >10 per high-power field
   - Cellular casts: erythrocyte or leukocyte
2. Proteinuria >2 g/day
3. Abnormal renal function
   - Associated with the above features of active nephritis
   - Particularly important if the duration of renal disease and/or rate of change are unknown
4. Document indications for use of high-risk therapeutic interventions

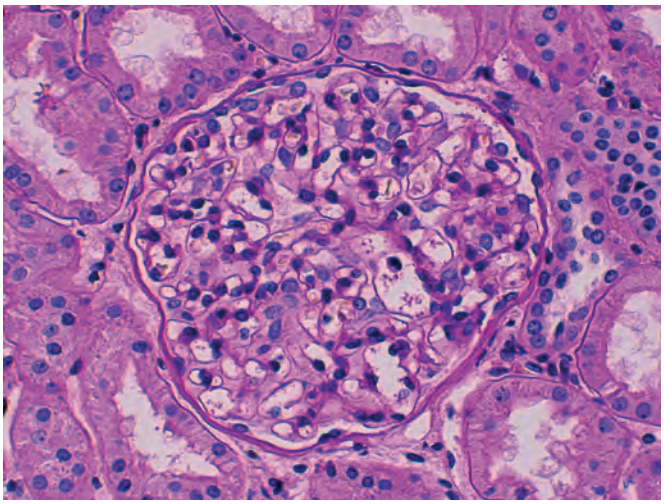


FIG. 68.3 Normal Glomerular Architecture. The glomerular capillary loops are patent and have normal thickness. Neither increased endocapillary cells nor expanded mesangial matrix encroach upon the patency of the capillary lumina (periodic acid–Schiff [PAS] stain).

RENAL BIOPSY

Renal biopsy is often indicated to establish or confirm a tissue diagnosis, clarify the type of renal injury, establish the stage of disease, formulate prognosis and direct therapy. Some important indications for renal biopsy are listed in Table 68.1. Light microscopy of a normal glomerulus is illustrated in Fig. 68.3. Some of the histological types of kidney injury will be discussed below.

MINIMAL CHANGE DISEASE

KEY CONCEPTS

Minimal Change Disease

- Most common cause of nephrotic syndrome in children
- High rate of response to glucocorticoids
- Cyclophosphamide is useful for those who relapse frequently
- Renal prognosis is characteristically excellent
- A subset may have or progress to unsampled focal segmental glomerulosclerosis

Childhood nephrotic syndrome is mainly caused by minimal change disease (MCD). MCD in adults with nephrotic syndrome is rarer.[3]

Clinical Features

MCD is almost always associated with precipitous and severe nephrotic syndrome without a systemic disease. Microscopic hematuria is not uncommon in adult patients. Proteinuria exclusively comprising albumin is characteristic; standard immunological screening tests are usually normal. Kidney biopsy establishes a diagnosis of MCD.

Etiology and Pathogenesis

The basic lesion of MCD is loss/neutralization of anionic proteoglycans in glomerular capillary loops. Dissipation of the negative-charge barrier allows anion-charged albumin to pass freely. Interdigitating podocyte foot processes form a second barrier to glomerular leakage of protein, called *slit diaphragms*. These podocyte foot processes and slit diaphragms are disrupted in MCD.

The cause of MCD remains idiopathic. There is evidence of immune dysregulation, involving cell-mediated immunity, supported by a tendency of MCD to become manifest or to relapse after a viral infection (*e.g.*, measles) or an allergic reaction. Some cases occur in association with Hodgkin lymphoma. MCD typically has a favorable response to immunosuppressants. Likewise, MCD relapses are associated with reduced regulatory T cells (Tregs).

Another hypothesis regarding causation of MCD invokes a circulating immune factor that causes increased glomerular permeability[4] (see "Focal Segmental Glomerulosclerosis"). Since patients respond to B-cell depletion with rituximab (a cytolytic monoclonal antibody against CD20 on B cells), a B-cell permeability factor cannot be ruled out.

Pathology

Generally, MCD biopsies appear normal by light microscopy. Electron microscopy shows characteristic pathological lesions of podocyte foot process fusion diffusely around glomerular capillaries.

Treatment

MCD is sensitive to glucocorticoids; greater than 90% of children enter remission within a few weeks of starting treatment. In adults the response to glucocorticoids is lower and more delayed. A substantial portion of those with MCD suffer long-term complications: some are steroid-resistant from the start; others are steroid-dependent for disease control; others still relapse frequently and suffer substantial steroid toxicity over time. Alkylating agents such as cyclophosphamide increase response rates and reduce relapse rates. The calcineurin inhibitors cyclosporine (CSA) and tacrolimus (Tac) are alternative steroid-sparing agents, but relapses frequently occur with withdrawal of these agents. Rituximab, mycophenolate mofetil (MMF) or azathioprine are alternative steroid-sparing agents, particularly in patients who are steroid-dependent or relapse frequently.

Risk of progression to end-stage kidney disease is extremely low in true MCD. Sampling errors on kidney biopsy account for mistaken diagnosis of MCD in some cases that are, in fact, focal segmental glomerulosclerosis (FSGS). It is debated whether MCD and FSGS represent different manifestations of one disease (with MCD progressing to FSGS), or if they are different diseases.

FOCAL SEGMENTAL GLOMERULOSCLEROSIS

KEY CONCEPTS

Focal Segmental Glomerulosclerosis

- Nephrotic syndrome with progressive renal insufficiency
- Glomerular permeability factor in plasma of some cases
- Unpredictable responses to glucocorticoids or cyclophosphamide
- Cyclosporine effective, but relapses common upon withdrawal
- Moderately high relapse rate in renal allografts

FSGS is a common primary glomerular disease in adults. Compared to patients with MCD, those with FSGS have higher frequency of microscopic hematuria, more persistent nephrotic syndrome, poorer response to immunosuppressants, and higher risk of progression to end-stage kidney disease. Incidence of FSGS as a cause of nephrotic syndrome is increasing, particularly among Black patients.

Etiology and Pathogenesis

Diverse etiologies may disrupt the podocytes, causing histological FSGS. Genetic mutations of the podocyte proteins podocin and nephrin cause a subset of FSGS in children and young adults. Circulating permeability factors, including soluble urokinase-type plasminogen activator receptor, have been proposed as mediators of podocyte injury in "primary" FSGS. Alternatively, podocytes may have increased surface expression of the transmembrane protein B7-1 (CD80) that stimulates T cells, leading to foot process effacement. "Secondary" FSGS may result from drug toxicities (*e.g.*, pamidronate), viral infection (*e.g.*, human immunodeficiency virus [HIV]), or maladaptive hemodynamic stress (*e.g.*, obesity or reduced nephron mass). A higher incidence of FSGS in African Americans is partially related to their increased frequency of *APOL1* gene polymorphisms, which may have been positively selected due to a protective role against trypanosomiasis.[5]

Pathology

The renal pathology of FSGS, as the name implies, is sclerosis of portions, or segments, of only some glomeruli in the sampled area. Segmental podocyte damage and detachment are seen, together with irregular foot process fusion, collapse of glomerular capillaries and marked increases in matrix and collagen accumulation. Segmentally sclerotic areas typically stain nonspecifically for IgM and C3 (but not immunoglobulin [Ig] G or IgA), particularly in areas of glomerular tuft hyalinosis (representing trapped plasma constituents), which do not represent classic immune complexes. One classification scheme divides FSGS into five variants: tip, perihilar, cellular, collapsing, and not otherwise specified[6] (Fig. 68.4). Collapsing FSGS, is associated with viral infections, notably HIV, and follows an aggressive course.

Treatment

Hypoalbuminemia and nephrotic range proteinuria are common in primary, but not secondary, FSGS, which typically cause sub-nephrotic proteinuria, with normal or near-normal serum albumin, no edema, and focal, rather than diffuse, foot process

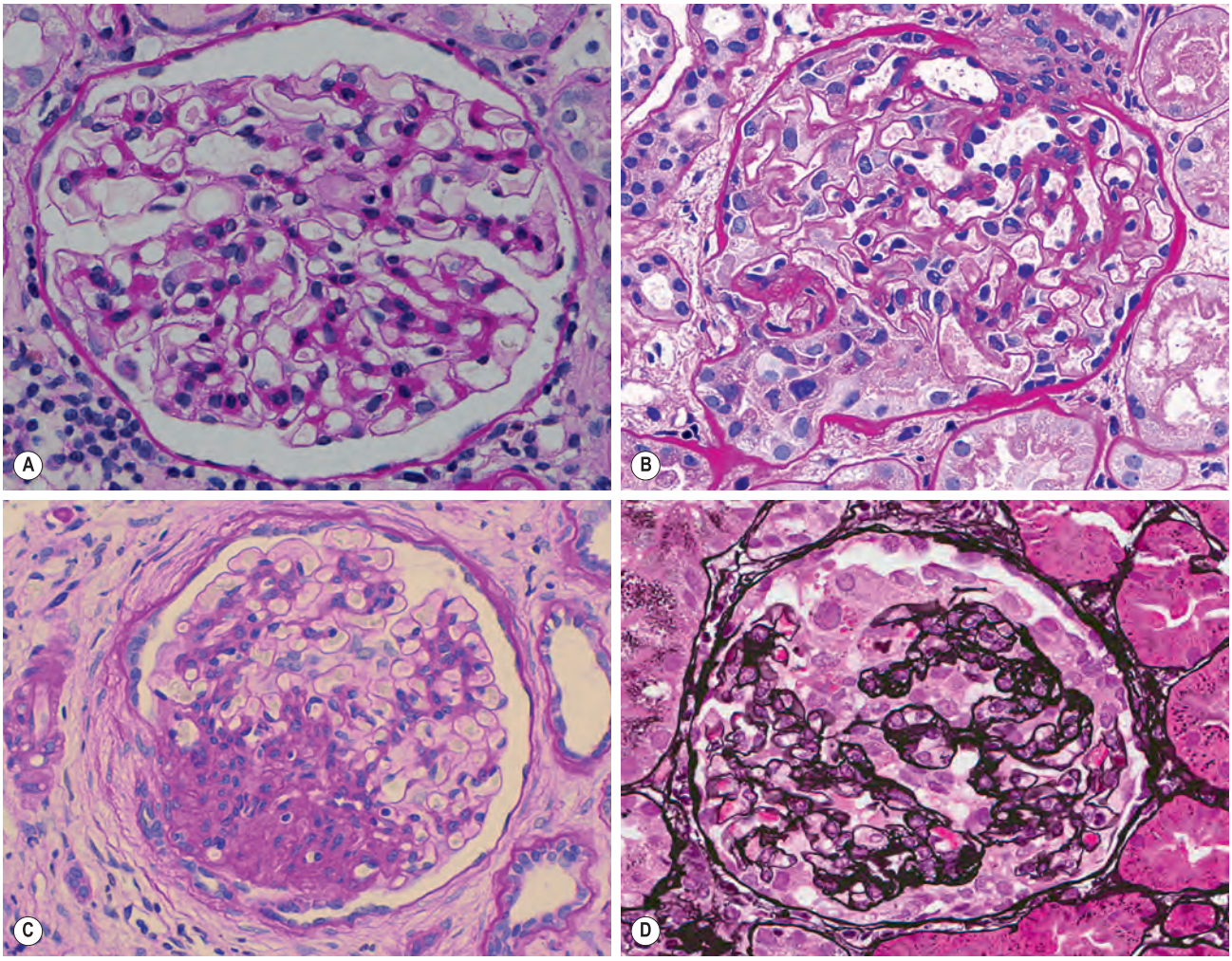


FIG. 68.4 Focal Segmental Glomerulosclerosis (FSGS). Several forms of glomerular lesions are seen in FSGS, often within the same biopsy: (A) minimal abnormality (periodic acid–Schiff [PAS] stain); (B) tip lesion manifested by segmental glomerular tuft lesion near the origin of the proximal tubule (PAS stain); (C) classical perihilar lesion (PAS stain); and (D) collapsing glomerulopathy; the glomerulus is globally contracted with wrinkling of the basement membranes; this is associated with hyperplasia of podocytes surrounding the glomerular capillaries (methenamine silver stain).

effacement. Genetic forms of FSGS should be considered in cases occurring at a very young age and in patients with steroid-resistance.

Treatment of genetic and hyperfiltration-induced FSGS focusses on reno-protection with angiotensin antagonists and lipid-lowering agents. Treatment of other forms of FSGS is similar to MCD, including immunosuppression with prednisone, cyclophosphamide, calcineurin inhibitors, or mycophenolate mofetil. There are mixed outcomes with rituximab in FSGS, particularly for steroid-resistant cases.[7] Complete remission of proteinuria is less common in FSGS than MCD. Relapses and progression to end-stage kidney disease (ESKD) remain a major concern, particularly in patients who are steroid-resistant and who frequently relapse. Several novel therapies targeting immunological, inflammatory, and costimulatory pathways are currently under investigation, including blockade of CD80 with cytotoxic T lymphocyte antigen 4-Ig (CTLA-4-Ig). Primary FSGS is at high risk of recurrence in patients receiving kidney transplants, reflecting the presence of an injurious circulating factor.

MEMBRANOUS NEPHROPATHY

KEY CONCEPTS

Membranous Nephropathy

- Common cause of nephrotic syndrome in adults
- Autoantibodies to several proteins, including phospholipase A_2 receptor (PLA_2R), detectable in large percentage of patients with primary MN
- Several secondary causes: SLE, drugs, chronic hepatitis, malignancies
- One-third of patients have spontaneous remission
- One-third of patients develop end-stage renal disease within a decade
- Protracted nephrotic syndrome confers risks of cardiovascular and thromboembolic events
- Therapies: steroids, alkylating agents, calcineurin inhibitors, rituximab, lipid-lowering drugs, angiotensin antagonists

MN is present in ~20% of adults with nephrotic syndrome. It is a less frequent cause of nephrotic syndrome in children. Primary MN was traditionally a diagnosis of exclusion after

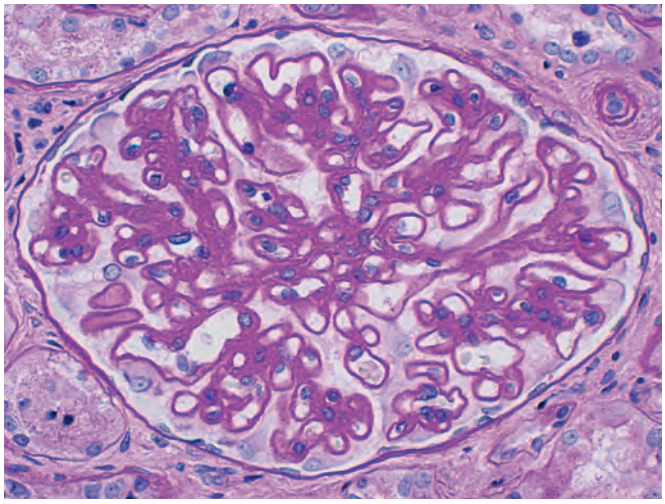


FIG. 68.5 Membranous Nephropathy. Capillary walls are nearly uniformly thickened but remain widely patent. Cellularity of the glomerulus is normal (periodic acid–Schiff [PAS] stain).

considering drugs, infections, neoplasms, and systemic illnesses (*e.g.*, SLE). Many cases of primary MN are now known to have antibodies against defined antigens, such as M-type phospholipase A2 receptor (PLA_2R), permitting diagnosis by screening for these antibodies.

Etiology and Pathogenesis

MN is characterized by subepithelial "*epimembranous*" immune deposits containing IgG and complement. Immune complexes are likely formed *in situ* by interaction between a pathogenic antibody and a constitutive glomerular antigen or an antigen ectopically planted in the glomerulus. The majority (70% to 80%) of patients with primary MN have circulating autoantibodies to PLA_2R, a receptor expressed in glomerular podocytes.[8] Antibodies to thrombospondin type-1 domain containing 7 A protein (THSDA7A), neutral endopeptidase and neural epidermal growth factor-like 1 protein (NELL-1) have been identified in smaller subsets of patients.[9,10]

Clinicopathological Features

MN generally presents with nephrotic syndrome. Renal biopsy is usually required, although serological testing for autoantibodies (*e.g.*, anti–PLA_2R) may be informative if renal biopsy is contraindicated. Light microscopy shows uniform thickening of the glomerular capillary walls without endocapillary cell proliferation (Fig. 68.5). Subepithelial and/or intramembranous deposits are seen on electron microscopy (Fig. 68.6). PLA_2R in glomerular immune deposits (by immunofluorescence or immunohistochemistry) favors a diagnosis of primary MN. A good proportion of secondary MN cases have mesangial deposits. However, diagnosing secondary MN relies on identifying clinical risk factors and finding abnormalities in laboratory and radiological data.

Natural History

The course of primary MN is highly variable. On average, one-third of patients remit spontaneously, one-third achieve partial remission and one-third progress to ESKD. Persistent proteinuria and impairment of renal function are long-term features of the latter two-thirds of patients.

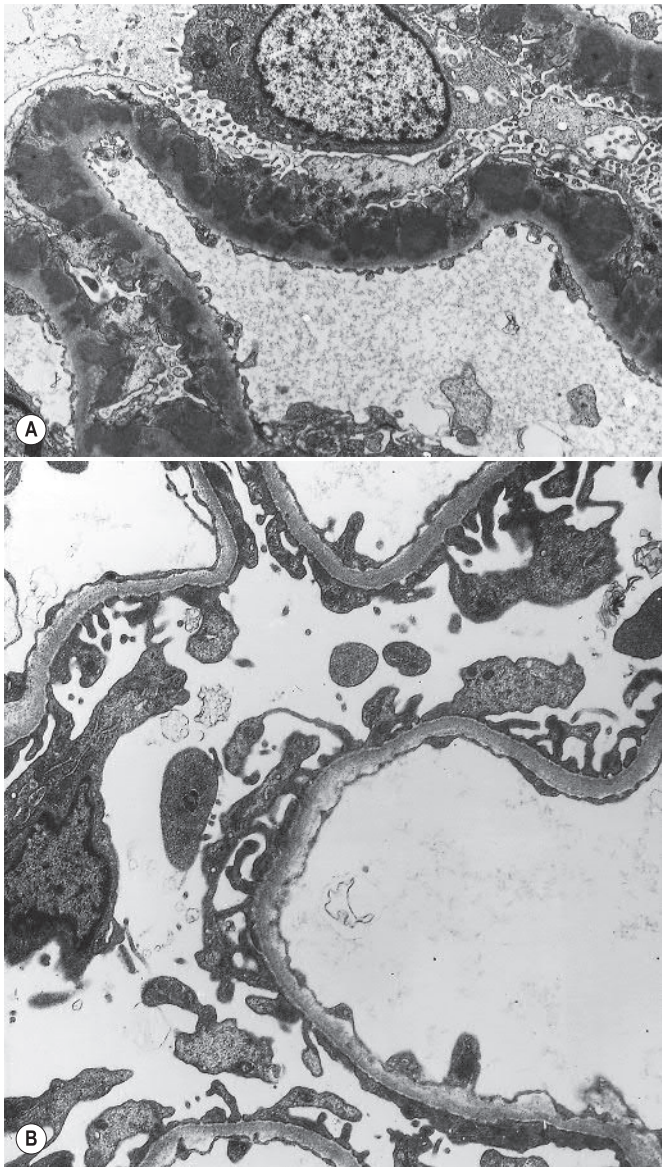


FIG. 68.6 Membranous Nephropathy (Ultrastructure). (A) Electron micrograph demonstrates heavy, dark-staining immune complex deposits along the outer surface of the glomerular basement membrane and beneath the epithelial foot processes (hence the terms *subepithelial* or *epimembranous deposits*). Note the thickening and projections of the gray-staining basement membrane between the electron-dense deposits. (B) Ultrastructure of a normal glomerular capillary wall for comparison.

Baseline characteristics, for example, severe nephrotic syndrome, hypertension, low GFR, are associated with poor outcomes. Protracted high-grade nephrotic-range proteinuria is a strong predictor of adverse renal outcomes. Some studies suggest that baseline PLA_2R autoantibody titers correlate with disease activity, with high titers a being a risk factor for decline of renal function.[11]

Treatment

Management of MN usually includes diuretics to reduce edema, lipid-lowering drugs, anticoagulants to prevent

thromboembolic complications, and anti-hypertensives. Angiotensin antagonists have a substantial antiproteinuric effect. All patients receive combinations of these measures. Low baseline, or decreasing, anti-PLA_2R antibody levels may predict spontaneous remission, favoring conservative therapy for 3 to 6 months, whereas high baseline, or increasing, anti-PLA_2R antibody levels encourage the rapid start of immunosuppression. Therapy with corticosteroids and cytotoxic drugs (usually cyclophosphamide) significantly increase rates of remission and slow renal function loss in patients with persistent nephrotic syndrome. There is increasing enthusiasm for first line use of rituximab based on two controlled trials demonstrating its superiority over anti-proteinuric therapy alone or cyclosporine.[12] Other anti-B-cell drugs are under evaluation, including belimumab.

MEMBRANOPROLIFERATIVE GLOMERULONEPHRITIS

KEY CONCEPTS

Membranoproliferative Glomerulonephritis

- Histologically classified into immune complex–mediated glomerulonephritis or complement-mediated glomerulonephritis based on immunofluorescence staining pattern
- C3 glomerulopathy characterized by C3 accumulation in the glomeruli in the form of electron-dense deposits.
- Genetic or acquired abnormalities in the activation of the alternate alternative complement pathway associated with C3 glomerulopathy
- Response to immunosuppressive drug treatment generally poor
- Tends to recur in renal allografts

MPGN is a morphological entity encompassing a heterogeneous group of diseases with similar appearance on light microscopy. The pattern of injury is mesangial matrix expansion, increased cellularity and thickening of the glomerular capillary walls, giving them a double contour appearance. These changes make the glomerulus appear lobulated (Fig. 68.7).

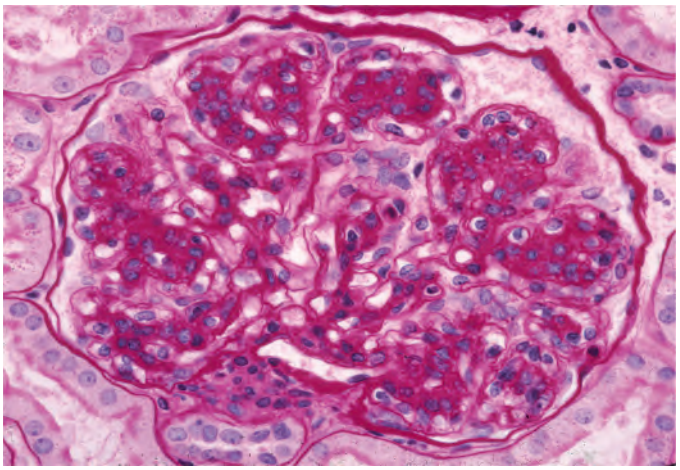


FIG. 68.7 Membranoproliferative Glomerulonephritis (MPGN). Glomerulus exhibits the typical lobulated appearance of this disease. Markedly increased mesangial cells and matrix in all of the lobules. Mesangium extends outward into the capillary loops and forms double contours with the glomerular basement membrane (periodic acid–Schiff [PAS] stain).

Etiology and Pathogenesis

Until recently MPGN was classified into three types: type 1 characterized by subendothelial deposits; type II by intramembranous electron dense deposits in a ribbon-like pattern (also called dense deposit disease [DDD]); and type III by subendothelial and subepithelial deposits (Fig. 68.8). This older classification also distinguished between secondary causes when an etiology was identifiable.

New insights into MPGN have resulted in re-classification of this disease into only two sub-types, Ig-mediated or

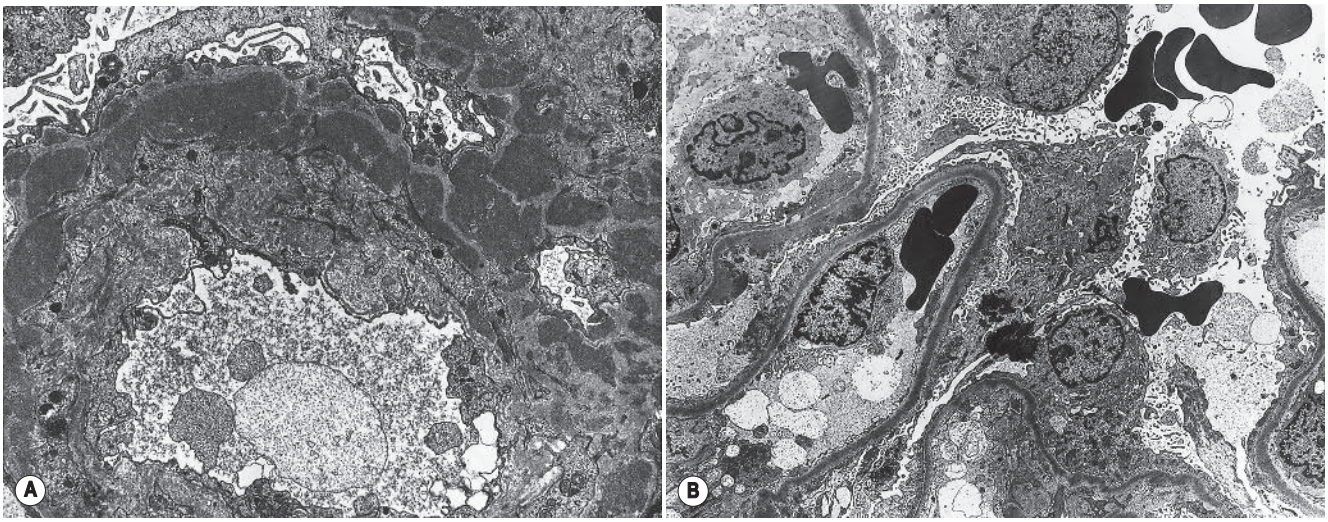


FIG. 68.8 Membranoproliferative Glomerulonephritis (MPGN) Ultrastructure. (A) Capillary wall is markedly thickened and contains heavy, dark-staining, electron-dense immune complexes in the subendothelial space. Mesangium (lighter material) extends into the capillary loop, where it is interposed between the basement membrane and the endothelium; the process gives the appearance of a massively thickened capillary loop on hematoxylin and eosin (H&E) staining, and the split appearance by periodic acid–Schiff (PAS) and silver stains. (B) Dense deposit disease; capillary loops contain smooth, continuous, linear dense material within the basement membrane.

complement-mediated disease, based on patterns and compositions of deposits assessed by immunofluorescence. This recognizes that some cases of MPGN result from immunoglobulin (Ig) deposition with secondary complement activation, whereas others arise from primary abnormalities of complement activation. The presence of both Ig and complement component 3 (C3) indicates an Ig-mediated process, in which immune complexes are deposited that activate the classical complement pathway. In contrast, C3 staining without significant Ig deposition implies an antibody-independent means of complement activation, indicating primary dysregulation of the alternative complement pathway (Fig. 68.9). Such diseases are now grouped under the umbrella term, "C3 glomerulopathy" (C3G), which encompasses both C3 glomerulonephritis (C3GN) and DDD.[13] A diagnosis of C3G prompts evaluation for genetic mutations of complement regulatory proteins (*e.g.*, factor H or factor I) or acquired autoantibodies to regulatory proteins (*e.g.*, C3 nephritic factor or anti-factor H). C3 nephritic factor, an autoantibody to C3 convertase, is frequently detected in C3G. C3 nephritic factor stabilizes C3 convertase, rendering it resistant to control by factor H, leading to persistent C3 activation and deposition of alternative pathway activation products in glomeruli.

The list of etiologies causing MCGN is long but includes infections (*e.g.*, with hepatitis C virus, see below), monoclonal gammopathies, autoimmune diseases (especially SLE) and complement disorders.

Pathology

MPGN shows endocapillary proliferation and positive findings on immunofluorescence. Double contour appearance to the GBM, best seen with silver stain, represents synthesis of GBM-like material from capillary wall remodeling. Irregular capillary wall subendothelial and mesangial deposits are seen by electron microscopy in Ig-mediated MPGN; small intramembranous and subepithelial deposits are also present in C3GN. The hallmark of C3GN is dominant staining for C3 by immunofluorescence in the mesangium and glomerular capillary walls (see Fig. 68.9).

Double contours of the GBM are also present in conditions with chronic endothelial injury, including thrombotic microangiopathy, transplant glomerulopathy, and preeclampsia, and show histological appearance of MPGN by light microscopy. However, in these settings there are no associated immune deposits and immunofluorescence is negative for Ig and C3.

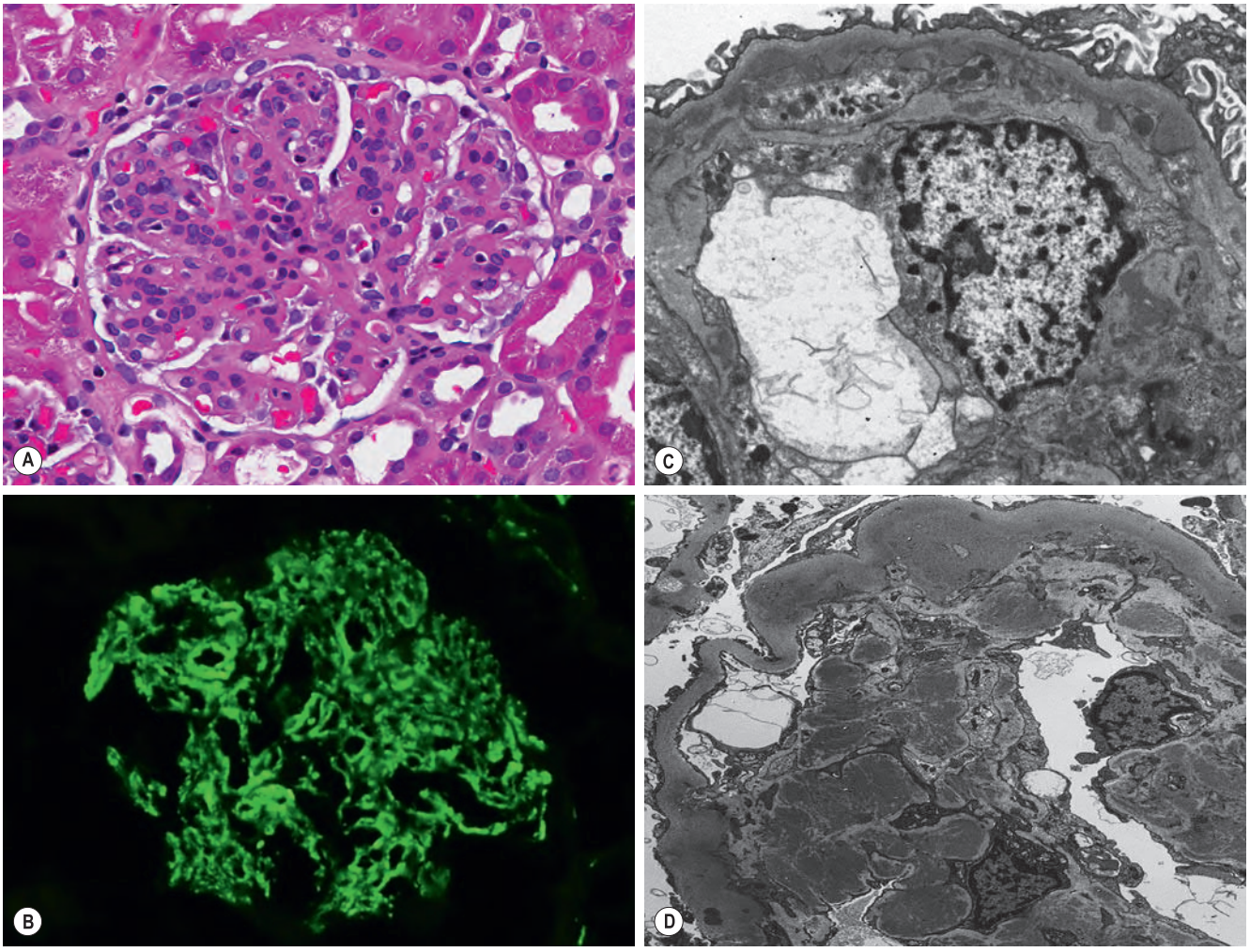


FIG. 68.9 C3 Glomerulopathy. (A) Glomerulus exhibiting the characteristic membranoproliferative pattern (hematoxylin and eosin [H&E] stain). (B) Immunofluorescence showing prominent glomerular staining for C3 complement. (C and D) Electron microscopy ultrastructure showing mesangial interposition giving a double contour of the glomerular capillary wall, as well as interposed electron dense deposits in the subendothelial space and the mesangium.

Clinical Presentation

Presentation of MPGN is characteristically as chronic low-grade nephropathy. Some patients experience nephrotic syndrome and even RPGN. Hypocomplementemia is common. In immune complex-mediated MPGN, serum C4 levels are generally low, reflecting classical pathway activation. C3 may be normal or mildly decreased. In complement-mediated MPGN C4 levels are normal, and C3 is typically low, consistent with alternative pathway activation. However, a normal serum C3 does not exclude the diagnosis. Renal outcomes in C3GN are variable, with up to 30% of patients progressing to ESKD. Prognosis for DDD is worse, with ~50% progressing to kidney failure. Idiopathic MPGN and MPGN from complement dysregulation tend to recur in renal allografts; recurrences of both types of MPGN worsen graft survival.

Treatment

Existing treatment trials for MPGN pre-date the new classification system and our appreciation of the distinct mechanisms of pathogenesis. Thus, their relevance is questionable because of the heterogeneity of patient populations. The new diagnostic categories will better define groups of patients with similar pathogenesis, enabling more rational approaches to targeted therapy in the future.

Management of MPGN relies on identifying any reversible/treatable underlying causes, such as hepatitis C virus (HCV) infection. Treatment of the underlying cause can reverse the kidney pathology and eliminate risk of aggravating occult etiologies, such as viral infection, with immunosuppression.

There is no consensus regarding treatment of "idiopathic" immune complex–mediated MPGN. Most immunosuppressants have been used with variable success, including prednisone alone or combined with either MMF or cyclophosphamide. There are limited data on rituximab for this disease. Optimal treatment of C3GN remains unclear but likely needs to be individualized based on correct identification of defects in the alternative complement system (genetic vs acquired autoantibodies). Infusions of fresh frozen plasma to replace missing complement factors may be beneficial for some. Immunosuppression with corticosteroids, rituximab, and MMF is theoretically beneficial in cases resulting from pathogenic antibodies to complement regulatory proteins, but is unproven. Eculizumab, a mAb blocking activation of C5 complement (C5b-9), has been tested in small series, but data remain inconclusive. New complement blocking agents may also hold promise.[14]

POSTINFECTIOUS NEPHROPATHIES

Many infections can cause immune-mediated nephropathies. Some of the more common ones are described here.

Viral Infections

Hepatitis B Virus

Hepatitis B virus (HBV) has several histological renal manifestations, the commonest being MN, which can occur in association with antibodies against PLA_2R,[15] MPGN, and vasculitides affecting the kidneys (causing RPGN or large vessel involvement). Clinical presentation is usually nephrotic syndrome, often with microscopic hematuria. Renal biopsy shows MN with positive immunofluorescence for Ig, C3, and some IgM. In some cases, HBV viral antigens are detectable within glomeruli. Electron microscopy shows subepithelial and intramembranous deposits, but there may also be mesangial and subendothelial deposits.

Therapy for HBV-associated kidney disease is focused on treating the virus using antivirals, because immunosuppressants may promote viral replication. Interferon-α (IFNα) or nucleoside/nucleotide analogues, such as lamividine, entacavir, adefovir, tenofovir, and telbivudine are commonly used. Treatment with nucleoside/nucleotide analogues is often for several years. Immunosuppression with corticosteroids with or without cyclophosphamide or rituximab is cautiously considered in patients with RPGN, but concomitant antiviral therapy is required.

Hepatitis C Virus

HCV is an established cause of kidney disease. The virus has a similar spectrum of pathogenicity as HBV, but the most common manifestation is MPGN, occurring in the context of type II mixed cryoglobulinemia, in which mixtures of antibodies of different classes and clonalities precipitate at temperatures below 37°C. Immune complex deposits within glomeruli are seen containing HCV, anti-HCV antibodies and virus-related (or unrelated) cryoglobulins. HCV also causes MN and vasculitides affecting the kidneys.

Treatment of HCV-associated nephropathy is aimed at the underlying cause with antiviral drugs. Prognosis has greatly improved with introduction of newer antiviral agents supplanting IFNα and ribavirin. HCV-specific antivirals are usually used in combinations, such as sofosbuvir plus ledipasvir, which are highly effective and in most cases curative.[16] Evidence for antiviral treatment in HCV-related kidney disease comes mostly from IFN-based regimens, which reported remission of proteinuria and hematuria and improvement of renal function. There are limited data regarding the use of the newer antiviral agents in HCV-associated glomerulonephritis, but these hold promise.[17] In severe progressive renal disease and/or vasculitis, treatment with rituximab or pulse intravenous steroids and cyclophosphamide may be warranted in conjunction with antivirals. Rituximab and plasma exchange may provide additional benefit in severe HCV-associated cryoglobulinemia refractory to antiviral therapy.[18]

Human Immunodeficiency Virus

HIV infection may cause a number of kidney manifestations. Chief is classical HIV-associated nephropathy (HIVAN), the first kidney disease associated with HIV infection. HIVAN is a collapsing form of FSGS accompanied by tubular microcysts and interstitial inflammation. It is usually seen in patients with AIDS but is occasionally diagnosed in less advanced HIV infection or even before acute HIV seroconversion. HIVAN classically has significant proteinuria and rapidly progressive kidney disease, often with hypertension and echogenic kidneys on ultrasound. HIVAN displays a striking racial predilection for Black patients, suggesting that genetic influences may be important.[19]

Other renal abnormalities have also been described with HIV, including IgA nephropathy, lupus-like glomerulonephritis, postinfectious glomerulonephritis, MPGN, MN, cryoglobulinemic glomerulonephritis, fibrillary and immunotactoid

glomerulopathy, and thrombotic microangiopathy.[20] Tubulointerstitial changes related to drug toxicity, acute interstitial nephritis, or superimposed viral, fungal, or mycobacterial infections may also be present.

The mainstay of treatment for HIVAN is antiretroviral therapy (ART), regardless of CD4 lymphocyte counts. Highly effective therapies for HIV have reduced the frequency of HIVAN and greatly improved the renal prognosis. Evidence for initiating ART in other HIV-associated immune complex glomerular diseases is inconclusive, but its use is a rational approach. Use of standard immunosuppressive drugs in this immunocompromised population is controversial.

Bacterial Infections

Poststreptococcal Glomerulonephritis

KEY CONCEPTS

Infection-Related Nephropathies

- *Viral:* Hepatitis B—membranous nephropathy; hepatitis C—cryoglobulinemic membranoproliferative glomerulonephritis; human immunodeficiency virus (HIV)—focal segmental glomerulosclerosis (HIV nephropathy)
- *Bacterial (mainly gram-positive):* Nephritogenic streptococcal infections, prosthetic device (shunt) infections, subacute bacterial endocarditis, chronic deep tissue abscesses—mainly diffuse or focal proliferative glomerulonephritis

Poststreptococcal glomerulonephritis (PSGN) is a classic immune complex–mediated glomerulonephritis caused by dermal or throat infection with nephritogenic strains of group A streptococci. PSGN has bimodal epidemiology, with peak incidence in children and in those over the age of 60. Latency between upper respiratory infection and nephritis is 7 to 10 days, and 2 to 4 weeks after skin infection. Anti-streptococcal antibody titers are usually measured to demonstrate a preceding streptococcal infection. Anti-streptolysin O titers and anti-DNase B titers are most frequently elevated in upper respiratory and skin infections, respectively.

PSGN is a nephritic syndrome, with smoky or rust-colored urine, generalized edema, hypertension, and nephritic urine sediment. Proteinuria is typically mild. Patients have rising titers of anti-streptolysin and depressed C3 early in nephritis but normal/minimally depressed C4, indicating alternative complement pathway activation. Significant numbers of patients also have rheumatoid factor. Other auto-reactive antibodies have also been described in PSGN.

Glomerular injury results from passive glomerular trapping of circulating immune complexes composed of nephritogenic bacterial antigens and IgG antibody, or by *in situ* formation of immune complexes. This is followed by immune cell recruitment, production of chemical mediators and cytokines, and local activation of complement and coagulation cascades that drive inflammatory responses.

Proliferative glomerulonephritis, with leukocyte infiltration, granular immune deposits of IgG and C3, and dome-shaped electron-dense subepithelial deposits (humps) are characteristic. Prognosis is excellent in children; with supportive care, almost all recover. Progressive renal failure accompanied by hypertension is more common in adults. Kidney biopsy is rarely needed in children but may be warranted if there is atypical presentation or evolution.

The classic childhood form is rarely seen in developed countries but is common in developing countries, thus remaining the commonest cause of acute childhood nephritis in the world. However, there is also an increase in incidence of non-streptococcal PSGN or "infection-related" glomerulonephritis in older patients with multiple comorbidities, especially diabetes, HIV and malignancy. These clinical variants are usually related to infections with bacteria such as *Staphylococcus aureus*, irrespective of methicillin-sensitivity, and may be characterized by IgA-dominant glomerular immune complex deposition.

Current therapies rely on culture-guided systemic antibiotics, especially in older patients, in whom MRSA may be the causative agent. Steroids are used in selected cases, in which crescents and severe interstitial inflammation are present.

IgA Nephropathy

KEY CONCEPTS

Immunoglobulin A Nephritis

- Common cause of asymptomatic microscopic hematuria, recurrent macroscopic hematuria, and/or low-grade proteinuria
- Spectrum of disease, including idiopathic IgA nephritis and IgA vasculitis (previously called Henoch-Schönlein purpura); IgA in skin and renal biopsy samples
- Mostly benign prognosis, especially in children
- Patients with progressive renal insufficiency and/or crescentic glomerulonephritis warrant trial of glucocorticoids and/or cytotoxic drug therapy

IgA nephropathy (IgAN) is the most common primary glomerulonephritis worldwide.[21] IgAN can affect patients of all ages, especially children and young adults, with a male preponderance.[22] There are geographical and ethnic differences in the prevalence of IgAN, with the highest frequency found among East Asians. IgAN is very uncommon in individuals of African ancestry. This observation and examples of familial clustering of IgA nephropathy favor an important element of genetic susceptibility.

The spectrum of presentations of IgAN is relatively broad. IgAN may be discovered during evaluation of asymptomatic microscopic hematuria. Alternatively, patients, especially children, can present with recurrent episodes of macroscopic hematuria that occur within 24 to 48 hours after an infection, usually an upper respiratory infection or gastroenteritis. A transient elevation in serum creatinine has been associated with macroscopic hematuria in about one third of cases. This has been attributed to tubular injury and obstruction caused by intraluminal RBC casts. A small percentage of patients present with either nephrotic syndrome or an acute RPGN characterized by edema, hypertension, renal insufficiency, and hematuria.

IgA vasculitis (IgAV), also known as Henoch-Schönlein purpura, which typically involves the gastrointestinal and dermatological systems, may also affect the kidneys and present with features similar to IgA nephropathy.

Pathology

The characteristic findings on light microscopy are mesangial cell proliferation and mesangial matrix expansion. Electron microscopy typically reveals electron-dense deposits that primarily limited to the mesangium, but a few deposits may also be present in subendothelial and subepithelial locations. The

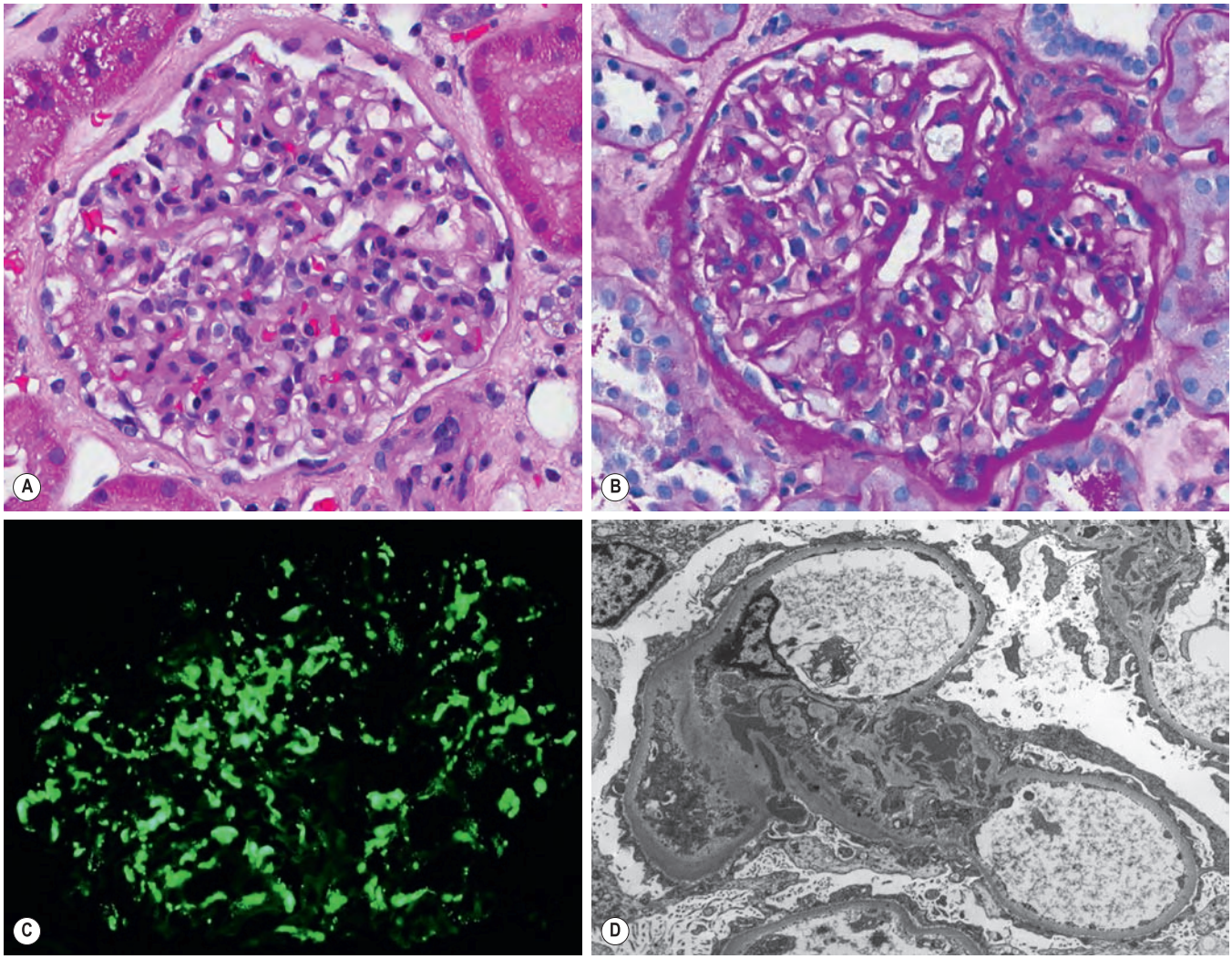


FIG. 68.10 Immunoglobulin A Nephropathy (IgAN). (A) hematoxylin and eosin (H&E) stain and (B) periodic acid–Schiff (PAS) stain: Glomeruli manifest modestly increased mesangial hypercellularity and expanded mesangial matrix. (C) Immunofluorescence showing predominantly IgA deposition within the mesangium. (D) Electron microscopy showing deposits mainly within the mesangium, but not along peripheral capillary loops.

pathognomonic finding on immunofluorescence microscopy is globular deposits of IgA, often accompanied by C3 and IgG, in the mesangium and, to a lesser degree along the glomerular capillary wall (Fig. 68.10).

Etiology and Pathogenesis

Aberrant glycosylation of *O*-linked glycans in the hinge region of IgA1 resulting in increased serum levels of galactose-deficient IgA1 (Gd-IgA1) plays a pivotal role in pathogenesis of IgAN. The aberrantly glycosylated IgA1 is recognized by antiglycan antibodies and leads to circulating IgA immune complexes that preferentially deposit in the mesangium, provoking local injury. A genetic predisposition to IgAN has been linked with polymorphisms involving both innate and adaptive immunity and the alternative complement pathway. A "second hit" may be needed in predisposed individuals. Infections may play a role because episodes of macroscopic bleeding often coincide with mucosal infections, including upper respiratory tract (synpharyngitic) or gastrointestinal infections.

Natural History

Patients with IgAN who have low-grade proteinuria (<1 g/day) have a good renal prognosis and a low risk of progression. However, at least one-third of patients with IgAN eventually progress to ESKD. Twenty years after apparent disease onset the probability of renal failure is 25%, and the probability of some renal dysfunction is 50%. Hypertension occurs frequently as the disease progresses and portends a poor prognosis. Other clinical features associated with poor prognosis include older age at disease onset, persistent proteinuria greater than 1 g/day, and persistent azotemia.

Treatment

The most effective treatment of progressive IgAN remains undefined and controversial. Angiotensin antagonists are recommended to control blood pressure, reduce proteinuria, and slow the rate of deterioration of renal function. A large trial (STOP-IgA nephropathy) showed that glucocorticoid monotherapy or a regimen that included prednisolone,

cyclophosphamide, and azathioprine did not have a significant beneficial effect on preservation of kidney function.[23] However, a course of oral steroids may be considered in patients with high-grade proteinuria, and cytotoxic drugs are indicated in a small subset of patients with crescentic rapidly progressive IgAN. A targeted-release formulation of budesonide added to optimized blockade of the renin-angiotensin system reduced proteinuria in a recent trial.[24] There are conflicting data about the value of fish oil dietary supplements (eicosanoids) in preventing renal progression in patients with IgAN. The practice of tonsillectomy in IgAN has not been clearly established. The lack of effective therapy has provided the impetus for several clinical trials testing novel therapies. Thus far, a small trial of rituximab has not shown compelling evidence for a benefit in IgAN.

Antineutrophil Cytoplasmic Antibodies–Associated Renal Vasculitis

KEY CONCEPTS

Antineutrophil Cytoplasmic Antibodies–Associated Renal Vasculitis

- Renal vasculitis with glomerular involvement includes microscopic polyangiitis (MPA), granulomatosis with polyangiitis (GPA), and necrotizing crescentic glomerulonephritis (renal-limited vasculitis)
- Associated with ANCAs
- Rapidly progressive glomerulonephritis is common; early treatment includes pulse methylprednisolone, cyclophosphamide, rituximab, or possibly plasma exchange
- Maintenance therapy: azathioprine, rituximab

ANCAs are associated with a distinct form of vasculitis that can affect many different types of vessels and any organ in the body. ANCA-associated vasculitis (AAV; see Chapter 59) is categorized into four main types: microscopic polyangiitis (MPA), granulomatosis with polyangiitis (GPA, formerly called Wegener granulomatosis), eosinophilic granulomatosis with polyangiitis (EGPA, formerly Churg-Strauss syndrome), and renal limited vasculitis. As shown in Table 68.2, the patterns of ANCAs differ, depending on the type of vasculitis.

Renal involvement is common in AAV but varies in type and severity. It occurs more frequently in MPA (90%) and in GPA (80%) and less frequently in EGPA (45%). Clinically, RPGN is a common manifestation of these renal vasculitides characterized by the presence of hematuria, sub-nephrotic proteinuria, active urinary sediments, and renal failure. Systemic disease may also produce clinical features pertaining to other organ involvement, for example hemoptysis from pulmonary hemorrhage or upper respiratory tract symptoms. Some patients present with overlap syndromes, showing features of both AAV and anti-GBM disease (see below).

TABLE 68.2 Prevalence of Antineutrophil Cytoplasmic Antibodies in Renal Vasculitis

	ANCAs TEST POSITIVITY (%)	
Type of Renal Vasculitis	P-ANCAs or Anti-Mpo	C-ANCAs or Anti-Pr3
Polyarteritis nodosa	10–20	10–20
Microscopic polyangiitis	50–80	10–20
Granulomatosis with polyangiitis (Wegener granulomatosis)	10–20	80–90
Necrotizing and crescentic GN	50–80	10–20

MPO, Myeloperoxidase; *PR3*, proteinase 3.

Pathology

The ANCA vasculitides produce similar types of kidney abnormalities, affecting predominantly, but not exclusively the glomeruli (Fig. 68.11). Histological examination usually reveals glomerular lesions that are characteristically focal and segmental in distribution, with fibrinoid necrosis and crescent formation. Breaks in the GBM may be seen. There may be accompanying necrotizing arteritis. In contrast to immune complex–mediated vasculitis, ANCA-associated vasculitis usually has little or no Ig deposition in injured glomerular vessels, with minimal or negative staining by immunofluorescence, a so-called pauci-immune pattern. Patients with large percentage of cellular crescents (>50%) typically present with severely reduced renal function but have a good chance for recovery of renal function with treatment, whereas those with a greater percentage of globally sclerotic (scarred) glomeruli are less likely to recover renal function.[25]

Treatment and Prognosis

ANCA-associated renal vasculitis tends to be severe and fulminant. The systemic features of these diseases impart a high mortality. Indeed, in the pre-steroid era these conditions had close to universal mortality. Thus, early detection is critically important in management. Even with early diagnosis, approximately one-third of patients will progress to renal failure within 5 years. Relapsing courses in patients with microscopic polyangiitis, underscore the importance of prompt induction treatment followed by maintenance therapy. Glucocorticoids are important in the early treatment of renal vasculitis. Cyclophosphamide-based regimens are very effective in inducing remission in patients with AAVs. Most advocate daily oral cyclophosphamide regimens, but intermittent pulse cyclophosphamide may be substituted to reduce the toxicity of extended therapy. In patients with severe pulmonary hemorrhage or rapidly progressive glomerulonephritis caused by renal vasculitis, pulse methylprednisolone, followed by prednisone and daily cyclophosphamide, is clearly indicated. Adjunctive plasma exchange is commonly used in cases of aggressive pulmonary–renal syndrome or life-threatening disease. Rituximab has become a valuable alternative to cyclophosphamide for induction treatment based on the results of two pivotal prospective randomized controlled trials.[26] Maintenance therapy is usually given for 12 to 18 months after achieving remission. Azathioprine has been the mainstay of maintenance therapy in AAVs after remission induction, but rituximab is likely to have a prominent role in maintenance therapy, as indicated by favorable results from recent trials.[27] The latest clinical evidence also suggests a role for inhibition of terminal complement components in the management of AAV, with phase 2 studies suggesting that avacopan (an inhibitor of the C5a receptor) can be effective in replacing high-dose steroids and a phase 3 study ("ADVOCATE") due to report shortly.

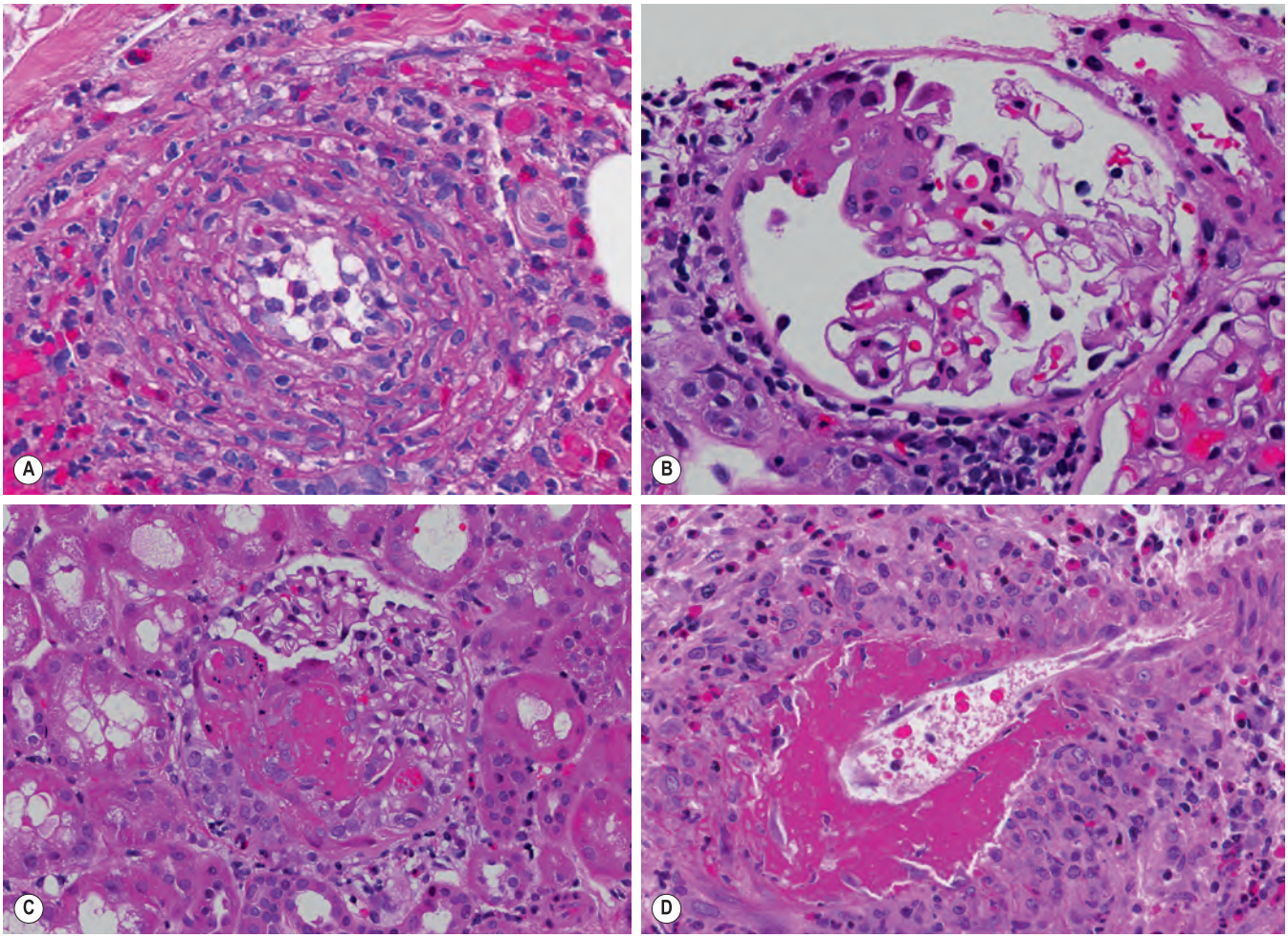


FIG. 68.11 Antineutrophil Cytoplasmic Antibody (ANCA)–Associated Systemic Vasculitis. (A) Skin biopsy showing leukocytoclastic vasculitis of a small dermal artery (hematoxylin and eosin [H&E] stain). (B) Glomerulus illustrating a characteristic segmental proliferative lesion forming an adhesion to the Bowman capsule in an early and mild case (H&E stain). (C) Glomerulus with classical fibrinoid necrosis and an associated cellular crescent of a more severe and rapidly progressive case (H&E stain). (D) Necrotizing vasculitis in a small renal artery in an extreme case of ANCA-associated renal vasculitis; large numbers of eosinophils are present in the perivascular inflammatory infiltrate (H&E stain).

Anti-Glomerular Basement Membrane Antibody-Mediated Nephritis (Goodpasture Disease)

KEY CONCEPTS

Anti–Glomerular Basement Membrane Disease

- Circulating anti-GBM antibody
- Associated with pulmonary hemorrhage
- Rapidly progressive glomerulonephritis with cellular crescents and linear deposits of IgG
- Treated with high-dose steroids, cyclophosphamide, plasma exchange

Anti-GBM disease is a rare, but classic immune-mediated cause of severe pulmonary-renal syndrome. The cardinal pathogenic factor is an autoantibody to a component of type IV collagen, specifically the noncollagenous domain 1 of the α_3 chain subunit. within the GBM or the alveolar basal membrane.[28] These autoantibodies lead to GBM rupture with development of crescentic glomerulonephritis (Fig. 68.12) and account for the pathognomonic finding of linear deposition of IgG along the glomerular capillaries as assessed by immunofluorescence microscopy. In contrast to AAV, the glomerular lesions of anti-GBM disease typically appear to be of the same age, suggesting a single moment of onset of the disease.

The genesis of anti-GBM antibodies in sporadic cases is unknown. It is postulated that an inciting event such as viral infection, or exposure to environmental factors, such as hydrocarbons or tobacco, perturbs the normal conformation of collagen in the basement membrane, exposing previously cryptic epitopes on the α_3 subunit and eliciting the pathogenic autoantibody response. This disease can occur paradoxically in normal kidneys transplanted into patients with Alport syndrome, the latter caused by mutations in the α-chains of type IV collagen, because the normal collagen is recognized as being foreign and elicits a humoral immune.

Clinically, malaise, weight loss, fever, or arthralgia may be the initial features of anti-GBM disease. Some patients present with isolated renal involvement, but more typically,

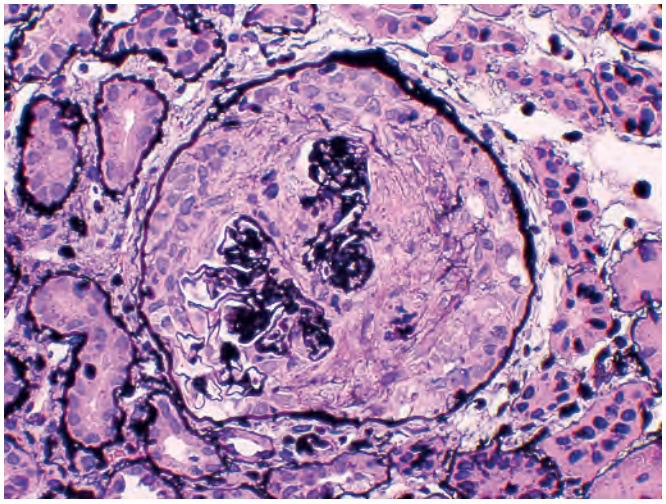


FIG. 68.12 Anti-Glomerular Basement Membrane (GBM) (Goodpasture) Disease. A circumferential cellular crescent fills the Bowman capsule and compresses the glomerular tuft (Silver stain).

pulmonary hemorrhage with hemoptysis accompanies acute renal failure. The devastating nature of the disease warrants aggressive treatment. Pulse methylprednisolone, cyclophosphamide, and plasma exchange are early indicated in the course of Goodpasture disease.[29] The role of rituximab in this disease requires further study. Reversibility of renal disease is unlikely if renal function is substantially impaired or oliguria ensues before treatment is started. Immunosuppressive treatment is normally continued until the patient has been in sustained clinical remission and anti-GBM titers are minimal or absent for at least 3 months. In contrast to AAV, anti-GBM is a "one-hit" disease and does not relapse.

Lupus Nephritis

KEY CONCEPTS

Lupus Nephritis

- *Class I:* normal glomeruli by light microscopy in patients with SLE
- *Class II, mesangial:* Immunosuppressive treatment is usually not indicated unless patient has proteinuria >1 g/day despite renin–angiotensin–aldosterone system blockade, especially if urine sediment is nephritic
- *Class III, focal nephritis and class IV, diffuse nephritis:* MMF or intravenous cyclophosphamide (IVCYC) plus glucocorticoids for induction followed by MMF or azathioprine plus low-dose corticosteroids as maintenance therapy
- *Class V, Membranous nephropathy:* Alternate-day prednisone with MMF, or alternatively bimonthly pulse cyclophosphamide or low-dose daily cyclosporine

SLE (see Chapter 52) is a systemic inflammatory disease of unknown etiology that can affect most organs in the body. The majority of patients with SLE have some degree of glomerular disease, but the presentation, course and outcome of lupus nephritis has a wide spectrum. Lupus nephritis occurs more often and is associated with less favorable outcomes among Hispanic, Asian, Native American, and especially African American populations compared with Caucasians. Nephritis is a major cause of morbidity and mortality and accounts for a large portion of all hospital admissions in patients with SLE.

Pathogenesis

Several different mechanisms may be involved in the pathogenesis of lupus nephritis, resulting in a wide spectrum of renal lesions. Deposition of immune complexes in the kidney appears to be the initiating event in proliferative lupus nephritis; however, only a subset of immune complexes appears to be nephritogenic. DNA and anti-DNA antibodies are known to be concentrated in glomerular deposits in the subendothelial space and are likely to play a central role in the pathogenesis of proliferative lupus nephritis. Unfortunately, there are fewer insights into the pathogenesis of lupus MN with its characteristic epimembranous immune deposits. Only a small proportion of patients with lupus MN have antibodies against the PLA_2R antigen. Although T cells are almost certainly involved in autoantibody production, it is unknown whether they have a direct role in the pathogenesis of lupus nephritis.

Clinical Features

Patients with lupus nephritis can have the full spectrum of renal presentations, reflecting the broad extent of renal involvement. Asymptomatic hematuria or proteinuria may be the presenting features, but they often progress to nephritic and/or nephrotic syndrome. Hypertension, azotemia, nephritic urine sediment (with hematuria and cellular casts), hypocomplementemia and high anti–double-stranded DNA (dsDNA) titers are more commonly found in patients with proliferative lupus nephritis. RPGN is usually associated with the appearance of cellular crescents and may be superimposed on severe proliferative or membranous forms of lupus nephritis.

Pathology

The former classification of renal biopsy in lupus nephritis by the World Health Organization (WHO) was revised by an international committee in 2004 (Figs. 68.13–68.15). A summary of the histological features in each class of lupus nephritis can be found in Table 68.3.

Treatment

In 2012, the American College of Rheumatology (ACR), the European League Against Rheumatism (EULAR), and the European Renal Association–European Dialysis and Transplantation Association (ERA-EDTA) published their recommendations for the management and treatment of patients with lupus nephritis.[30,31] Immunosuppressive treatment of mesangial classes of lupus nephritis (classes I and II) is usually not indicated (ACR). However, the distinction between early mesangial lesions that are in transition to more ominous classes that reflect mild and stable nephropathy is difficult. Consequently, treatment with prednisone alone or in combination with azathioprine has been recommended for patients with proteinuria that exceeds 1 g/day despite blockade of the renin–angiotensin–aldosterone (RAA) system, especially if the patient also has a nephritic urinary sediment (EULAR/ERA-EDTA). For patients with focal or diffuse proliferative glomerulonephritis (classes III and IV), ACR and EULAR/ERA-EDTA recommended mycophenolate mofetil (MMF) or intravenous cyclophosphamide (IVCYC) plus glucocorticoids. Low-dose IVCYC (500 mg IV every 2 weeks in six doses) offers a favorable balance of efficacy and relatively low toxicity for Caucasian patients with Western or Southern

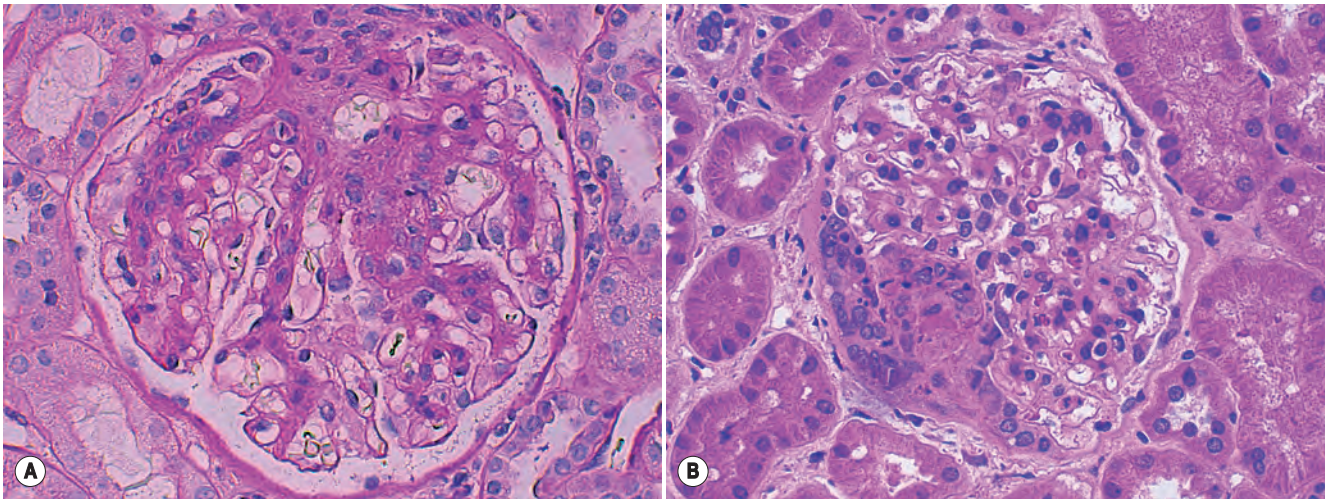


FIG. 68.13 Classes of the Pathology of Lupus Nephritis (1). (A) Class II, mesangial proliferative lupus nephritis; mesangial areas are expanded by cells and matrix but the peripheral capillary loops remain widely patent (periodic acid–Schiff [PAS] stain). (B) Class III, focal lupus nephritis; solid lesion at the lower right portion of this glomerulus demonstrates segmental fibrinoid necrosis. Note the nuclear fragments (karyorrhexis) in the fibrinous exudate (hematoxylin and eosin [H&E] stain.)

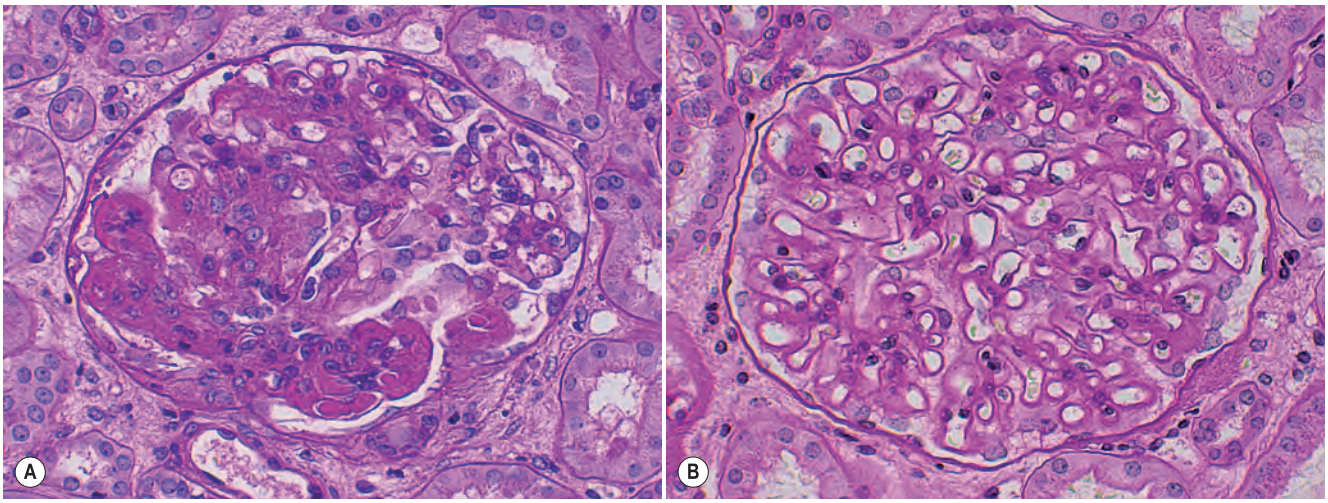


FIG. 68.14 Classes of the Pathology of Lupus Nephritis (2). (A) Class IV, diffuse lupus nephritis; glomerulus with irregular but nearly global changes, including obliteration of many capillary loops resulting from endocapillary hypercellularity, "wire loop" thickening and hyaline thrombi (periodic acid–Schiff [PAS] stain). (B) Class V, membranous lupus nephritis; glomerulus shows minimally increased mesangial cellularity with thickened but widely patent capillary loops (PAS stain).

European ancestry. Higher dose, monthly IVCYC for 6 months, plus 3 daily IV infusions of methylprednisolone initially, followed by prednisone has also been recommended for patients with ominous clinical and histological prognostic indicators, including cellular crescents and fibrinoid necrosis. ACR and EULAR/ERA-EDTA recommend either azathioprine (AZA) or MMF as maintenance therapy for patients showing a favorable response after initial immunosuppressive therapy. Although AZA and MMF appeared to be equally effective maintenance therapies in a European study, patients randomized to maintenance therapy with MMF had better outcomes than those randomized to AZA in a larger study conducted worldwide. The duration of maintenance immunosuppressive therapy involves careful consideration of the risks of another renal flare-up versus the risks of drug toxicity. EULAR/ERA-EDTA recommend at least 3 years of therapy in patients showing improvement after initial treatment. In general, we have offered a comparable recommendation that treatment should continue for at least 1 year after remission of renal disease to prevent exacerbations.

Neither IVCY nor MMF is universally effective in the management of lupus nephritis, hence the search for more efficacious treatment regimens, including rituximab, belimumab (binds to BAFF), immunomodulators (*e.g.*, laquinimod), cytokine inhibitors, immunoablation without or with stem-cell reconstitution, and immunological co-stimulation inhibitors (*e.g.*, CTLA-4-Ig), continues.[32,33] The importance of these ongoing efforts to investigate novel approaches to treatment is underscored by observations that although the risk of renal failure caused by lupus nephritis decreased from the 1970s to the mid-1990s (coincident with the increased use of cyclophosphamide), that risk has shown a reverse trend and increased slightly in the last two decades.[34]

For patients with "pure" class V lupus MN and nephrotic-range proteinuria, both the ACR and the EULAR/ERA-EDTA

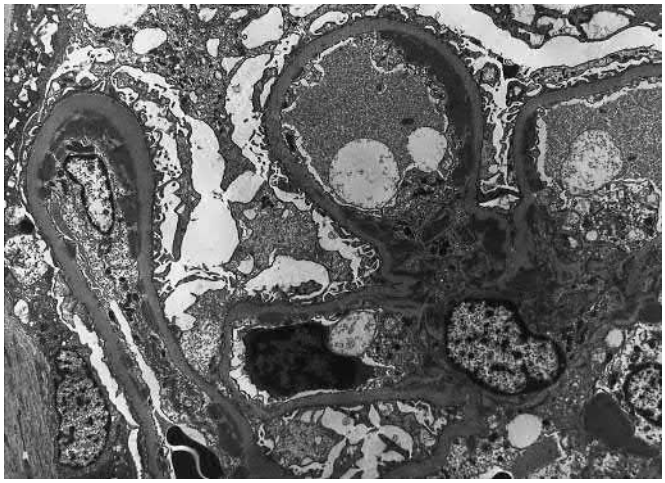


FIG. 68.15 Ultrastructure of Proliferative Lupus Nephritis. Electron microscopy demonstrates the characteristic mesangial deposits (dark materials interspersed within the centrally located amorphous, gray mesangial matrix) and subendothelial deposits (dark materials extending along the peripheral capillary loops).

TABLE 68.3 International Society of Nephrology/Renal Pathology Society 2004 Classification of Lupus Nephritis

Class	Histologic Features/Comments
I. Minimal mesangial	Normal light microscopy (LM); mesangial deposits by immunofluorescence (IF) and electron microscopy (EM)
II. Mesangial proliferative	Pure mesangial hypercellularity and matrix expansion *IF* and *EM:* mesangial immune deposits
III. Focal	Glomerular capillary obliteration in <50% of nephrons as a result of proliferation or sclerosis *LM:* Increased numbers of mesangial, endothelial, and/or hematogenous cells. Active inflammatory lesions (karyorrhexis, fibrinoid necrosis, adhesion to the Bowman capsule, cellular crescents, interstitial inflammatory infiltrates). Wire loop lesions. Hyaline thrombi *IF* and *EM:* Mesangial and peripheral capillary loop (subendothelial) immune complex deposits
IV. Diffuse	Qualitatively similar histologic lesions as in class III. Glomerular capillary obliteration involving >50% of nephrons. Subsets defined as primarily global (class IV-G) or primarily segmental (class IV-S) involvement
V. Membranous	*LM:* Regular thickening of the peripheral capillary loops of the glomerulus. Mesangial expansion *EM:* Subepithelial, intramembranous, mesangial (but no or very rare subendothelial) immune complex deposits
VI. Advanced	>90% global sclerosis without residual active lesions

have recommended oral prednisone and MMF. A prospective controlled trial showed that both IVCYC and cyclosporine A (CSA) were more effective than steroids alone in inducing remission of proteinuria in lupus MN, but that relapse of nephrotic syndrome occurred significantly more often in the CSA group than in the IVCYC treatment group.

Systemic Sclerosis (SSc, Scleroderma) Renal Disease

KEY CONCEPTS

Nephropathies of Selected Connective Tissue Diseases

- *Scleroderma renal crisis:* Predominantly renal vasculopathy; moderate to severe (high renin) hypertension with progressive renal failure—treated with angiotensin-converting enzyme inhibitors (ACEIs); additional antihypertensive agents may be needed
- *Primary Sjögren syndrome:* Distal renal tubular acidosis, nephrogenic diabetes insipidus, interstitial nephritis, hypokalemia, and/or renal calculi; rarely glomerulonephritis

The most common and potentially devastating renal manifestation of systemic sclerosis (see Chapter 56) is scleroderma renal crisis (SRC).[35] Most cases of SRC occur within 4 years of the onset of systemic sclerosis in patients with diffuse cutaneous scleroderma affecting the proximal extremities and the trunk, although some may have a scleroderma renal crisis without dermatological features. Several characteristics of the disease, including rapid progression of skin thickening, palpable tendon friction rubs, anti-RNA polymerase III antibody, recent-onset cardiac events (*e.g.*, pericardial effusion or heart failure), new-onset anemia (especially if associated with microangiopathic hemolysis and thrombocytopenia), and recent treatment with high-dose corticosteroids, help identify patients at increased risk for developing SRC. A classic clinical presentation may obviate the need for renal biopsy. However, renal biopsy may be necessary in atypical cases. For example, about 20% of SRC cases occur before the diagnosis of scleroderma has been established. Patients with scleroderma can rarely also develop other renal diseases, such as ANCA-associated vasculitis, which are important to recognize because they require treatments different from those usually recommended for SRC.

Scleroderma renal crisis is characterized by abrupt onset of renin-mediated moderate to severe hypertension, rapid deterioration of renal function, and proteinuria, which is usually nonnephrotic. Associated findings may include microangiopathic hemolysis, hypertensive encephalopathy including seizures, hypertensive retinopathy, acute left ventricular failure, and pulmonary edema. Normotensive renal crisis occurs infrequently and may be recognized by the presence of microangiopathic hemolysis with or without unexplained azotemia. The primary pathogenic process appears to be a renal vasculopathy involving predominantly the interlobular arteries and arterioles. Marked intimal thickening with an attendant "mucoid" appearance, and fibrinoid necrosis in the absence of vasculitis, are common and characteristic of the disease (Fig. 68.16). Immune deposits are rarely observed by fluorescence or electron microscopy studies.

Although a variety of treatments have been proposed for patients with scleroderma, none has been proven to be consistently efficacious. The most significant therapeutic advance in the treatment of SRC is the use of angiotensin-converting enzyme (ACE) inhibitors (ACEIs), which have dramatically increased the 1-year survival of patients with SCR. Prompt treatment with ACEIs is recommended for hypertensive and normotensive SRC because these agents may reverse the process by interfering with angiotensin II–mediated vasoconstriction and by inhibiting the degradation of bradykinin, a potent vasodilator, by ACE. Despite this impressive impact of ACEIs on clinical outcomes, approximately one-third to one-half of patients still progress to early death or renal failure. This has prompted continued investigation of additional treatment strategies, including blockade

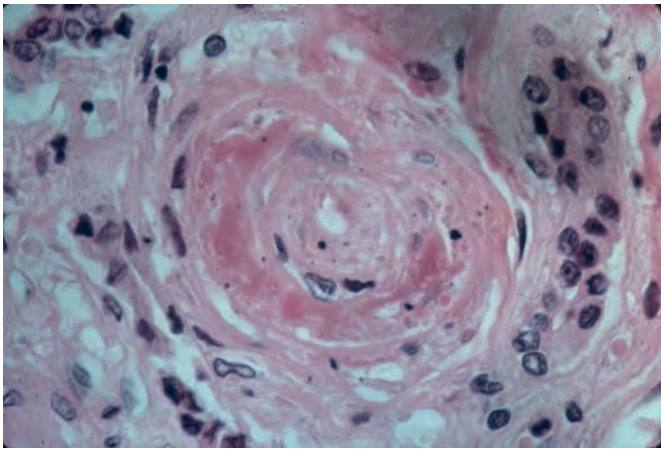


FIG. 68.16 Scleroderma Renal Crisis. Renal arteriole demonstrates extensive fibrin deposition (dark material) within multiple layers of its wall. The lumen is further compromised by severe swelling and intimal hyperplasia (Masson trichrome stain).

of the RAA system at multiple sites, for example, with a direct renin inhibitor as well as a potent vasodilator such as prostacyclin or an endothelin receptor antagonist.

Renal Disease of Primary Sjögren Syndrome

A variety of renal manifestations occur in approximately one-third of patients with primary Sjögren syndrome (see Chapter 55), including tubular dysfunction (proximal or, more often, distal renal tubular acidosis), Fanconi syndrome, hypokalemia that may be profound and associated with paralysis, Bartter syndrome, Gitelman syndrome, nephrogenic diabetes insipidus), as well as renal calculi, nephrocalcinosis, interstitial nephritis, pseudolymphoma, necrotizing vasculitis, and glomerulopathy.[36] Studies of renal pathology among patients with primary Sjögren syndrome have shown that acute and/or chronic tubulointerstitial nephritis is the predominant lesion. A diverse range of glomerular diseases, including MN and proliferative glomerulonephritis (focal or diffuse, and membranoproliferative), have been reported in primary Sjögren syndrome. In such cases, the possibility of overlap with SLE should be considered. Immunosuppressive therapy for renal manifestations of primary Sjögren syndrome should be individualized on the basis of activity and severity of the glomerular and/or tubulointerstitial nephritis. Patients with serious complications of tubular dysfunction may warrant immunosuppressive therapies based on preliminary evidence that autoantibodies (against, for example, carbonic-anhydrase-II) as well as cell-mediated injury likely contribute to at least some of those disorders.

Thrombotic Microangiopathies

KEY CONCEPTS

Thrombotic Microangiopathies

- Thrombotic microangiopathies (TMA) can be hereditary or acquired and the main primary forms are complement-mediated TMA (atypical hemolytic uremic syndrome, aHUS), Shiga-toxin-mediated TMA (ST-HUS) and ADAMTS13 deficiency–mediated TMA (TTP, thrombotic thrombocytopenic purpura)
- Key clinical findings are microangiopathic hemolytic anemia (MAHA), thrombocytopenia, and acute kidney injury
- Treatment according to the etiology, plasma exchange being the mainstay of acute therapy

Thrombotic microangiopathies (TMA) are a heterogeneous group of diseases of either hereditary or acquired etiology that share the clinical findings of microangiopathic hemolytic anemia (MAHA) and thrombocytopenia.[37] Peripheral blood smears reveal fragmented erythrocytes, which result from shearing injury in the microvasculature. The common pathologic features are arterial and capillary microthrombi formation and characteristic endothelial and vessel wall damage. Based on a number of advances in pathophysiological insight, diagnostic tests and treatment in the recent years the nomenclature has been modified to reflect causality. Below are the main primary causes of TMA. In addition, a range of secondary causes of TMA, among them pregnancy, severe hypertension, glomerular diseases or certain drugs, have been described.

ADAMTS13 deficiency–mediated TMA (thrombotic thrombocytopenic purpura, TTP) features predominantly neurologic symptoms, with acute kidney injury occurring rarely. ADAMTS13 deficiency can be hereditary or acquired based on autoantibodies against the metalloprotease. Treatment consists of acute plasma exchange and prophylactic plasma transfusions and glucocorticoids in the course of the disease.

In Shiga-toxin-mediated TMA (ST-HUS) a prodromal diarrheal disease is followed by TMA with acute kidney injury and occasionally neurological abnormalities. *Escherichia coli*, serotype O157:H7 and *Shigella dysenteriae* are the most common toxin-producing bacteria and often occur in outbreaks. Treatment is supportive, with patients commonly requiring dialysis.

Complement-mediated TMA (atypical hemolytic uremic syndrome, aHUS) is a prototypic disease of complement dysregulation. The complement system is a highly conserved part of the innate immune system, with its key components C3 and C5. It mediates immune cell and platelet activation, endothelial cell damage and pathogen opsonization. Activation can happen through the classical pathway with an antibody signal, the lectin-pathway through a pattern recognition signal, or through the alternative pathway (tick-over). Complement-mediated TMA is thought to be based on uncontrolled alternative pathway activation, either caused by loss-of-function mutations in regulatory complement genes (*e.g.*, *CFH*, *CFI*, *CD46*) or gain-of-function mutations in effector genes (*e.g.*, *CFB*, *C3*). By far the most common defect is in complement factor H (*CFH*), which constitutes 10% of the cases. Importantly, most genetic mutations show incomplete penetrance. Environmental triggers, such as pregnancy or infection, are often necessary to trigger the disease. Rarely, antibodies against CFH or CFI proteins are found.

Clinical features of complement-mediated TMA are commonly acute kidney injury and hypertension. Treatment is based on plasma exchange and the complement factor C5 inhibitor, Eculizumab. Treatment duration with eculizumab remains unproven. Treatment failure should prompt screening for genetic causes or autoantibodies. The other TMAs also feature complement activation, but it is less clear if this is a primary event or bystander effect. The use of complement inhibitors is controversial here.

The success of eculizumab in treating aHUS has prompted a new interest in complement inhibitors and the complement system itself. Drugs under investigation include inhibitors targeting C5AR1 and MASP2, and long-acting formulations of C5 inhibitors.[38] Recent evidence points to the existence of intracellular C3 and C5 complement systems in many cells.[39] An unresolved field remains the understanding of the integration and inter-connection of the intracellular and extracellular complement pathways.

ON THE HORIZON

- The advent of new genetic and molecular techniques and new disease models has led to exciting progress in our understanding of the biology of the glomerulus, the pathogenesis of many glomerular diseases, and the influence of genetic variants on disease predisposition and progression.
- These insights should lead to better noninvasive diagnostic techniques, biomarkers, and predictors of prognosis and relapse and facilitate a more personalized approach to therapy rather than a one-size-fits-all approach.
- Novel targeted therapies are on the horizon that will interrupt or modulate the underlying pathophysiology of the individual diseases and also halt the downstream pathways of injury and fibrosis common to all of the glomerular diseases.

REFERENCES

1. Couser WG, Johnson RJ. The etiology of glomerulonephritis: roles of infection and autoimmunity. *Kidney Int*. 2014;86:905–914.
2. Floege J, Amann K. Primary glomerulonephritides. *Lancet*. 2016;387:2036–2048.
3. Waldman M, Crew RJ, Valeri A, et al. Adult minimal-change disease: clinical characteristics, treatment, and outcomes. *Clin J Am Soc Nephrol*. 2007;2:445–453.
4. Maas RJ, Deegens JK, Wetzels JF. Permeability factors in idiopathic nephrotic syndrome: historical perspectives and lessons for the future. *Nephrol Dial Transplant*. 2014;29:2207–2216.
5. Genovese G, Friedman DJ, Ross MD, et al. Association of trypanolytic ApoL1 variants with kidney disease in African Americans. *Science*. 2010;329:841–845.
6. D'Agati VD, Alster JM, Jennette JC, et al. Association of histologic variants in FSGS clinical trial with presenting features and outcomes. *Clin J Am Soc Nephrol*. 2013;8:399–406.
7. Iijima K, Sako M, Nozu K, et al. Rituximab for childhood-onset, complicated, frequently relapsing nephrotic syndrome or steroid-dependent nephrotic syndrome: a multicentre, double-blind, randomised, placebo-controlled trial. *Lancet*. 2014;384:1273–1281.
8. Beck LH, Bonegio RGB, Lambeau G, et al. M-type phospholipase A2 receptor as target antigen in idiopathic membranous nephropathy. *N Engl J Med*. 2009;361:11–21.
9. Tomas NM, Beck LH, Meyer-Schwesinger C, et al. Thrombospondin type-1 domain-containing 7A in idiopathic membranous nephropathy. *N Engl J Med*. 2014;371:2277–2287.
10. Sethi S, Debiec H, Madden B, et al. Neural epidermal growth factor-like 1 protein (NELL-1) associated membranous nephropathy. *Kidney Int*. 2020;97:163–174.
11. Hoxha E, Thiele I, Zahner G, et al. Phospholipase A2 receptor autoantibodies and clinical outcome in patients with primary membranous nephropathy. *J Am Soc Nephrol*. 2014;25:1357–1366.
12. Fervenza FC, Appel GB, Barbour SJ, et al. Rituximab or Cyclosporine in the Treatment of Membranous Nephropathy. *N Engl J Med*. 2019;381:36–46.
13. Fakhouri F, Frémeaux-Bacchi V, Noël L-H, et al. C3 glomerulopathy: a new classification. *Nat Rev Nephrol*. 2010;6:494–499.
14. De Vriese AS, Sethi S, Van Praet J, et al. Kidney Disease Caused by Dysregulation of the Complement Alternative Pathway: An Etiologic Approach. *J Am Soc Nephrol*. 2015;26:2917–2929.
15. Kupin WL. Viral-Associated GN: Hepatitis B and Other Viral Infections. *Clin J Am Soc Nephrol*. 2017;12:1529–1533.
16. Bertino G, Ardiri A, Proiti M, et al. Chronic hepatitis C: This and the new era of treatment. *World J Hepatol*. 2016;8:92–106.
17. Sise ME, Bloom AK, Wisocky J, et al. Treatment of hepatitis C virus-associated mixed cryoglobulinemia with direct-acting antiviral agents. *Hepatology*. 2016;63:408–417.
18. Sneller MC, Hu Z, Langford CA. A randomized controlled trial of rituximab following failure of antiviral therapy for hepatitis C virus-associated cryoglobulinemic vasculitis. *Arthritis Rheum*. 2012;64:835–842.
19. Lan X, Rao TKS, Chander PN, et al. Apolipoprotein L1 (APOL1) Variants (Vs) a possible link between Heroin-associated Nephropathy (HAN) and HIV-associated Nephropathy (HIVAN). *Front Microbiol*. 2015;6:571.
20. Nobakht E, Cohen SD, Rosenberg AZ, et al. HIV-associated immune complex kidney disease. *Nat Rev Nephrol*. 2016;12:291–300.
21. Working Group of the International IgA Nephropathy Network and the Renal Pathology Society Cattran DC, Coppo R, et al. The Oxford classification of IgA nephropathy: rationale, clinicopathological correlations, and classification. *Kidney Int*. 2009;76:534–545.
22. Wyatt RJ, Julian BA. IgA nephropathy. *N Engl J Med*. 2013;368:2402–2414.
23. Rauen T, Eitner F, Fitzner C, et al. Intensive Supportive Care plus Immunosuppression in IgA Nephropathy. *N Engl J Med*. 2015;373:2225–2236.
24. Fellström BC, Barratt J, Cook H, et al. Targeted-release budesonide versus placebo in patients with IgA nephropathy (NEFIGAN): a double-blind, randomised, placebo-controlled phase 2b trial. *Lancet*. 2017;389:2117–2127.
25. Berden AE, Ferrario F, Hagen EC, et al. Histopathologic classification of ANCA-associated glomerulonephritis. *J Am Soc Nephrol*. 2010;21:1628–1636.
26. Stone JH, Merkel PA, Spiera R, et al. Rituximab versus cyclophosphamide for ANCA-associated vasculitis. *N Engl J Med*. 2010;363:221–232.
27. Guillevin L, Pagnoux C, Karras A, et al. Rituximab versus azathioprine for maintenance in ANCA-associated vasculitis. *N Engl J Med*. 2014;371:1771–1780.
28. Pedchenko V, Bondar O, Fogo AB, et al. Molecular architecture of the Goodpasture autoantigen in anti-GBM nephritis. *N Engl J Med*. 2010;363:343–354.
29. Pusey CD. Anti-glomerular basement membrane disease. *Kidney Int*. 2003;64:1535–1550.
30. Hahn BH, McMahon MA, Wilkinson A, et al. American College of Rheumatology guidelines for screening, treatment, and management of lupus nephritis. *Arthritis Care Res (Hoboken)*. 2012;64:797–808.
31. Bertsias GK, Tektonidou M, Amoura Z, et al. Joint European League Against Rheumatism and European Renal Association-European Dialysis and Transplant Association (EULAR/ERA-EDTA) recommendations for the management of adult and paediatric lupus nephritis. *Ann Rheum Dis*. 2012;71:1771–1782.
32. Karras A, Jayne D. New biologics for glomerular disease on the horizon. *Nephron Clin Pract*. 2014;128:283–291.
33. Mok CC. Towards new avenues in the management of lupus glomerulonephritis. *Nat Rev Rheumatol*. 2016;12:221–234.
34. Tektonidou MG, Dasgupta A, Ward MM. Risk of End-Stage Renal Disease in Patients With Lupus Nephritis, 1971-2015: A Systematic Review and Bayesian Meta-Analysis. *Arthritis Rheumatol*. 2016;68:1432–1441.
35. Mouthon L, Bussone G, Berezné A, et al. Scleroderma renal crisis. *J Rheumatol*. 2014;41:1040–1048.
36. François H, Mariette X. Renal involvement in primary Sjögren syndrome. *Nat Rev Nephrol*. 2016;12:82–93.
37. George JN, Nester CM. Syndromes of thrombotic microangiopathy. *N Engl J Med*. 2014;371:654–666.
38. Mastellos DC, Ricklin D, Lambris JD. Clinical promise of next-generation complement therapeutics. *Nat Rev Drug Discov*. 2019;18:707–729.
39. Hess C, Kemper C. Complement-Mediated Regulation of Metabolism and Basic Cellular Processes. *Immunity*. 2016;45:240–254.

69

Immunologic Mechanisms of Atherosclerosis and Myocarditis

Peter Libby and Andrew H. Lichtman

Immunologic mechanisms contribute critically to many cardiovascular diseases. This chapter will focus on atherosclerosis—an example of a common vascular disease—and myocarditis—an immune-mediated disease of the heart itself.

IMMUNOLOGICAL MECHANISMS THAT CONTRIBUTE TO ATHEROSCLEROSIS

Immunological mechanisms participate in all phases of atherosclerosis from lesion initiation through the long clinically silent or stable phase of lesion progression, culminating in the thrombotic complications that bring patients to the attention of physicians most dramatically.[1–3]

Atherosclerosis Initiation

The normal artery has a tri-laminar structure (Fig. 69.1).[4] The outermost layer, the adventitia, contains fibroblasts, nerve endings, and mast cells. Occasionally, lymphoid collections localize in the adventitia. The middle layer, the tunica media, consists of multiple layers of arterial smooth muscle cells (SMCs) embedded in layers of extracellular matrix. In medium-sized muscular arteries where atherosclerosis occurs most commonly, a collagen and elastin-rich matrix predominates. In larger elastic arteries, such as the aorta, lamellae of elastin surround layers of SMC.

The innermost layer of all arteries consists of a monolayer of endothelial cells (ECs) that abut a basement membrane. The internal elastic lamina separates the tunica media from the intima. In most human arteries affected by atherosclerosis, the intima contains extracellular matrix with occasional resident SMCs. Portions of arteries particularly predilected to develop atherosclerotic plaques often contain cushions of intimal SMCs even early in life. This complex structure of the tunica intima in humans differs from many experimental animals, such as the mouse or the rabbit, in which the intima is thin and lacks resident SMCs.

The EC monolayer in contact with the blood displays remarkable homeostatic properties under normal circumstances (Fig. 69.2, *left hand column*). The endothelial monolayer can maintain blood in a liquid state during prolonged contact, a property seldom found at other natural or synthetic surfaces. The ability to resist thrombus accumulation depends on a tightly orchestrated series of functions depicted in Fig. 69.2. The normal anti-coagulant mechanisms associated with the undisturbed endothelial surface include expression of thrombomodulin, and heparan sulfate-like molecules that can activate antithrombin III. The enzymes urokinase-type plasminogen activator (uPA) and tissue-type plasminogen activator (tPA) regulate fibrinolysis at the surface of the endothelial monolayer should a thrombus form. The normal endothelium resists prolonged contact with leukocytes that circulate in blood. Normal ECs produce nitric oxide, a gas that promotes smooth muscle relaxation and hence vasodilatation.

These salutary functions can change when EC encounter danger signals, such as pro-inflammatory cytokines, pathogen-associated molecular patterns (PAMPs), or danger-associated molecular patterns (DAMPs) (see Fig. 69.2, *middle column*).[5] Engagement of selective receptors in the case of cytokines or pattern-recognition receptors, such as the toll-like receptors (TLRs) in the case of PAMPs and DAMPs, instigates a program of functions that disrupts the normal homeostatic properties of the endothelium. Pro-coagulant and anti-fibrinolytic mechanisms prevail over anti-coagulant, anti-thrombotic, and endogenous fibrinolytic mechanisms as shown in Fig. 69.2. The endothelial cell activated by inflammatory mediators can produce the potent pro-coagulant tissue factor and plasminogen-activator inhibitor-1 (PAI-1), the inhibitor of the endogenous pro-fibrinolytic enzymes uPA and tPA.

Stimulation by pro-inflammatory mediators elicits the expression by EC of molecules that mediate the adhesion of various classes of leukocytes (Fig. 69.3, *left hand*).[4] E-selectin, a prototypical endothelial-leukocyte adhesion molecular, participates in acute inflammatory responses by recruiting polymorphonuclear leukocytes. Vascular cell adhesion molecule-1 (VCAM-1) expressed by EC interacts with the integrin cognate ligand very late antigen-4 (VLA-4) on leukocytes involved in chronic inflammatory responses, such as atherosclerosis, notably monocytes and lymphocytes. EC themselves and subjacent SMC in the arterial intima can produce chemotactic and chemoattractant cytokines that can stimulate the directed migration of adherent leukocytes into the intima.[6] The recruitment of leukocytes represents an early step in the formation of atherosclerotic lesions. In animals that develop hypercholesterolemia, either through dietary or genetic manipulation, lipid deposits—composed of free cholesterol or, within cells, cholesteryl esters—also accumulate. Many have ascribed a pro-inflammatory role to low-density lipoprotein (LDL) particles that have undergone oxidation in the environment of the intima, shielded from plasma antioxidants. While various forms of oxidized lipoproteins and lipids localize within the nascent atheroma and atherosclerotic plaques, their causal role in inciting inflammation remains incompletely established.[7]

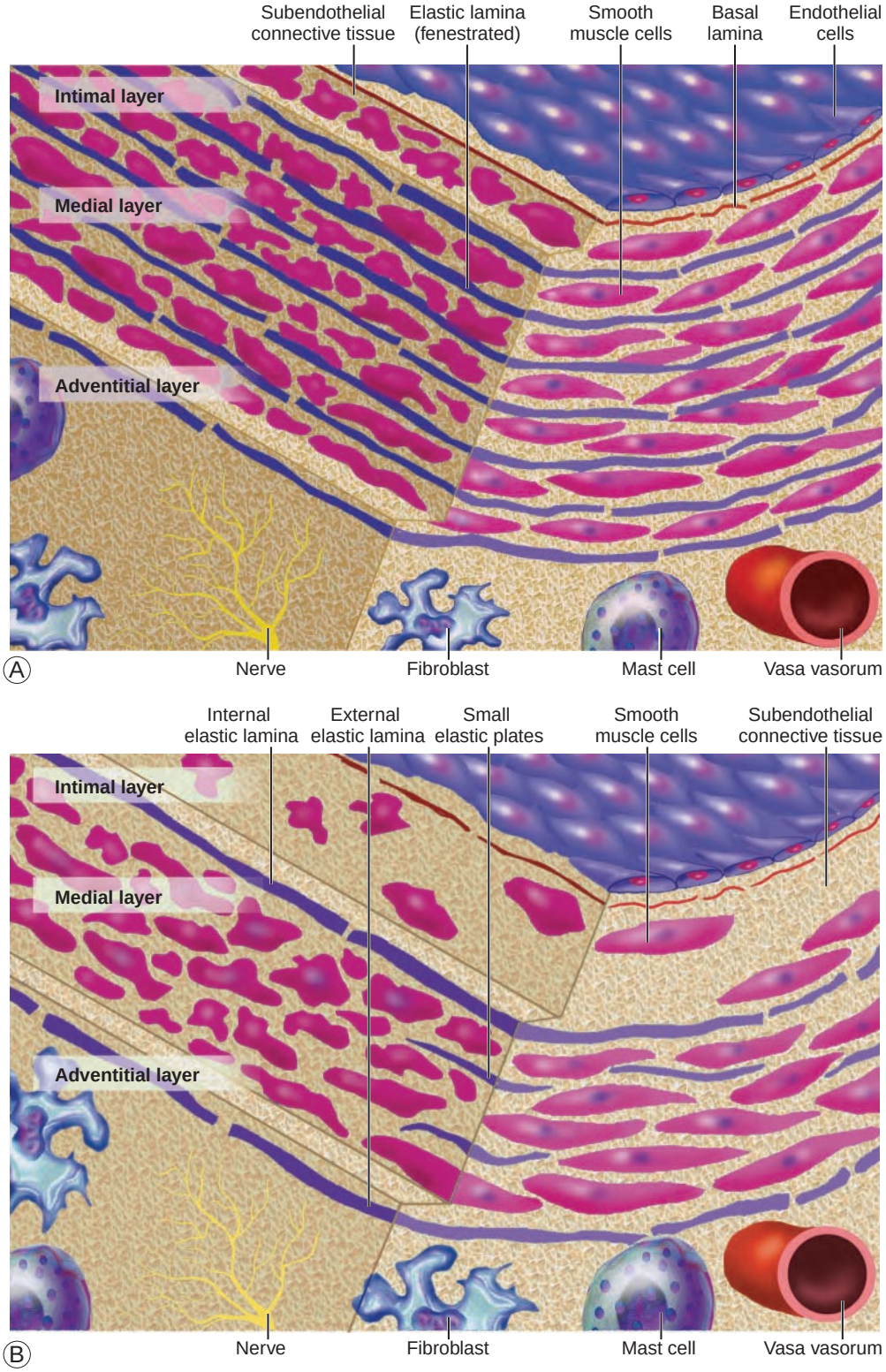


FIG. 69.1 The Structures of Normal Arteries. (A) Elastic artery. Note the concentric laminae of elastic tissue that form sandwiches with successive layers of smooth muscle cells (SMCs). Each level of the elastic arterial tree has a characteristic number of elastic laminae. (B) Muscular artery. In the muscular artery, a collagenous matrix surrounds the SMCs, but the architecture lacks the concentric rings of the well-organized elastic tissue characteristic of larger arteries. From Libby P. The Vascular Biology of Atherosclerosis. In: Libby P, Bonow RO, Mann DL, Tomaselli GF, Bhatt DL, Solomon SD, eds. *Braunwald's Heart Disease* (12th ed). Philadelphia: Elsevier 2022:425–441.

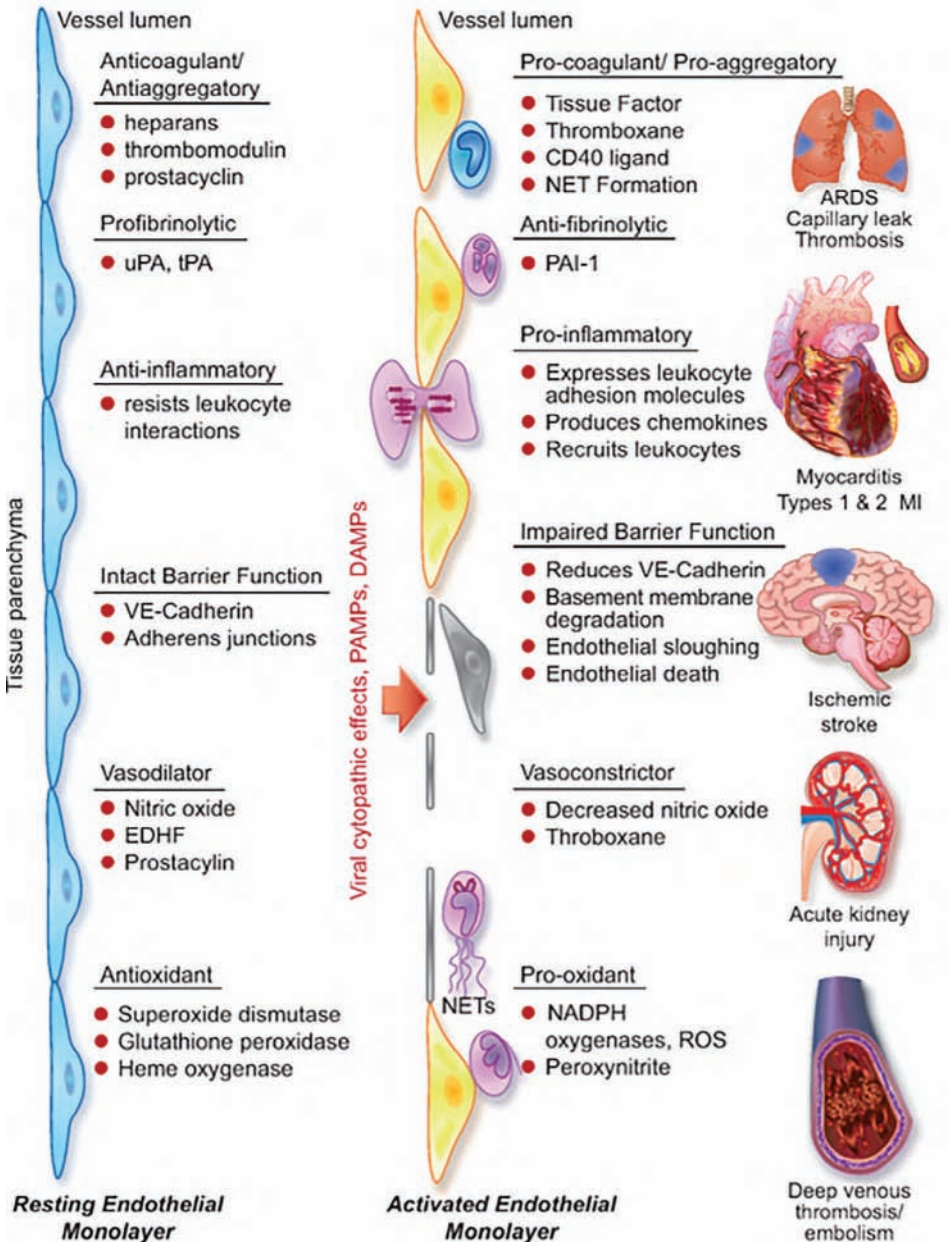


FIG. 69.2 Endothelial Activation During Acute Inflammation and Mechanisms of Organ Damage. The left side of the diagram shows a resting endothelial monolayer with the endothelial cells of squamous morphology resting on an intact basement membrane. The homeostatic mechanisms displayed by the resting endothelium include the listed properties as detailed in the text. When the endothelial cells undergo the cytopathic effect of a viral infection such as SARS-CoV-2, or encounter pathogen-associated molecular patterns (PAMPs) derived from viruses or bacteria such as lipopolysaccharide, proinflammatory cytokines such as IL-1 or TNF, or damage-associated molecular patterns (DAMPs) derived from dead or dying cells, the endothelial cells become activated. The endothelial cells display more columnar morphology. They can express adhesion molecules that attract leucocytes and chemokines that direct their migration into the subendothelial space. Sloughing of endothelial cells uncovers the thrombogenic basement membrane. Adherent neutrophils can undergo formation of neutrophil extracellular traps (NETs) that provide an amplifier for endothelial damage mediated in part by IL-1α. Inflammatory activation of endothelial cells can disrupt VE-cadherin, largely responsible for the integrity of the endothelial barrier function. Activated endothelial cells can also express matrix metalloproteinases that can degrade the basement membrane and further interrupt endothelial barrier function. In small vessels, such as those that embrace alveoli in the lung, this impaired barrier function can lead to capillary leak. These various disturbances in endothelial function, depicted in the middle part of the diagram, lead to end-organ damage including adult respiratory distress syndrome and thrombosis in the lungs, predispose to plaque rupture and thrombosis in coronary arteries, and affect the microvasculature leading to myocardial ischemia and damage. The thrombotic diathesis provoked by endothelial dysfunction can also predispose towards strokes. Microvascular as well macrovascular injury can potentiate acute renal failure. Hepatic dysfunction can also result from microvascular thrombosis among other mechanisms. Deep venous thrombosis can occur as endothelial disfunction represents an important part of Virchow's triad, and sets the stage for pulmonary embolism. Thus, loss of the endothelial protective and unleashing of the mechanisms depicted can lead to multiorgan system failure that characterizes the advanced stages of COVID-19. (From Libby P, Lüscher T. COVID-19 is, in the end, an endothelial disease. *European Heart J.*2020;41(32):3038–3044.)

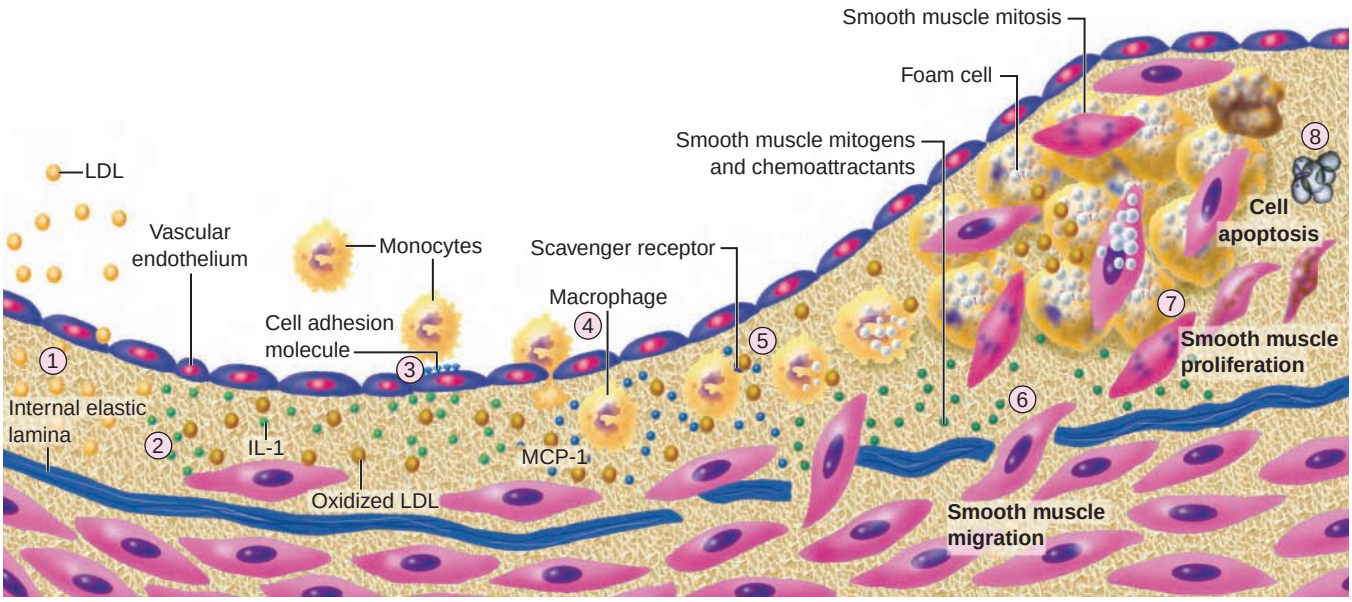


FIG. 69.3 Schematic of the Evolution of the Atherosclerotic Plaque. (1) Accumulation of lipoprotein particles in the intima *(yellow spheres)*. The modification of these lipoproteins is depicted by the darker color. Modifications include oxidation and glycation. (2) Oxidative stress, including products found in modified lipoproteins, can induce local cytokine elaboration *(green spheres)*. (3) The cytokines thus induced increase expression of adhesion molecules (*blue stalks* on endothelial surface) for leukocytes, which cause their attachment and chemoattractant molecules that direct their migration into the intima. (4) Blood monocytes, on entering the artery wall in response to chemoattractant cytokines such as monocyte chemoattractant protein 1 (MCP-1), which encounter stimuli such as macrophage colony-stimulating factor that can augment their expression of scavenger receptors. (5) Scavenger receptors mediate the uptake of modified lipoprotein particles and promote the development of foam cells. Macrophage foam cells are a source of mediators, such as cytokines and effector molecules, like hypochlorous acid, superoxide anion (O_2^-), and matrix metalloproteinases. (6) Smooth muscle cells (SMCs) migrate into the intima from the media. (7) SMCs can then divide and elaborate extracellular matrix, promoting ECM accumulation in the growing atherosclerotic plaque. In this manner, the fatty streak can evolve into a fibrofatty lesion. (8) In later stages, calcification can occur (not depicted) and fibrosis continues, sometimes accompanied by SMC death (including programmed cell death or apoptosis), yielding a relatively acellular fibrous capsule surrounding a lipid-rich core that also may contain dying or dead cells and their detritus. *IL*, Interleukin; *LDL*, low-density lipoprotein. (From Libby P. The Vascular Biology of Atherosclerosis. In: Libby P, Bonow RO, Mann DL, Tomaselli GF, Bhatt DL, Solomon SD, eds. *Braunwald's Heart Disease* (12th ed). Philadelphia: Elsevier 2022:425–441.)

Once leukocytes have taken up residence within the intima, they—as well as the intrinsic arterial wall cells—ECs and SMCs can elaborate mediators that stimulate the migration and proliferation of SMCs (see Fig. 69.3, *right hand*). Proteins such as platelet-derived growth factor (PDGF) can beckon SMC usually resident in the tunica media to enter the intima. In the normal media and under basal circumstances, the generally quiescent SMCs contain considerable contractile proteins. The SMCs that have undergone stimulation by mediators such as PDGF not only migrate and proliferate but contain less contractile protein and produce more extracellular matrix macromolecules, such as interstitial collagens and elastin as well as proteoglycans and glycosaminoglycans. These modulated SMCs elaborate much of the extracellular matrix that comprises a substantial portion of the volume of atherosclerotic plaques.

The mononuclear phagocytes are not the only leukocytes that enter the intima during atherogenesis. T lymphocytes and B lymphocytes, while less numerous than mononuclear phagocytes, also populate nascent atheromata.[8] These T lymphocytes can orchestrate pro-inflammatory and pro-oxidant functions of the more numerous macrophages through elaboration of mediators such as interferon gamma (IFN-γ). In addition to T helper 1 cells (Th1) that can elaborate IFN-γ when activated, lesions contain Th2 cells that elaborate IL-4 and IL-10. During the evolution of the atheroma, a prolonged tug-of-war between pro-inflammatory and anti-inflammatory lymphokines plays out. The role of Th17 cells in atherosclerosis remains controversial.

Lesion Progression

During these initial phases of atherogenesis described above, different classes of mononuclear leukocytes and modulated SMCs have taken up residence in the intima, accompanied by cholesterol or cholesteryl ester. Lesion complication (Fig. 69.4) involves elaboration of extracellular matrix by the activated SMCs. Transforming growth factor-beta (TGF-β), elaborated by regulatory T cells (T_{reg}), potently stimulates the production of interstitial forms of collagen by SMCs but may also limit smooth muscle proliferation. The blood monocytes mature into tissue macrophages when recruited into the atherosclerotic lesion. In mice, these mononuclear phagocytes can also proliferate during later stages of lesion development.[9] The tissue macrophages display considerable heterogeneity. Indeed, the complexity of macrophage populations determined from initial forays into single-cell messenger RNA sequencing appears considerable.[10] The tissue macrophages can express scavenger receptors that can take up modified lipoproteins promoting intracellular accumulation of cholesterol esters in lipid droplets within the cytoplasm.[4] The classical

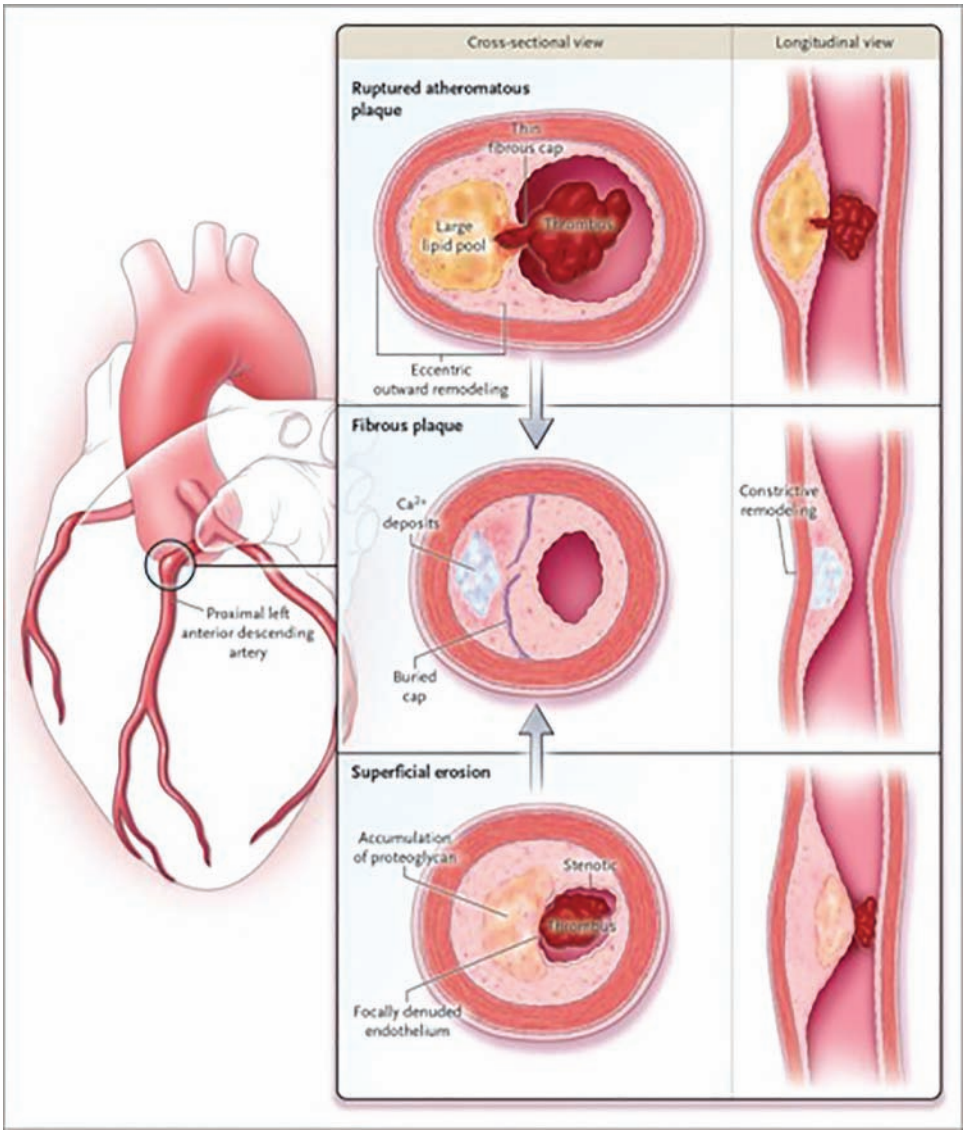


FIG. 69.4 Characteristics of Atherosclerotic Plaques Associated with Various Presentations of Coronary Artery Disease. This figure depicts the characteristic morphologic features of coronary atherosclerotic plaques that associate with three distinct clinical presentations. The representation of the cardiac surface *(left)* shows an atherosclerotic plaque affecting the proximal left anterior descending coronary artery. The cross-sectional and longitudinal views of the artery portray typical lesion types in greater detail. The two top images show an eccentric, positively remodeled atheroma with a thin fibrous cap that has fractured and caused a thrombus to form. The healing of plaques that have disrupted can yield the formation of a more fibrous lesion (shown in the middle pair of images). Stenotic, fibrous plaque can cause stable ischemic syndromes (*e.g.*, demand angina pectoris) as a result of narrowing of the arterial lumen. In this situation, plaques can display "buried caps," the result of a prior disruption of the fibrous cap that triggered thrombosis, followed by healing, fibrosis, and often constrictive (inward) remodeling. This process can promote the progression of a nonocclusive, atheromatous, lipid-rich plaque to a stenotic, more fibrous, calcified plaque. The bottom pair of images shows a proteoglycan-rich plaque that instigated thrombosis due to superficial erosion of the endothelial monolayer. (Libby P. Mechanisms of acute coronary syndromes and their implications for therapy. *N Engl J Med.* 2013;368:2004–2013.)

LDL receptor undergoes tight regulation by intracellular cholesterol levels under the control of the sterol regulatory element proteins (SREBPs). Thus, it is difficult to load cells with cholesterol through the classical LDL receptor as the very accumulation of cholesterol inhibits the expression of the usual LDL receptor. Hence, the importance of the scavenger receptors, which are lower affinity and high-capacity receptors that can mediate the internalization of modified cholesterol, bypassing the tightly regulated LDL receptor.

Lipid-laden macrophages can undergo cell death by apoptosis or pyroptosis. When these cells die, they extrude their lipid content into the extracellular space. Cholesterol crystals composed of cholesterol monohydrate can form both within macrophages and in the extracellular space. The cholesterol crystals can coactivate the inflammasome that activates interleukin (IL)-1-beta and can thus promote pyroptosis. Generally, apoptotic cells do not tarry in tissues because of rapid phagocytosis, and therefore, do not incite inflammatory responses as do necrotic cells. However, within

atherosclerotic plaques, it appears that phagocytosis of dead or dying cells, a process denoted efferocytosis, works inefficiently leading to the buildup of inflammatory cellular debris within the plaque. Ultimately, the accumulation of lipid and cellular detritus lead to formation of a central lipid or necrotic core in the evolving atherosclerotic plaque. Typically, a fibrous cap comprised of SMC and interstitial collagen overlies the lipid core. These various mechanisms lead to the formation of the mature lipid-rich atherosclerotic plaque (see Fig. 69.4).

Under normal circumstances, microvessels perfuse only the outer third of the adventitia of arteries. As atherosclerotic plaques evolve, they can acquire a microcirculation due to angiogenesis.[11,12] The neovessels that form at the base of plaques in this manner exhibit friability and can have impaired barrier function, favoring the extravasation of erythrocytes and stimulating local thrombi within the vascular walls, providing another source of platelet-derived mediators that can stimulate SMC proliferation and extracellular matrix production.

Lesion Complication

Atherosclerotic plaques can progress to form arterial narrowings or stenoses that can limit blood flow, causing lack of supply of oxygen (ischemia) due to impeded blood flow in the areas subtended by the stenosed vessel (see Fig. 69.4, *middle*). Such stenoses can arise by SMC proliferation and accumulation of extracellular matrix molecules, such as interstitial collagens, elastins, proteoglycan, and glycosaminoglycans. Yet, such progressive growth to flow-limiting lesions, usually accounts for chronic ischemia, producing chest discomfort (angina pectoris), under conditions of increased oxygen demand due, for example, to physical effort or mental stress. Such gradual narrowing of the artery *per se* seldom leads to the acute thrombotic complications of atherosclerosis that cause acute coronary syndromes, such as myocardial infarction. Rather, a physical disruption of the atherosclerotic plaque most frequently provokes the blood clots that precipitate acute coronary syndromes (see Fig. 69.4).[13] The most common form of plaque disruption involves a frank rupture or fissuring of the plaque's fibrous cap that overlies the lipid core (see Fig. 69.4, *top*). Indeed, lesions that have provoked a fatal thrombus often have a fractured, thin fibrous cap that overlies a lipid core rich in procoagulants, such as tissue factor produced by plaque macrophages and SMC. When the coagulation proteins in blood gain access to the procoagulants within the plaque due to a fissure, a thrombus can rapidly ensue.

The macrophages within the plaque can produce enzymes that degrade the constituents of the arterial extracellular matrix when stimulated by pro-inflammatory mediators. These enzymes include the matrix metalloproteinases (MMPs) specialized in breakdown of interstitial collagens (MMPs 1,8, and 13) and cysteinyl proteinases such as cathepsins S, K, or L that can degrade elastin. The overproduction of these enzymes by macrophages within the plaque can disrupt the usually protective fibrous cap, rendering it thin, friable, and susceptible to rupture, thus setting the stage for a fracture and precipitation of an acute thrombotic event (Fig. 69.5).[13] IFN-γ elaborated by activated plaque T lymphocytes can inhibit the production of new collagen by SMC required to repair and maintain the protective extracellular matrix of the fibrous cap. IFN-γ can also enhance the proteolytic, oxidative, and pro-inflammatory functions of lesional macrophages, further destabilizing the plaque. A rupture of the plaque's fibrous cap causes two-thirds to three-quarters of acute coronary syndromes.

Another mechanism known as superficial erosion, can also cause acute thrombotic complications of atherosclerotic plaques (see Fig. 69.4, *bottom*).[14] Indeed, in an era of effective lowering of LDL, and with improved treatment of hypertension and more widespread smoking cessation, plaque rupture due to the thin-capped lipid-rich atheroma appears to be becoming less prevalent.[15] Superficial erosion now accounts for a quarter to a third of acute coronary syndromes.

Superficial erosion also appears to depend on innate immune processes. For example, activation of Toll-like receptor-2 (TLR-2) on endothelial cells, perhaps in response to fragments of hyaluronic acid that form within the subjacent intima, can sensitize SMC to apoptosis, sloughing, and impaired ability to repair a rent in the monolayer.[16] When patches of EC slough, they expose basement membrane, which can activate platelet accumulation. Polymorphonuclear leukocytes can accumulate at these sites and form neutrophil extracellular traps (NETs). These filamentous structures consist of strands of the granulocyte's nuclear DNA that unwind when relieved of constraints into nucleosomes.[17] NETs bind the constituents of polymorphonuclear leukocyte granules and also capture tissue factor from the blood, presenting these pro-inflammatory, pro-oxidant, and pro-thrombotic mediators at the surface of the endothelium.[18] NETs thus serve as a solid-state reactor that favors thrombus formation and accumulation. Plaque disruption due to fracture of the fibrous cap or to superficial erosion due to desquamation of lumenal EC underly the vast majority of acute coronary syndromes. Similar mechanisms may pertain to thromboses responsible for many ischemic strokes that arise from the extracranial cerebral arteries and may also complicate aortic atheromata.

Therapeutic Implications of Inflammation in Atherosclerosis

The scientific literature is replete with experimental studies that support a role for various pro-inflammatory pathways—both activation of innate and adaptive immunity—in atherosclerosis in all phases. Moreover, biomarker studies in humans support a relationship between inflammatory status and the manifestations of atherosclerosis, including thrombotic complications. C-reactive protein—measured by a highly sensitive assay (hsCRP)—remains the best validated of these biomarkers of inflammation, but the literature with other indices of inflammation build a highly congruent case that inflammation correlates with risk of atherosclerosis and its complications. Yet, until recently, the causality of inflammation in atherosclerosis remained an open question. The Canakinumab Anti-Inflammatory Thrombosis Outcome Study (CANTOS) was the first to close this loop of causality in humans.[19] CANTOS enrolled over 10,000 individuals who had sustained an acute coronary syndrome and received state-of-the art treatment mandated by guidelines for preventing recurrent events. Yet, despite treatment with highly effective LDL-lowering statins, aspirin, beta-adrenergic blocking agents, and when appropriate inhibitors of the renin angiotensin system, CANTOS selected individuals who had an indication of residual inflammation as disclosed by an hsCRP above median for the population, 2 mg/L. The participants randomly received placebo or one of three doses of canakinumab, a human monoclonal antibody that selectively neutralizes IL-1β. After a median follow up of 3.7 years, canakinumab produced a significant 15% reduction in recurrent myocardial infarction or stroke or cardiovascular death. This result substantiated the causality of

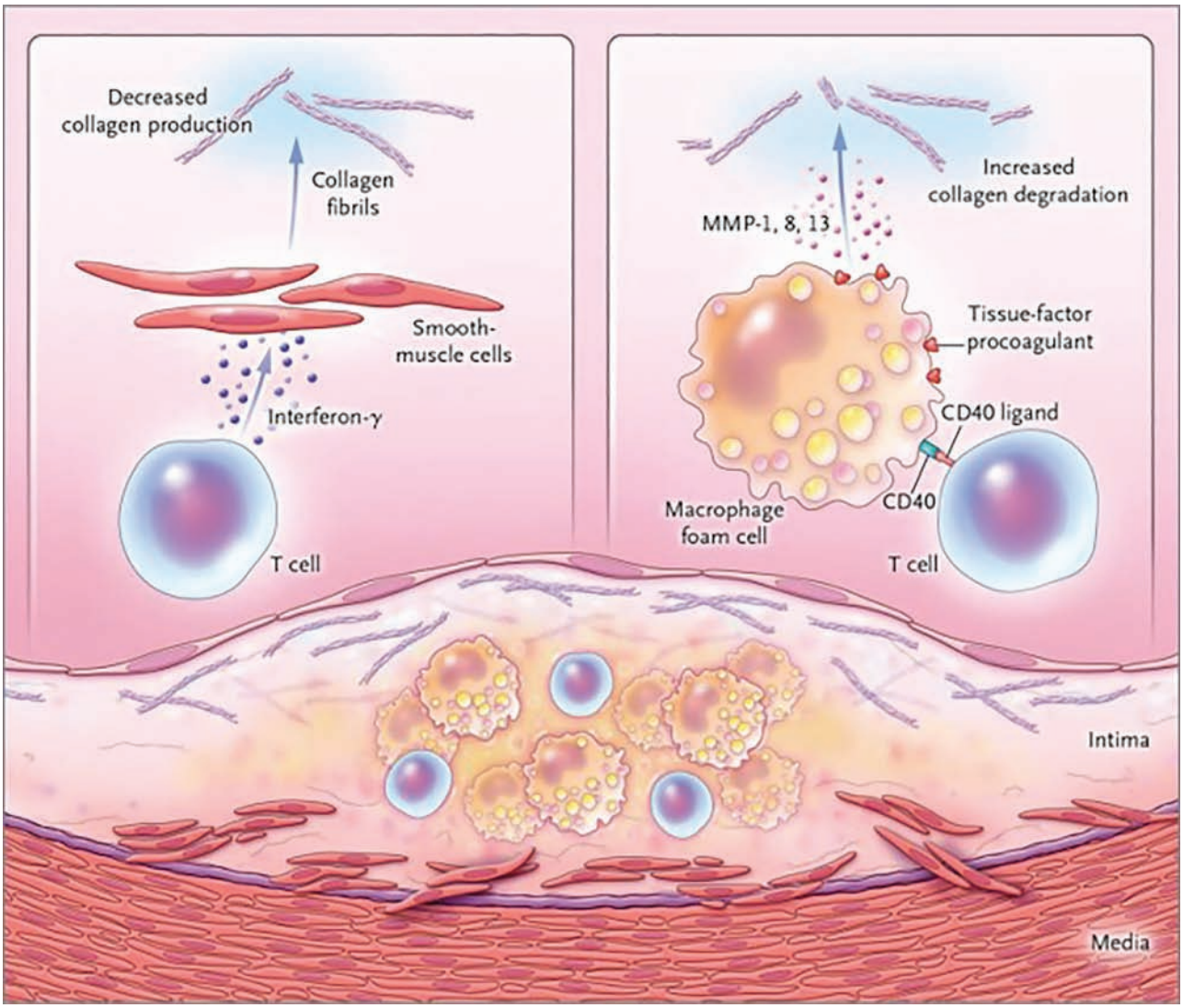


FIG. 69.5 **Inflammatory Pathways Predisposing Coronary Arteries to Rupture and Thrombosis.** A cross-section of an atheromatous plaque at the bottom of the figure shows the central lipid-rich core that typically contains macrophage foam cells *(yellow)* and T lymphocytes *(blue)*. The intimal and medial layers also contain arterial smooth-muscle cells *(red)*, which elaborate the arterial extracellular matrix including interstitial collagen (depicted as triple helical coiled structures). Activated T cells (of the Th1 helper T-cell subtype) secrete interferon-γ. This cytokine inhibits the synthesis of the new collagen that confers strength upon the plaque's protective fibrous cap *(upper left)*. The T lymphocytes can also stimulate lesional macrophages to express CD40 ligand (CD154), which ligates its receptor (CD40). This inflammatory signal provokes the overproduction of interstitial collagenases (matrix metalloproteinases [MMPs] 1, 8, and 13) that can attack the usually very stable collagen triple helix that lends tensile strength to the plaques cap *(top right)*. CD40 binding can also elicit overproduction of the potent procoagulant tissue-factor by macrophages. In this manner, inflammatory signaling places the collagen in the plaque's fibrous cap in double jeopardy—decreased synthesis and increased degradation—a recipe for rendering the fibrous cap liable to rupture. The augmented tissue-factor production in response to inflammatory stimulation favors thrombus formation in the ruptured plaque. These mechanisms contribute to inflammations ability to trigger the thrombotic complications of atherosclerosis such as the acute coronary syndromes. (Libby P. Mechanisms of acute coronary syndromes and their implications for therapy. *N Engl J Med*. 2013;368(21):2004–2013.)

pro-inflammatory immune responses in human atherosclerotic complications. In a pre-specified on-treatment analysis, individuals treated with canakinumab who achieved a greater than median drop in hsCRP showed an over 30% reduction in cardiovascular and total mortality (Fig. 69. 6). Canakinumab was associated with a small, but statistically significant, increase in infections, including fatal infections. The infections probably arose from impaired host defenses due to IL-1β blockade. The increase in infections were balanced by a remarkable reduction in the incidence and deaths due to cancer, primarily lung cancer, in the enrolled population. This observation has spawned numerous studies of neutralization of IL-1β as an adjutant therapy to combat lung cancer. A number of mechanisms may contribute to these various benefits of IL-1β neutralization observed in CANTOS (Fig. 69.7).

The canakinumab-treated patients in CANTOS exhibited many other benefits, including reduced hospitalizations for heart failure and less frequent incidence of anemia. Arthritis

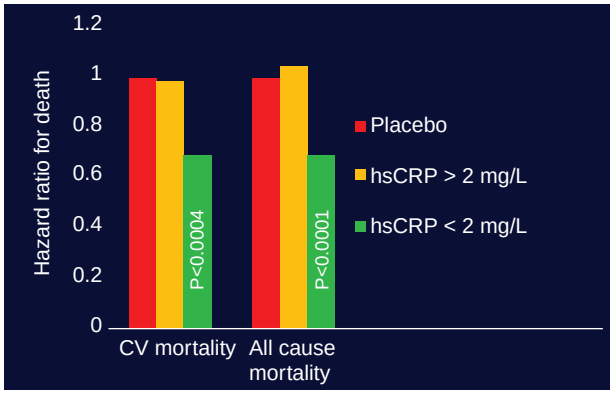


FIG. 69.6 Responders to Canakinumab had Mortality Benefit in the On-Treatment Analysis from CANTOS. Multivariable adjusted hazard ratios (HR) for prespecified cardiovascular outcomes according to on-treatment, high-sensitivity C-reactive protein (hsCRP) levels above or below 2 mg/L after drug initiation. HRs adjusted for age, gender, smoking, hypertension, diabetes, body mass index, baseline hsCRP, baseline low density lipoprotein-cholesterol. This on-treatment analysis underwent numerous sensitivity analyses to assess possible confounding that affirmed the conclusion. (Data from Ridker PM, MacFadyen JG, Everett BM, et al. Relationship of C-reactive protein reduction to cardiovascular event reduction following treatment with canakinumab: a secondary analysis from the CANTOS randomised controlled trial. *Lancet*. 2018;391(10118):319–328.)

symptoms declined strikingly in the canakinumab-treated patients, and acute gouty arthritis fell by over 50%. The CANTOS trial thus opened a new window on atherosclerosis therapy by proving that a directed anti-inflammatory intervention that did not alter atherogenic lipoproteins could provide cardiovascular benefit on top of all standard treatments.

Another anti-inflammatory intervention, colchicine, has also shown benefit in patients with coronary artery disease. The COLCOT study demonstrated a greater than 25% reduction in recurrent cardiovascular events in patients within a month of acute myocardial infarction who randomly received low-dose colchicine 0.5 mg daily.[20] The LoDoCo2 study showed a similar reduction in events in patients with stable coronary artery disease with a like colchicine treatment.[21] The mechanisms of colchicine's anti-inflammatory effects remain incompletely understood.

Thus, we now possess information that supports the efficacy of two anti-inflammatory interventions in atherosclerosis. Numerous possibilities and questions remain in moving the field of anti-inflammatory therapies in atherosclerosis forward.[22,23] Could allocation of anti-inflammatory therapies to individuals who have not yet sustained an acute coronary syndrome prove beneficial? What biomarkers would permit the precise allocation of an anti-inflammatory intervention in this circumstance?

One possibility arises from the recent observation that individuals who have clones of leukocytes that arise from somatic mutations in a subset of known leukemia driver genes have a markedly accentuated risk of atherosclerotic events.[24]

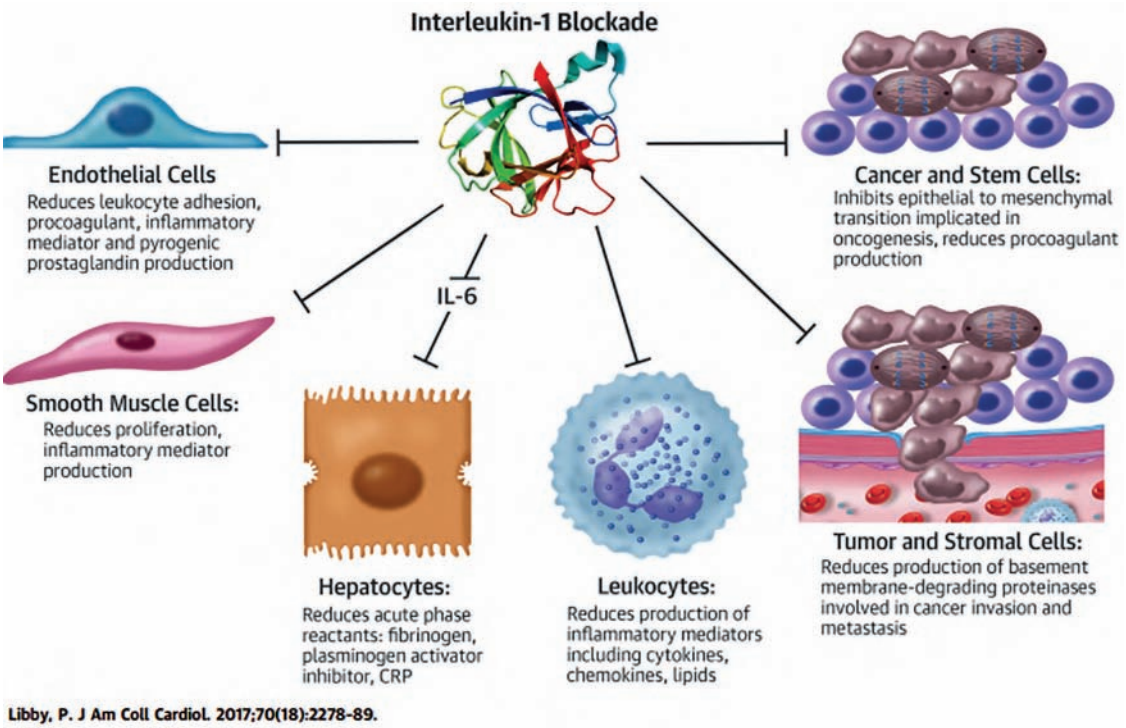


FIG. 69.7 Some Effects of IL-1 Blockade on Cellular Functions. Interleukin (IL)-1 alters many functions of cells that can participate in disease pathogenesis, including such conditions as atherosclerosis, thrombosis, oncogenesis, and invasion and metastasis of tumors. A number of the actions of IL-1 on hepatocytes, among them the induction of the acute phase response, depend on IL-6, a cytokine downstream of a protein induced by IL-1. Concordant evidence from human genetics supports the causality of IL-6 in atherothrombosis. CRP = C-reactive protein. (From Libby, P. Interleukin-1 beta as a target for atherosclerosis therapy: biological basis of CANTOS and beyond. *J Am Coll Cardiol*. 2017;70(18):2278–2289.)

Experimental studies have documented an increase in the inflammasome–IL-1β–IL-6 pathway in animals engineered to resemble humans with clonal hematopoiesis due to somatic mutations. Individuals who inherit a variant in the IL-6 receptor that interferes with IL-6 signaling have a remarkable decrease in atherosclerosis events compared to individuals who lack the low-function variant, but only if they bear clonal hematopoiesis mutations.[25] Thus, individuals found to have clonal hematopoiesis due to these mutations could benefit from a therapy that interrupts IL-6 signaling. Such a genotype-directed anti-inflammatory intervention would represent an example of a precision therapy that would limit treatment to those with particular propensity to benefit.

Many other anti-inflammatory therapies have become candidates for clinical evaluation in the context of atherosclerosis. These include pro-inflammatory cytokines, as well as a number of inhibitors of other pro-inflammatory pathways or mediators.[23] Yet, pharmacologic inhibition of p38MAP kinase and inhibitors of various phospholipases that give rise to pro-inflammatory phospholipids have not reduced cardiovascular events in rigorous randomized trials. A pilot trial is evaluating manipulation of adaptive immunity in atherosclerosis. Low-dose IL-2 can stimulate the activity of regulatory T cells that may limit inflammation during atherosclerosis.[26] Vaccination furnishes another strategy that manipulates adaptive immunity that might modulate atherosclerosis.[27] Indeed, a number of approaches to vaccinating against apolipoprotein B, the major apolipoprotein of LDL, have undergone development, although no clinical trials have yet proven efficacy of a vaccination approach.

In conclusion, we now have abundant experimental evidence that both adaptive and innate immunity operate in atherosclerosis. Recent clinical and human genetic studies provide direct support for the operation of innate immune pathways in human atherosclerosis. A number of therapeutic avenues that target inflammation merit further investigation to make inroads against the unacceptable residual burden of risk despite contemporary standard of care. Such approaches might ultimately prevent recurrent events and those with established atherosclerosis, but also forestall the initial manifestations of atherosclerosis in selected individuals.

MYOCARDITIS: IMMUNE RESPONSE THAT DAMAGES THE HEART

Immune Status of the Heart Under Homeostatic Conditions

Human life depends on uninterrupted electrochemical function of the heart, and that function depends on a non-redundant anatomy of myocardial cells that have essentially no regenerative capacity. Innate and adaptive immune responses invariably lead to some degree of tissue damage, even if they are initiated by microbial infections. Therefore, these responses pose a substantial danger to uninterrupted cardiac function. The dense microvascularity of the heart—needed for efficient delivery of oxygen to contracting myocytes—carries the liability of easy access for circulating immune cells to enter the myocardium. These risks have likely exerted selective pressures for the coevolution of immune-suppressive mechanisms that raise the bar for initiation of immune responses in the heart. Such mechanisms would block the accidental activation of circulating T cells, including autoreactive T cells, that may migrate into the heart at a low frequency. A relative "immune-privileged" state of the heart is reflected in the low number of T cells present in the healthy myocardium, unlike the chronic "physiologic" inflammation characteristics of many other organs and tissues.

Although the mechanisms for the relatively privileged state of immunity in the healthy heart is not well understood, several cellular and molecular components likely contribute. One mechanism is the presence of tissue-resident macrophages, which populate the myocardium during fetal development and acquire anti-inflammatory homeostatic properties from cardiac tissue-specific signals.[28] Furthermore, mediators generated during an innate immune response or early during adaptive T cell–mediated responses rapidly induce immunoregulatory mechanisms in the myocardium. These include the induction of endothelial expression of the immune checkpoint molecule PD-L1 by type 1 interferons and interferon-γ, which serves to limit reactivation of effector T cells that enter the heart.[29] These mechanisms may fail in the setting of certain cardiotropic microbial infections, a genetic susceptibility for autoimmunity, or a combination of both. When adaptive immune responses in the heart do occur, they can manifest as a heterogeneous spectrum of clinical syndromes all grouped under myocarditis. Initial innate immune responses can evoke dysfunctional adaptive immune responses, including effector T cells and antibodies that may then engage innate immune effector molecules and cells to contribute to structural damage.

Inflammation in the myocardium in older adults most often arises from innate immune responses to ischemic injury resulting from coronary artery disease, less commonly at all ages by innate and adaptive responses to infection and more rarely as a manifestation of autoimmunity. Adaptive immune responses in the heart, mediated by antibodies and T cells, may be specific for microbial antigens (*e.g.*, the Streptococcus in rheumatic heart disease), or in the case of autoimmunity, myocardial antigens. In all cases, myocardial immune responses may lead to acute loss of heart function and often result in structural changes that lead to dilated cardiomyopathy with chronic heart failure.

Myocarditis

Myocarditis includes several different clinical-pathological conditions with the common characteristic of myocardial inflammation, most often involving lymphocytes and macrophages, that causes injury and functional impairment of the myocardium (Table 69.1).[30,31] The term myocarditis is not used to describe the inflammatory response that rapidly develops after ischemic injury. The clinical presentation of myocarditis is highly variable, depending on the underlying cause and when the patient comes to medical attention. Signs and symptoms often include fatigue, shortness of breath, chest pain, palpitations, electrocardiographic (ECG) findings of arrythmias and atrioventricular block, echocardiographic or MRI findings of right or left ventricular wall motion defects, and valvular dysfunction, elevated serum markers (troponins, erythrocyte sedimentation rate and C-reactive protein). Most of these findings can reflect ischemic heart disease unrelated to myocarditis, and often, the presumptive diagnosis of myocarditis is one of exclusion after coronary artery disease, myocardial infarction, and heart failure due to other causes have been ruled out. Details of different modes of presentation are described in Cardiology textbooks. Constellations of symptoms, physical exam, ECG, imaging, and clinical laboratory tests can only lead to a diagnosis of clinically suspected myocarditis. A definitive diagnosis of myocarditis,

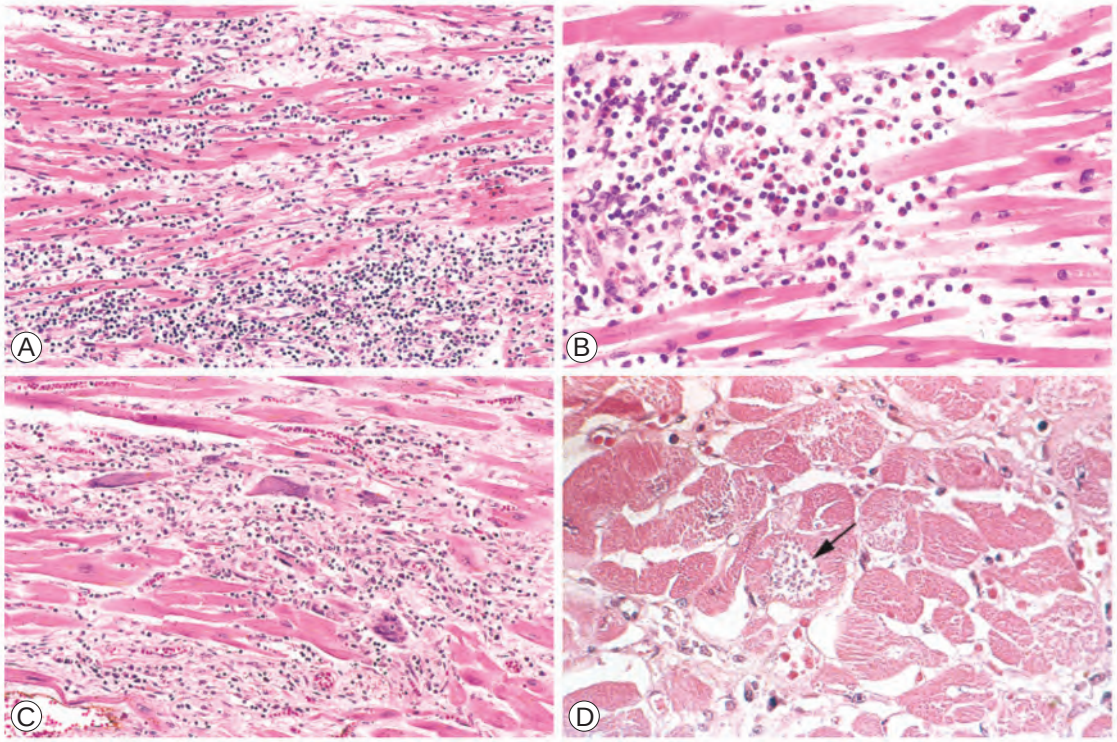


FIG. 69.8 Myocarditis. (A) Lymphocytic myocarditis, associated with myocyte injury. (B) Hypersensitivity myocarditis, characterized by interstitial inflammatory infiltrate composed largely of eosinophils and mononuclear inflammatory cells, predominantly localized to perivascular and expanded interstitial spaces. (C) Giant-cell myocarditis, with mononuclear inflammatory infiltrate containing lymphocytes and macrophages, extensive loss of muscle, and multinucleated giant cells. (D) The myocarditis of Chagas disease. A myofiber distended with trypanosomes *(arrow)* is present along with inflammation and necrosis of individual myofibers. (From Kumar V, Abbas AK, Fausto N, Aster, JC. *Robbins and Cotran pathologic basis of disease*. Chapter 12, 8th ed. Philadelphia: Elsevier; 2020.)

and subclassification of the type of myocarditis, requires histopathologic examination of endomyocardial biopsies (EMBs) or heart issue obtained at autopsy (Fig. 69.8). Specific criteria for the pathologic diagnosis of different types of myocarditis, also called inflammatory cardiomyopathies, have been published and updated (Table 69.2).[32] These criteria include the immunohistochemical identification and quantification of infiltrating T and B lymphocytes, monocytes or macrophages, neutrophils, and eosinophils, accompanied by assessment of infectious agents by polymerase chain reaction (PCR) or *in situ* hybridization. However, EMB is not done routinely in many medical centers, and while generally quite safe in experienced hands, may have increased risks in very ill patients with acutely dilated hearts. Furthermore, EMB-based diagnosis can have low sensitivity due to sampling issues, if the inflammatory process in the myocardium is not diffuse. Therefore, our knowledge of the etiology and immune mechanisms of human myocarditis still relies in large part on correlations with extracardiac immune variables (blood cell counts and phenotype, plasma proteins) and extrapolations from animal research. The following discussion will focus on the immunological features of different forms of human myocarditis, including those of known association with infections, and those without known infectious causes.

Infectious Myocarditis

Infections of the myocardium appropriately fall into a discussion of immune-mediated cardiovascular diseases for two principal reasons. First, much of the pathophysiology of infections in the heart result from the immune response to the etiologic agent, beyond their direct cytopathic effects. Second, autoimmune myocarditis, in which the adaptive immune system specifically targets cardiac antigens, appears to frequently arise as a complication of myocardial infections, especially virial myocarditis. All classes of microbes can infect cardiac tissue and cause an inflammatory response that impairs cardiac function (Table 69.1). Most cases of myocarditis in North America, Europe, and Asia are a result of viral infections, but in South and Central America, *Trypanosoma cruzi* is the most frequent cause of acute and chronic myocarditis.

Viral myocarditis is diagnosed in patients with clinical findings generally associated with myocarditis (discussed above) paired with evidence of acute viral infection, such as increased titers of virus-specific antibodies and detection of viral antigens or nucleic acids in the blood, or molecular detection of viral nucleic acid or immunohistochemical detection of viral proteins in EMBs. Historically, the most common viruses associated by serology with acute myocarditis have been Coxsackievirus, other enteroviruses, adenoviruses, hepatitis C virus, cytomegalovirus, and influenza virus. Currently, the most frequently identified viral genomes by PCR of EMBs are parvovirus B19 and Epstein-Barr virus, hepatitis C virus, and Human Immunodeficiency Virus (HIV). PCR-based detection of viral nucleic acids in myocarditic hearts does not prove the etiologic role of those viruses—especially common endemic viruses, such as parvovirus—but ongoing viral genome studies of EMB samples may shed further light on this question.

TABLE 69.1 Causes of Myocarditis

Cause	Clinical/Pathologic Syndrome
Infections of Myocardium	
Viruses (e.g., coxsackievirus, ECHO, influenza, HIV, Epstein-Barr virus, cytomegalovirus, parvovirus)	Acute and chronic viral myocarditis, chronic; dilated cardiomyopathy
Bacteria (e.g., rickettsia)	Cardiac rickettsial vasculitis
Helminths (e.g., trichinosis)	Cardiac trichinellosis
Fungi (e.g., Candida)	Cardiac candidiasis
Protozoa (e.g., *Trypanosoma cruzi*, *Toxoplasma gondii*	Acute and chronic Chagasic myocarditisCardiac toxoplasmosis
Effects of non-cardiac infection on the heart	
Corynbacterium diphtheriae	Diphtheriae toxin myocarditis
Neisseria meningitidis	Meningococcal myocarditis
Immune-mediated	
Post-viral autoimmune	Chronic lymphocytic myocarditis, dilated cardiomyopathy
Poststreptococcal autoimmune	Rheumatic fever/carditis, chronic rheumatic heart disease
Hypersensitivity (e.g., drug reactions)	Eosinophilic myocarditis
Unknown trigger	Sarcoidosis, isolated cardiac
Unknown trigger	Giant cell myocarditis
Unknown trigger	Necrotizing eosinophilic myocarditis
Manifestations of systemic autoimmune disease	Systemic lupus erythematosus cardiitis, Cardiac involvement in systemic sarcoidosis,Giant cell myocarditis with IBD, other autoimmune diseases, Necrotizing eosinophilic myocarditis, Eosinophilic granulomatosis with polyangiitis (EGPA)

TABLE 69.2 Histopathological Classifications of Myocarditis

Histopathological Type	Histopathology	Most Common Clinical Associations
Acute lymphocytic myocarditis	Patchy infiltration of CD3+ T lymphocytes, with minimal fibrosis	Viral and autoimmune
Chronic lymphocytic myocarditis	Myocardial fibrosis accompanied by lymphocytic infiltration.	Viral and autoimmune
Giant cell myocarditis	Extensive areas of inflammation, predominantly with CD68+ macrophages, often multinucleated giant cells, as well as T cells and eosinophils, with associated myocardial necrosis.	Isolated or autoimmune disease
Eosinophilic myocarditis	Mixed lymphocytic and myeloid infiltrate with significant numbers of eosinophils; little myocardial damage in hypersensitivity form of disease; marked necrosis in rare form of disease.	Drug hypersensitivity or associated with autoimmune disease with eosinophilia (*e.g.*, Churg Strauss)
Granulomatous: Sarcoidosis	Noncaseating granulomas and lymphocytic infiltrate and fibrosis	Isolated or systemic sarcoidosis
Neutrophilic myocarditis	Neutrophilic abscesses often with bacteria	Systemic bacterial infection

The many different virus species identified as causative agents in myocarditis cases exert highly diverse cytopathic effects, including variability in cell types that are infected and injured. Nonetheless, the fundamental steps in the initiation and progression of anti-viral immune responses, which contribute to the pathophysiology of infection, appear to be similar across different viral infections of the heart.[33,34] The innate response to viruses in any tissue begins with recognition of viral nucleic acids by pattern recognition receptors (PRRs), including endosomal Toll-like receptors (TLRs)-3,-7,-8, and -9; cytosolic RNA sensors such as Rig-like receptors (RLRs) and nucleotide-binding oligomerization domain 2 (NOD2); and cytoplasmic DNA sensors that engage the cGAS-STING pathway.

Mouse studies have shown that either RNA and DNA viruses can cause essentially indistinguishable forms of myocarditis in the same strain of mice, which argues for the stereotypical innate anti-viral response that initiates disease in response to different viruses. Some experimental evidence implicates many of these receptors in both protection against viral infection, as well as pathological inflammation caused by virus, mostly in mouse Coxsackievirus (CV) B3 myocarditis.[35] The role of these innate-pattern recognition pathways in human viral myocarditis derives support from circumstantial evidence of expression and increases in several of the components of the pathways in human myocarditic hearts. Each of these pathways can provide anti-viral protection through activation of interferon-response factors (IRFs), which increase expression of type-1 interferon genes. In addition, most of these pathways also enhance inflammation through activation of Nuclear Factor kappa B (NF-κB), which enhances expression of acute inflammatory cytokines and factors that initiate adaptive immune responses. Such mouse studies illustrate a precarious balance between early innate immune protection against the virus and inflammation that results in acute and chronic myocardial injury.

The histopathological form of myocarditis that associates most frequently with viral cardiac infection is lymphocytic myocarditis. The adaptive immune response to viruses in mouse and human hearts is characterized by infiltrating effector T cells, B cells, and monocyte-derived macrophages, with foci of myocyte necrosis. Immunohistochemical interrogation of endomyocardial biopsies (EMBs) has established accumulation of both CD8 and CD4 T cells into hearts of patients with viral myocarditis, consistent with studies in mouse CVB3 myocarditis.[36] Mouse studies have implicated a pathogenic role of perforin/granzyme mediated CTL killing of myocytes, as well as inflammatory injury driven by CD4 Th1 and Th17 cells. CD8 and CD4 T cells are inferred to be pathogenic in human viral myocarditis hearts as well, based on the detection of these effector cells in tissue samples. Antibodies specific for many different myocardial proteins are found in the blood and in heart tissues of mice and humans with viral myocarditis (discussed below), but there is little evidence that they mediate substantial injury.

SARSCoV2, the etiologic agent of COVID-19, associates strongly with cardiac damage, as determined by cardiac troponin elevation and indices of edema and fibrosis in cardiac magnetic resonance studies.[37] Cardiac pericytes—the cells that invest the coronary microvasculature—express angiotensin-converting enzyme 2 (ACE-2), the enzyme that SARSCoV2 uses as a receptor. Some reports describe a fulminant myocarditis associated with SARSCoV2. Yet, actual infection of cardiac myocytes and a clear-cut myocarditis due to SARSCoV2 seems at present less common as a cause of myocardial damage in those with COVID-19 than an imbalance between oxygen supply and demand. Dysfunction of the endothelium does contribute to many complications of COVID-19.

T. cruzi infection (Chagas disease) is the most common infectious cause of myocarditis worldwide, with most cases occurring in South and Central America and Mexico.[38] Approximately 30% of infected people will develop some degree of chronic Chagasic cardiomyopathy (CCC). Acute myocarditis caused by *T. cruzi* infection of cardiac myocytes is usually mild or asymptomatic when transmitted by the insect vector (triatomine bug) bite but is often severe when transmitted by ingestion of contaminated food or juice, often leading to fulminant disease with heart failure and up to 10% mortality. In this acute form of the disease, there appear to be strong innate and adaptive immune responses, likely initialized by direct myocytopathic effects of the organism. Inflammatory infiltrates in acute chagasic myocarditis include neutrophils, NK cells, macrophages, T cells, and plasma cells. Most Chagas heart disease does not begin with symptomatic acute myocarditis but proceeds chronically after asymptomatic infection with progressive scarring of the myocardium over 10 to 20 years, eventually causing arrhythmias and biventricular heart failure without a known history of acute disease. CCC appears to be driven by persistent infection, documented by histology and molecular techniques, causing chronic inflammation, which damages tissue but fails to eradicate the infection. In murine studies, CD8 cytotoxic T lymphocytes are required to control infection, and an effective CTL response requires Th17 cells or other sources of IL-17. In humans, the CTL response becomes ineffective over time and the T cells acquire a dysfunctional or exhausted phenotype. Although autoantibodies specific for cardiac antigens can be detected in CCC patients, there is no evidence that they are pathogenic.

Autoimmune Myocarditis

Autoimmune myocarditis can develop as a consequence of known microbial infection, as one component of multiorgan/systemic autoimmune disease, as an isolated heart-only autoimmune disease with unknown connection with infection, or as a complication of cancer immunotherapy. In all cases, there is a failure in tolerance to cardiac antigens, but as is the case for most autoimmune diseases, the mechanisms of that failure are not well understood.[39] Self-tolerance to cardiac antigens requires a combination of central and peripheral tolerance mechanisms. Central tolerance to cardiac antigens is induced during B- and T-cell development in bone marrow or thymus, respectively, by deletion of immature lymphocyte clones specific for self-antigens. Peripheral tolerance is needed to prevent activation of naïve cardiac-specific lymphocyte clones that escaped central tolerance and depends on immune checkpoint regulatory molecules and regulatory T cells (T_{reg}). These peripheral tolerance mechanisms may inhibit priming of naïve lymphocytes in secondary lymphoid organs (SLOs) or block activation of effector lymphocytes that are generated in the SLOs.[40]

Autoimmune Myocarditis Associated With Infection

Viruses are the most common class of microbes implicated worldwide in infections that precede autoimmune myocarditis. The identification of a virus or other microbes responsible for myocarditis is accomplished only in a minority of cases. Therefore, although infection in the heart is a likely antecedent cause of many cases of chronic autoimmune myocarditis or idiopathic dilated cardiomyopathy, the causal connection is rarely proven. A progression from acute viral myocarditis to an acute autoimmune phase progressing to chronic dilated cardiomyopathy has been well characterized in certain inbred strains of mice infected with Coxsackievirus (Fig. 69.9), and this has served as a model for understanding the pathogenesis of autoimmune myocarditis in humans. The distinction between active infectious myocarditis and antecedent autoimmune myocarditis in both experimental animals and humans is often blurred because circulating autoantibodies specific for various cardiac antigens, such as myosin heavy chains and beta-1 adrenergic receptors are often found concurrently with biopsy-derived evidence of microbial infection, but the pathogenicity of such antibodies in human disease has not been established. Even if there is persistent detection of viral nucleic acid the relative contribution of autoimmune versus antiviral responses to cardiac injury is unclear.

Rheumatic heart disease is a cardiac manifestation of acute rheumatic fever (ARF), which is a systemic autoimmune complication of group A streptococcal pharyngeal infections. ARF begins 2 to 3 weeks after infection and includes carditis, arthritis, chorea, and cutaneous lesions. The cardiac component is an acute pancarditis affecting the pericardium, myocardium, and endocardium. A subset of patients will suffer from scarring of valves (most commonly mitral and aortic), leading to regurgitation and/or stenosis and chronic heart failure. Acute rheumatic carditis is a T-cell and antibody-driven inflammatory reaction that apparently arises as a result of antigen-mimicry between streptococcal antigens and human cardiac antigen. Antibodies that recognize conformational epitopes of streptococcal M protein and N-acetyl-beta-D-glucosamine cross-react with human myosin epitopes, and rodent studies show that streptococcal M protein immunization can cause inflammation of myocardium and cardiac valves. There is also evidence for the presence of myosin-specific autoantibodies in acute rheumatic fever patients that cross-react with N-acetyl-beta-D-glucosamine and M protein. A pathognomonic feature of this condition is the presence of granuloma-like Aschoff bodies in the myocardium, consisting of T cells, activated macrophages, fibrin, and necrosis. There is no definitive identification of peptide epitopes shared by streptococcal and cardiac proteins.

Autoimmune Myocarditis Without Known Infection

Idiopathic lymphocytic myocarditis shares clinical and histopathological features of many cases of viral myocarditis and is usually associated with markers of autoimmunity directed against the heart—including circulating anti-alpha-myosin and anti-beta adrenergic receptor antibodies—without evidence of causal viral infection.[39] As is the case in viral myocarditis, the

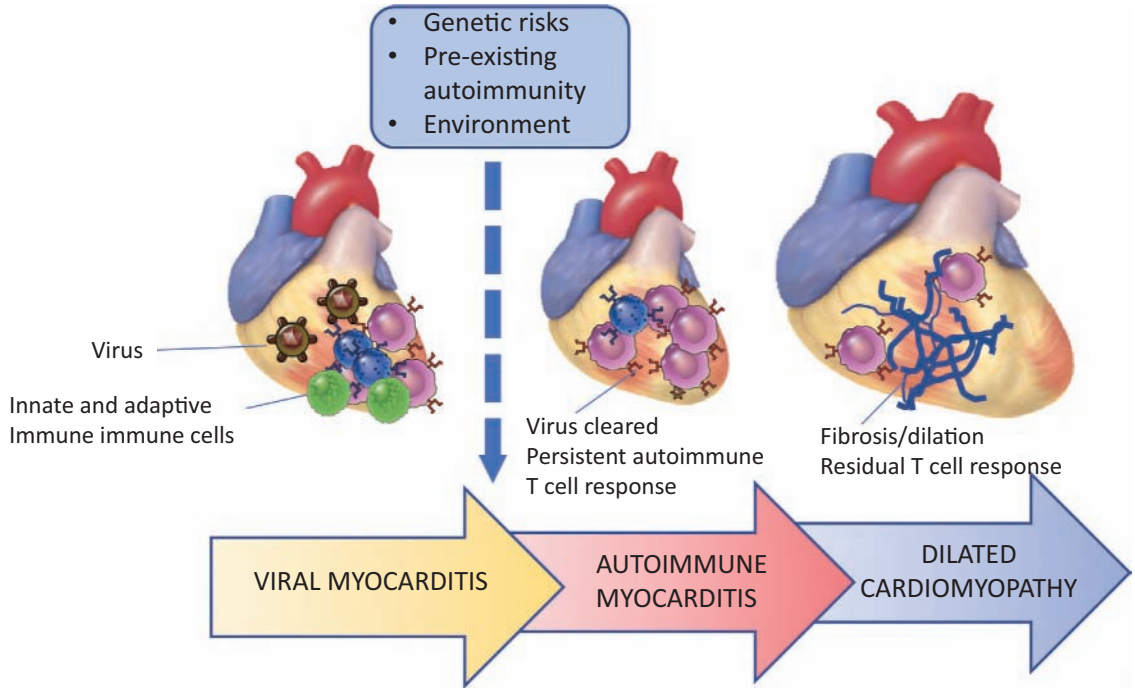


FIG. 69.9 Progression from Viral Myocarditis to Autoimmune Myocarditis and Dilated Cardiomyopathy. This scheme is based on experimental evidence from mice with coxsackievirus-induced myocarditis and experimental autoimmune myocarditis, as well as correlative human clinical data. Virus infection of myocardium evokes an innate immune response, which in turn promotes virus-specific T-cell responses, characterized by CD4 and CD8 effector T-cell infiltration accompanied by macrophages. A subset of patients, with as of yet poorly identified risk factors, go on to develop a T-cell driven autoimmune myocarditis, independent of persistent viral infection, likely induced by myocyte injury with release of self- antigens in a inflammatory milieu. The autoimmune phase in mouse models is largely a CD4 helper T-cell cytokine-driven process. Over time—weeks in mice, years in humans—progressive fibrosis of injured myocardium leads to heart failure with chamber dilation.

pathogenic role of the auto-antibodies is uncertain. Pathologic diagnosis is often made in late stages of the disease, usually in patients with a clinical diagnosis of dilated cardiomyopathy, rendering it difficult to rule out that viral infection initiated the disease.

Idiopathic giant cell myocarditis (GCM) is a rare disorder that most often presents with heart failure and is fatal or requires transplantation in about 90% of patients within 5 months of diagnosis. GCM is a diagnosed by histopathologic examination of endomyocardial biopsies, from myocardial cores removed for ventricular assist device (VAD) placement, or autopsy. Characteristic findings include diffuse infiltration by lymphocytes, macrophages, plasma cells and eosinophils, numerous multinucleated giant cells expressing macrophage markers, and foci of myocardial necrosis.[41] Giant cells also characterize cardiac sarcoidosis, where they usually localize within well-formed, non-necrotic granulomas. There is no evidence for microbial infection in GCM. Circumstantial evidence suggests autoimmunity or other forms of immune dysregulation are central to the pathogenesis GCM. About 20% of cases occur in patients diagnosed with other autoimmune/inflammatory diseases, and early treatment with calcineurin inhibitors slows disease progression and reduces mortality. Cardiac myosin immunization of rodents can induce myocarditis with giant cells. To date, no specific antigen has been identified that drives the human disease.

Eosinophilic myocarditis includes several different clinical entities in which eosinophils provoke cardiac inflammation.[42] The most common cause of myocardial eosinophilic inflammation appears to be hypersensitivity reaction to any one of a large number of drugs. These cases of hypersensitivity myocarditis rarely result in myocyte necrosis, but some patients present with heart failure. Acute necrotizing eosinophilic myocarditis is a rare and frequently lethal disorder that often presents with fulminant heart failure and demonstrates diffuse eosinophilic inflammation with necrosis. Thrombosis appears to be more frequent in eosinophilic myocarditis compared to viral myocarditis. This disorder may occur as an isolated cardiac disease or more commonly in patients with systemic eosinophilic autoimmune disorders, such as eosinophilic granulomatosis with polyangiitis (formerly Churg-Strauss syndrome). Other conditions in which eosinophilic inflammation is seen in the heart include Loeffler endomyocardial disease, parasite infections, and idiopathic hypereosinophilic syndrome.

Myocarditis Associated With Systemic Autoimmune Diseases

Myocarditis reportedly occurs more frequently in patients with systemic autoimmune diseases than the general population. Up to 10% of patients with systemic lupus erythematosus (SLE) will develop clinically diagnosed myocarditis, and autopsy studies indicate there is myocarditis in up to 50% of SLE patients. However, the limited information available from histopathologic examination in these cases suggests that the myocardial inflammation may be a response to ischemic damage caused by immune complex vasculitis rather than cardiac-antigen specific

autoimmunity. Clinical data suggest that patients with Sjögren syndrome, vasculitis, and polymyositis are also at higher risk for myocarditis, although in most of these cases histopathologic examination does not confirm the diagnosis. An example of autoimmune myocarditis associated with risk for another autoimmune disorder has been described in autoimmune (type 1) diabetes (T1D) patients who have a myocardial infarction. Most of these patients additionally suffer from a post-infarction syndrome that clinically resembles myocarditis. Almost 90% of these patients develop antibody and T-cell responses to cardiac alpha-myosin, while post-MI T2D patients do not.

Immune Checkpoint Blockade-Associated Myocarditis

Immune checkpoint blockade (ICB)-associated myocarditis is a recently emerged clinical entity which occurs in cancer patients treated with ICB drugs.[43,44] ICBs include monoclonal antibodies that block the function of CTLA-4, PD-1, or PD-L1 molecules that are required to regulate T-cell responses and prevent autoimmunity. More than 50% of ICB-treated cancer patients develop one or more immune-related adverse events (IRAEs), involving any one of many different organs and tissues. Patients with ICB-myocarditis present with acute onset heart failure or arrythmias, at times ranging from days to months after initiation of ICB therapy. ICB-associated myocarditis can occur in patients treated with only anti-CTLA, only anti-PD-1, or both antibodies. Histopathologic examination of tissue from autopsies and endomyocardial biopsies show CD4 and CD8 T-cell infiltration along with macrophages, and some myocardial necrosis. Although ICB myocarditis is relatively rare (~0.1% to 1% of ICB-treated patients) compared to other ICB-induced IRAEs, it appears to be one of the most lethal autoimmune T-cell-mediated processes. Preclinical studies showing the importance of CTLA-4 and PD-1/PD-L1 in protecting the heart from autoimmune myocarditis support this pathophysiologic scheme.[45] Some studies demonstrate oligoclonality of T-cell infiltrates, shared clones of T cells in the heart and tumor of individual patients and shared genes expressed by tumor and myocardium. However, the target cardiac antigens are not known. Skeletal myopathy can also occur in conjunction with ICB-associated myocarditis, as can a myasthenia gravis-like syndrome, indicating shared epitopes between these two types of striated muscle.

Cardiac Allograft Rejection

Heart transplantation is performed over 5000 times per year worldwide to treat end-stage heart failure due to many causes, including myocarditis. Certain forms of rejection of transplanted hearts involve immune effector mechanisms seen in myocarditis of native hearts, and transplantation medicine has to some degree informed the clinical approach to treating myocarditis. Other than testing for blood group compatibility and the presence of preformed allo-antigen-specific antibodies in the recipients' blood, both of which might cause hyperacute rejection and immediate graft failure, there are no other measures routinely taken to minimize allogeneic differences between the donor and recipient. Nonetheless, current immunosuppressive regimens allow for ~90% one-year patient survival and an average allograft half-life of ~10 years. The frequency of acute rejection episodes is highest early after transplantation, peaking at about 1 month and affecting from 10% to 30% of patients. Acute rejection is often diagnosed by routine serial EMBs of asymptomatic patients but may present with signs and symptoms of left ventricular dysfunction and heart failure. A standardized pathologic grading system for cardiac allograft rejection published by the International Society for Heart and Lung Transplantation[46] is widely used. Acute rejection is most often mainly T-cell-mediated (ACR), but antibody mediated (AMR) or a combination of both, also occur. The most likely target antigens are donor HLA proteins not shared by the host. ACR is characterized by an infiltrate of activated CD4 and CD8 T cells in the myocardium, with myocyte necrosis in severe cases. The appearance is similar to many cases of idiopathic autoimmune myocarditis. Myocytes and endothelial cells in the vicinity of lymphocyte infiltrates will often show evidence of activation by T-cell cytokines, including expression of HLA-DR and adhesion molecules. Often, significant left ventricular dysfunction may not be associated with significant myocyte damage on biopsy, perhaps reflecting disruptive effects of T-cell cytokines on the contractile function of myocytes, without overt cell necrosis. AMR is diagnosed by the presence of microvascular damage, thrombi, hemorrhage, neutrophilic inflammation, myocyte necrosis and the detection of Ig and complement fragments bound to capillary walls. These findings are consistent with allo-antibody-mediated complement activation in and damage to capillaries, leading to ischemic injury to myocytes. Acute rejection is prevented by chronic administration of various combinations of immunosuppressive drugs including calcineurin inhibitors, mTOR inhibitors, anti-metabolites, and corticosteroids. Treatment of an acute rejection episode usually involves changes in the dose or type of drugs, including the temporary use of T- or B-cell-depleting antibodies. For a discussion of allograft vasculopathy, please see Libby, Hasan, and Nohria in Bhatt Cardiovascular Therapeutics.[47]

CLINICAL IMPLICATIONS OF CARDIAC INFLAMMATION

The broad range of etiologies and clinical manifestations of myocardial inflammation pose unmet diagnostic and therapeutic challenges. Current diagnostic approaches short of myocardial biopsies do not reliably distinguish different immunopathological processes taking place in the heart, and pathological diagnoses of biopsies remain largely descriptive. Experimental forms of myocarditis in mice have provided robust evidence for the progression from viral infections in the heart to cardiac-specific autoimmunity to chronic injury and dilated cardiomyopathy, and mouse studies have shown that helper T-cell responses to defined epitopes of myocardial proteins can drive this progression. However, translation of the findings about T-cell subsets, cytokines, and autoantibodies involved in mice to treat effectively and specifically most forms of human myocarditis has lagged. There are no consistently effective treatments for acute lymphocytic myocarditis. Corticosteroid treatment is inconsistently effective in treating active lymphocytic myocarditis, including ICB-related myocarditis. Immunosuppressive drugs that target T-cell activation, such as calcineurin inhibitors, can delay progression of giant cell myocarditis, but such drugs have no proven benefit in more common forms of lymphocytic myocarditis. In the case of viral myocarditis, such treatments may impair anti-viral immunity and, thus, enhance cardiac injury due to virial cytopathic effects. For a fuller discussion of these therapies, please consult Libby, Hasan, and Nohria in Bhatt Cardiovascular Therapeutics.[47]

Progress in this clinical arena will require further characterization and understating of the immune responses in the human heart. This will involve application of new approaches for interrogating those responses, including identifying relevant immune-cell subsets by gene- and protein-expression technologies, determining the relevant specificities of lymphocytes that drive many of these disorders and clinical trials of specific immunomodulatory drugs.

REFERENCES

1. Hansson GK, Jonasson L. The discovery of cellular immunity in the atherosclerotic plaque. *Arterioscler Thromb Vasc Biol*. 2009;29(11):1714–1717.
2. Libby P. History of discovery: inflammation in atherosclerosis. *Arterioscler Thromb Vasc Biol*. 2012;32(9):2045–2051.
3. Saigusa R, Winkels H, Ley K. T cell subsets and functions in atherosclerosis. *Nat Rev Cardiology*. 2020
4. Libby P, et al. Atherosclerosis. *Nat Rev Dis Prim*. 2019;5(1).
5. Libby P, et al. Inflammation, Immunity, and Infection in Atherothrombosis: JACC Review Topic of the Week. *J Am Coll Cardiol*. 2018;72(17):2071–2081.
6. Noels H, Weber C, Koenen RR. Chemokines as Therapeutic Targets in Cardiovascular Disease. *Arterioscler Thromb Vasc Biol*. 2019 p. ATVBAHA118312037.
7. Libby P. The changing landscape of atherosclerosis. *Nature*. 2021;592(7855):524–533.
8. Lichtman AH, et al. Adaptive immunity in atherogenesis: new insights and therapeutic approaches. *J Clin Invest*. 2013;123(1):27–36.
9. Swirski FK, Hilgendorf I, Robbins CS. From proliferation to proliferation: monocyte lineage comes full circle. *Semin Immunopathol*. 2014;36(2):137–148.
10. Zernecke A, et al. Meta-Analysis of Leukocyte Diversity in Atherosclerotic Mouse Aortas. *Circulation Res*. 2020;127(3):402–426.
11. van Hinsbergh VWM, Eringa EC, Daemen MJAP. Neovascularization of the atherosclerotic plaque: interplay between atherosclerotic lesion, adventitia-derived microvessels and perivascular fat. *Curr OpLipidol*. 2015;26(5).
12. Sedding DG, et al. Vasa Vasorum Angiogenesis: Key Player in the Initiation and Progression of Atherosclerosis and Potential Target for the Treatment of Cardiovascular Disease. *Front Immunol*. 2018;9:706.
13. Libby P. Mechanisms of acute coronary syndromes and their implications for therapy. *N Engl J Med*. 2013;369(21):2004–2013.
14. Kolte D, Libby P, Jang I-K. New Insights Into Plaque Erosion as a Mechanism of Acute Coronary Syndromes. *JAMA*. 2021;325(11):1043.
15. Pasterkamp G, den Ruijter HM, Libby P. Temporal shifts in clinical presentation and underlying mechanisms of atherosclerotic disease. *Nat Rev Cardiol*. 2017;14(1):21–29.
16. Quillard T, et al. TLR2 and neutrophils potentiate endothelial stress, apoptosis and detachment: implications for superficial erosion. *Eur Heart J*. 2015;36(22):1394–1404.
17. Martinod K, Wagner DD. Thrombosis: tangled up in NETs. *Blood*. 2014;123(18):2768–2776.
18. Franck G, et al. Roles of PAD4 and NETosis in Experimental Atherosclerosis and Arterial Injury: Implications for Superficial Erosion. *Circ Res*. 2018;123(1):33–42.
19. Ridker PM, et al. Antiinflammatory Therapy with Canakinumab for Atherosclerotic Disease. *N Engl J Med*. 2017;377(12):1119–1131.
20. Tardif JC, et al. Efficacy and Safety of Low-Dose Colchicine after Myocardial Infarction. *N Engl J Med*. 2019;381(26):2497–2505.
21. Nidorf SM, et al. Colchicine in Patients with Chronic Coronary Disease. *N Engl J Med*. 2020(383):1838–1847.
22. Zhao TX, Mallat Z. Targeting the Immune System in Atherosclerosis: JACC State-of-the-Art Review. *J Am Coll Cardiol*. 2019;73(13):1691–1706.
23. Libby P. Inflammation in Atherosclerosis—No Longer a Theory. *Clin Chem*. 2021;67(1):131–142.
24. Jaiswal S, et al. Clonal Hematopoiesis and Risk of Atherosclerotic Cardiovascular Disease. *N Engl J Med*. 2017;377(2):111–121.
25. Bick AG, et al. Genetic Interleukin 6 Signaling Deficiency Attenuates Cardiovascular Risk in Clonal Hematopoiesis. *Circulation*. 2020;141(2):124–131.
26. Zhao TX, et al. Low-dose interleukin-2 in patients with stable ischaemic heart disease and acute coronary syndromes (LILACS): protocol and study rationale for a randomised, double-blind, placebo-controlled, phase I/II clinical trial. *BMJ Open*. 2018;8(9):e022452.
27. Nilsson J, Hansson Göran K. Vaccination Strategies and Immune Modulation of Atherosclerosis. *Circulation Res*. 2020;126(9):1281–1296.
28. de Couto G. Macrophages in cardiac repair: Environmental cues and therapeutic strategies. *Exp Mol Med*. 2019;51(12):1–10.
29. Lichtman AH. The heart of the matter: protection of the myocardium from T cells. *J Autoimmun*. 2013;45:90–96.
30. Blyszczuk P. Myocarditis in Humans and in Experimental Animal Models. *Front Cardiovasc Med*. 2019;6:64.
31. Leone O, et al. The spectrum of myocarditis: from pathology to the clinics. *Virchows Arch*. 2019;475(3):279–301.
32. Caforio A, et al. Immune-Mediated and Autoimmune Myocarditis: Clinical Presentation, Diagnosis and Management. *Heart Fail Rev*. 2012:18.
33. Maisch B. Cardio-Immunology of Myocarditis: Focus on Immune Mechanisms and Treatment Options. *Front Cardiovasc Med*. 2019;6:48.
34. Tschöpe C, et al. Myocarditis and inflammatory cardiomyopathy: current evidence and future directions. *Nat Rev Cardiology*. 2021;18(3):169–193.
35. Bottermann M, James LC. Intracellular Antiviral Immunity. *Adv Virus Res*. 2018;100:309–354.
36. Rose NR. Viral myocarditis. *Curr Opin Rheumatol*. 2016;28(4):383–389.
37. Lala A, et al. Prevalence and Impact of Myocardial Injury in Patients Hospitalized With COVID-19 Infection. *J Am Coll Cardiology*. 2020;76(5):533.
38. Perez-Molina JA, Molina I. Chagas disease. *Lancet*. 2018;391(10115):82–94.
39. Bracamonte-Baran W, Cihakova D. Cardiac Autoimmunity: Myocarditis. *Adv Exp Med Biol*. 2017;1003:187–221.
40. Grabie N, Lichtman AH, Padera R. T cell checkpoint regulators in the heart. *Cardiovasc Res*. 2019;115(5):869–877.
41. Blauwet LA, Cooper LT. Idiopathic giant cell myocarditis and cardiac sarcoidosis. *Heart Fail Rev*. 2013;18(6):733–746.
42. Brambatti M, et al. Eosinophilic Myocarditis: Characteristics, Treatment, and Outcomes. *J Am Coll Cardiol*. 2017;70(19):2363–2375.
43. Moslehi JJ, Brinkley BM, Meijers WC. Fulminant Myocarditis: Evolving Diagnosis, Evolving Biology, Evolving Prognosis. *J Am Coll Cardiol*. 2019;74(3):312–314.
44. Bonaca MP, et al. Myocarditis in the Setting of Cancer Therapeutics: Proposed Case Definitions for Emerging Clinical Syndromes in Cardio-Oncology. *Circulation*. 2019;140(2):80–91.
45. Salem JE, et al. Abatacept for Severe Immune Checkpoint Inhibitor-Associated Myocarditis. *N Engl J Med*. 2019;380(24):2377–2379.
46. Stewart S, et al. Revision of the 1990 working formulation for the standardization of nomenclature in the diagnosis of heart rejection. *J Heart Lung Transpl*. 2005;24(11):1710–1720.
47. Libby P, Hasan AA, Nohria A. Drugs Targeting Inflammation, in Opie's Cardiovascular Drugs: A Companion to Braunwald's Heart Disease. In: Bhatt DL, ed. Elsevier; 2020.

70

Autoimmune Thyroid Diseases

Laura C. Lane, Simon H.S. Pearce, and Anna L. Mitchell

GRAVES HYPERTHYROIDISM

KEY CONCEPTS

Classification of Autoimmune Thyroid Disease

Autoimmune hyperthyroidism	Graves disease
Autoimmune thyroiditis	Hashimoto thyroiditis
	Atrophic thyroiditis
	Postpartum thyroiditis

Graves disease (GD) is a common autoimmune condition that accounts for the majority of cases of hyperthyroidism in the developed world. Its pathogenesis is unique among the autoimmune endocrinopathies because a key feature is the presence of stimulating autoantibodies directed against the thyrotropin (thyroid-stimulating hormone [TSH]) receptor, which mimic the action of TSH, a native hormone produced in the pituitary, to drive thyroid overactivity. Interestingly, thyroid dysfunction is commonly associated with other extrathyroidal manifestations of GD, the most common being Graves ophthalmopathy (GO).

Epidemiology

GD is one of the most common autoimmune diseases, with a prevalence of approximately 1% in women in the developed world.[1] It is more common in iodine-sufficient countries, where it accounts for 60% to 90% of cases of hyperthyroidism.[2] GD is seven times more common in women than in men and may affect individuals at any age; however, the peak incidence occurs between the ages of 35 and 40 years.

Etiology

GD is a complex genetic condition, implying that environmental stimuli precipitate disease in genetically predisposed individuals who harbor multiple susceptibility alleles. A large Danish twin study estimated that approximately 80% of the propensity to develop GD is attributable to genetic factors.[3] Further evidence for the heritability of GD comes from the observation that it clusters within families. Up to one-quarter of individuals with GD have a first-degree relative with the condition or with another autoimmune thyroid disease, such as autoimmune hypothyroidism (AH).[4] Should an individual have a sibling with GD, it is estimated that the relative risk (λ_s) of that individual developing GD is approximately 10-fold that of the background population, which is comparable with that of other heritable autoimmune conditions, such as type 1 diabetes, which has a λ_s of 15.

A number of genetic loci have been shown to contribute to GD susceptibility (Fig. 70.1). These genes encode proteins in biologic pathways that regulate immune system activity or thyroid biology.[4,5] The major histocompatibility complex (MHC) region on chromosome 6p21 is associated with multiple autoimmune conditions. Human leukocyte antigen (HLA) genes found within the MHC region play a vital role in pathogen and self-peptide recognition and therefore have a clear role in immunity and in establishing and maintaining immune tolerance (see Chapters 10 and 90).

In European populations, the primary association between MHC and GD is with alleles of the class II MHC genes. The *HLA-DR3* allele is detected twice as frequently in subjects with GD as in healthy controls (i.e., 50% of GD subjects vs. 25% of controls). At the protein level, neutral amino acids alanine or glutamine are substituted for positively charged arginine at position 74 in the HLA-DR peptide-binding pocket, which is thought to alter the binding-pocket configuration, more readily allowing self-peptides to enter the antigen binding site.[5] Importantly, 50% of individuals with GD do not have the *HLA-DR3* allele, implying that there is unlikely to be a single antigenic epitope responsible for GD.

The cytotoxic T-lymphocyte antigen-4 *(CTLA4)* gene (chromosome 2q33) encodes a costimulatory molecule expressed on the surface of activated T cells, which plays a pivotal role in downregulating T-cell responses and in checking T-cell activation, emphasizing the contribution of inhibitory signals in setting immune response thresholds (see Chapter 10). The *CT60* single nucleotide polymorphism (SNP) downstream from the 3′ untranslated region (3′UTR) was found to influence GD susceptibility (odds ratio [OR] 1.5) and has been suggested as a possible etiologic variant; however, contributions from other variants, such as at codon 17 in the CTLA-4 signal peptide remain likely.[6] Approximately 50% of individuals from healthy European populations carry the autoimmune "susceptible" *CTLA4* haplotype; therefore other important factors are clearly at play. CTLA-4 polymorphisms also contribute to susceptibility to type 1 diabetes, autoimmune adrenal insufficiency, celiac disease, and several other autoimmune conditions.

Products of several other genes involved in immunoregulation have allelic variants associated with GD, including CD25, CD40, PTPN22, PD-L1, IFIH1, BACH2, SCGB3A2, and FCRL3. In addition to these variants in immune regulatory pathways, loci specific to GD have been identified on the basis of known thyroid pathophysiology. The gene encoding the TSH receptor (TSH-R) on chromosome 14q31 is an obvious candidate for GD, as the TSH-R is directly stimulated by autoantibodies in affected individuals. Although initial studies reported conflicting results, a definite association between a number of SNPs in intron 1 of the *TSHR* gene and GD has now been confirmed in Whites.[5] The mechanism by which these intronic SNPs confer disease susceptibility is unknown.

Despite years of research into the genetic etiology of GD, 30% to 40% of the inherited susceptibility has yet to be accounted for. This "hidden" heredity is a common theme in genetically complex traits, but it is likely to be owing to rare genomic variants, polymorphisms in regulatory DNA

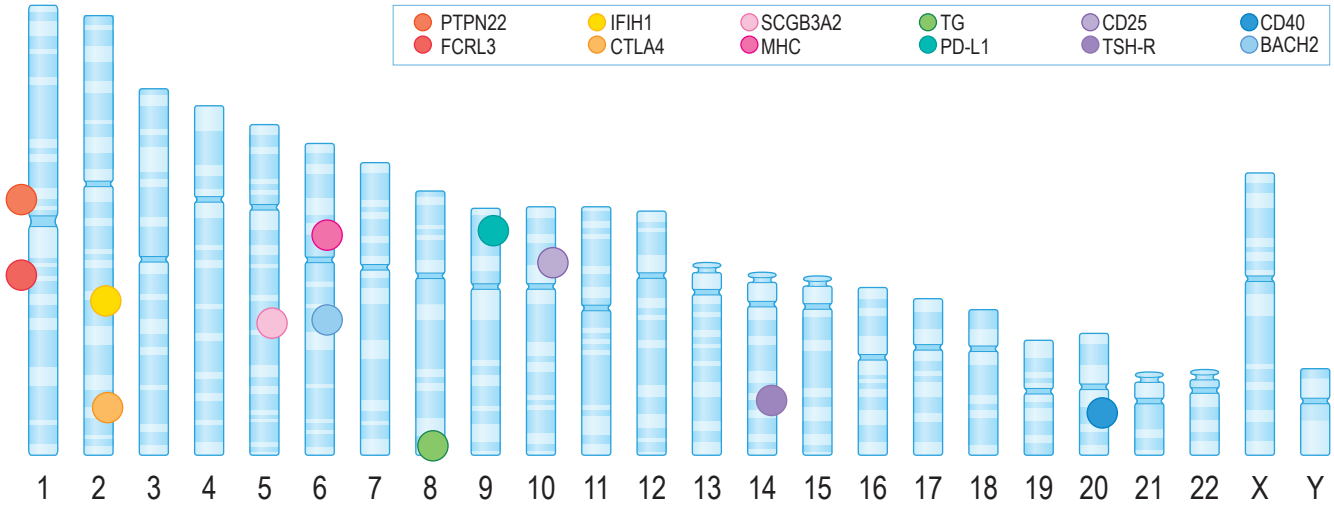


FIG. 70.1 Schematic Diagram to Illustrate the Loci that have been Associated with Graves Disease to Date. Each locus is shown on its respective chromosome, with chromosome 1 depicted on the left and chromosome Y on the right. *MHC*, Major histocompatibility complex; *PD-L1*, programmed death 1/ligand 1; *TSH-R*, thyroid-stimulating hormone receptor.

sequences, including noncoding or microRNAs (miRNAs), and/or epigenetic factors.

GD is one of the few autoimmune conditions for which links to environmental factors have been definitively established.[7] Iodine is one of the most common precipitants of thyroid dysfunction. With regard to GD, more cases are observed in iodine-sufficient areas. A study of individuals in Iceland, where iodine intake is high, and age-matched individuals in East Jutland, Denmark, where iodine intake is low, showed that the incidence of GD was more than double in the higher-iodine environment.[2]

Cigarette smoking also influences GD susceptibility and severity, in particular of GO. Meta-analysis of eight studies showed the OR for developing GD was 3.3 for current smokers compared with that for lifelong nonsmokers.[8] The same study also revealed that current smokers are more likely to develop GO compared with nonsmokers (OR 4.4). Cigarette smoke extract has been shown in vitro to stimulate adipogenesis and glycosaminoglycans (GAGs) production by orbital fibroblasts (OFs), which accumulate in the orbital tissues in GO.[9]

Stress appears to influence both GD disease onset and clinical course. Individuals with GD retrospectively reported more negative life events in the preceding year compared with matched controls.[10] In a population-based study, individuals with clinically diagnosed stress disorders were found to have an increased risk of many different forms of autoimmunity, including autoimmune thyroid diseases.[11]

Changes in immune system function appear to influence the onset of GD. During pregnancy, which is a relatively immunosuppressed state, hyperthyroid GD is often mild and can be managed with smaller doses of antithyroid drugs. In some cases, it remits entirely, allowing the individual to stop medication in the short term. However, in the postpartum period, when the immune system normalizes, there is typically worsening or relapse of GD. A similar phenomenon is seen in individuals who have been significantly immunosuppressed and then have recovered. For example, new-onset GD has been reported in people who have been successfully treated with highly active antiretroviral therapy (HAART) for human immunodeficiency virus (HIV) infection. During treatment, as the immune system recovers, immune activation increases. T cells are exposed to thyroid antigens, resulting in an autoimmune response at the time of immune reconstitution.[12] A similar phenomenon has been seen in individuals with multiple sclerosis (MS) treated with alemtuzumab (ALZ), the lymphocyte-depleting anti-CD52 antibody. An alternative explanation is the weakening of physiologic antiinflammatory pathways, which unleashes the immune system. As more novel biologic agents become available, particularly for cancer treatment, this phenomenon is likely to become more common.

KEY CONCEPTS

Environmental Factors Known to Influence Graves Disease Susceptibility

- Smoking
- Iodine
- Stress
- Immune system reconstitution states
  - Postpartum state
  - Successful treatment of human immunodeficiency virus (HIV) with highly active antiretroviral therapy (HAART)
  - T-cell depletion therapy (e.g., alemtuzumab)
- Infections (e.g., hepatitis C)

Immunopathogenesis

Histologically, the thyroid in GD is characterized by a diffuse lymphocytic infiltrate, consisting of both T and B cells, associated with thyrocyte hyperplasia (Fig. 70.2). Although T cells play a major role in inflammatory cell recruitment, cytokine secretion, antigen recognition, and thyrocyte damage, infiltrating B cells also produce antibodies, including those that drive hyperthyroidism. The major autoantigens in GD are the TSH-R, the thyroid peroxidase (TPO) enzyme, and thyroglobulin (Tg). More than 95% of patients with GD have detectable circulating TSH-R autoantibodies (TRAbs),[13] which are necessary for hyperthyroidism, and approximately 90% have detectable TPO autoantibodies.[14] Antibodies directed against the sodium iodide symporter and the apical iodide transporter, pendrin, have been reported in smaller numbers of patients.[14]

GD is unique among autoimmune conditions in that the TRAbs directly stimulate thyroid gland activity. This is

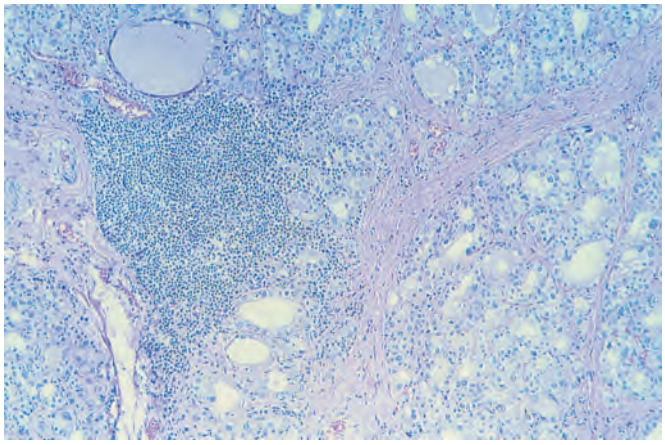


FIG. 70.2 Diffuse lymphocyte Infiltrate and Thyrocyte Hyperplasia in a Patient with Graves Disease.

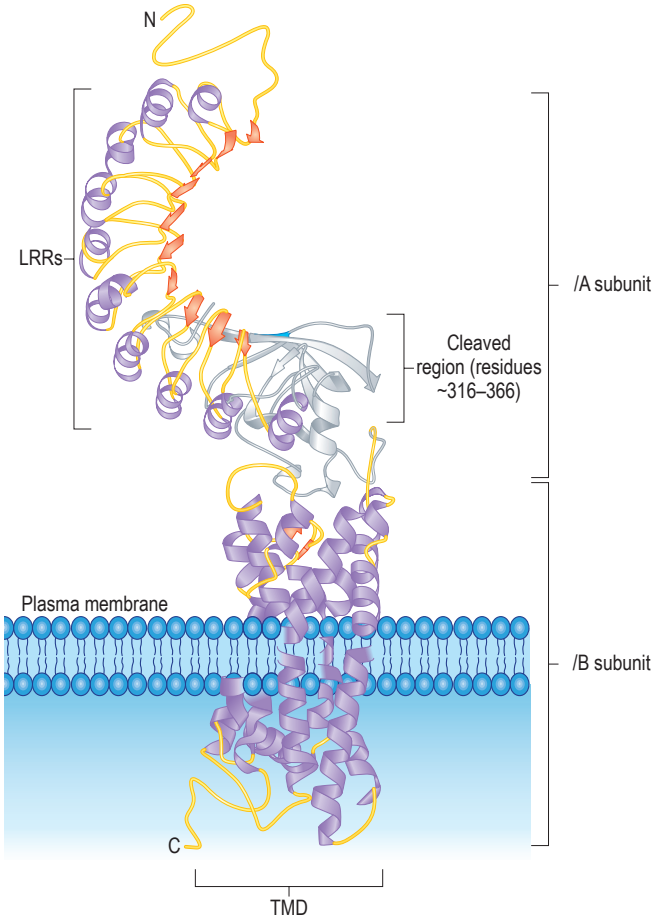


FIG. 70.3 Structure of the Thyroid-Stimulating Hormone Receptor. *LRRs*, Leucine-rich repeats; *TMD*, transmembrane domain. (Courtesy R. Latif; adapted from Davies TF, Ando T, Lin RY, et al. Thyrotropin receptor-associated diseases: from adenomata to Graves' disease. *J Clin Invest*. 2005;115:1972–183.)

exemplified by neonatal GD, where maternal TRAbs cross the placenta, resulting in transient hyperthyroidism in the newborn. However, although TRAbs are sufficient to result in transient hyperthyroidism in these infants, these autoantibodies are not sufficient per se to result in the persistent thyroid autoreactivity of true GD. TRAbs are classically immunoglobulin G1 (IgG1) subclass and target an epitope in the amino terminal region of the leucine-rich repeat motif in the extracellular domain of the TSH-R (Fig. 70.3).[15] When the autoantibody binds to the TSH-R, this activates intracellular G proteins, which, in turn, induce transcription of genes, such as *TPO* and *TG*, via the cyclic adenosine monophosphate (c-AMP) and phospholipase C pathways. This results in thyrocyte hyperplasia and increased thyroid hormone synthesis. The TSH-R antibody–induced expression of TPO and Tg, which are also thyroid antigens, may be a mechanism for disease perpetuation. TRAbs can also be "blocking" in nature and prevent receptor activation, causing hypothyroidism. These two types of autoantibodies can also coexist, resulting in fluctuating thyroid function. Thus, although it is common to equate GD with hyperthyroidism, individuals with GD may occasionally be hypothyroid or euthyroid (presenting with GO).

TPO antibodies can be of IgG subclass 1, 2, or 4 and typically circulate in concentrations 1000-fold higher than those of TRAbs. They are directed against two structurally complex regions of the TPO molecule, the epitopes involving residues from both the myeloperoxidase-like and the complement control protein-like domains. TPO antibodies may have pathogenic significance in that they can fix complement and target the thyrocyte for cell-mediated cytotoxicity. In contrast to TRAbs, they do not appear to either stimulate or block the enzymatic activity of TPO.

Clinical Presentation

Hyperthyroid GD can present with manifestations affecting almost any organ system in the body, and, as with many endocrine conditions, affected individuals may report a gradual onset of nonspecific symptoms, typically over a period of months. This often leads to a delay in seeking medical attention and in the initial diagnosis being made. The signs and symptoms of GD can be divided into those associated with hyperthyroidism in general and those specific to GD. These are summarized in Table 70.1.

CLINICAL PEARLS

Clinical Signs Specific to Graves Disease

- Graves orbitopathy
- Thyroid bruit
- Thyroid acropachy
- Pretibial myxedema

Investigation and Diagnosis of Graves Disease

The diagnosis of GD is a clinical one, supported by laboratory investigations. Imaging is occasionally required if the diagnosis is in doubt. Thyrotoxicosis is diagnosed biochemically on the basis of an elevated serum free triiodothyronine (fT_3) or free thyroxine (fT_4) in the presence of a completely suppressed TSH. TRAbs are highly sensitive for GD, and assays are now widely available. TRAbs should be measured in all thyrotoxic patients in order to reach a positive diagnosis. They are clinically useful, even if a patient has extrathyroidal signs of GD, such as GO or pretibial myxedema, because they can be monitored to ascertain response to treatment. TPO antibodies are also often measured as a surrogate for thyroid autoimmunity.

Imaging is reserved for individuals in whom the diagnosis is not clear (e.g., hyperthyroid patients with negative TRAb). Radionuclide scanning (e.g., ^{99}Tc or ^{123}I) is favored over ultrasonography, because the former gives functional information on the activity of the thyroid gland, although Doppler flow studies

TABLE 70.1 Common and Rarer Clinical Manifestations of Hyperthyroidism

Common	Rare
Neuropsychiatric	
Anxiety	Chorea
Fatigue and exhaustion	Collapse (periodic paralysis)
Fine tremor	Pseudobulbar palsy
Restlessness and fidgeting	Spasticity
Gastrointestinal	
Increased appetite	Hepatosplenomegaly
Loose stools	
Increased frequency of defecation	
Nausea	
Weight loss	
Cardiorespiratory	
Palpitations	Congestive cardiac failure
Shortness of breath on exertion	
Tachycardia (sinus, atrial fibrillation)	
Peripheral vasodilation, flushing	
Systolic hypertension	
Genitourinary	
Menstrual irregularities	
Cutaneous	
Itch	Thyroid acropachy
Heat intolerance	Pretibial myxedema
Hair loss	Onycholysis
Musculoskeletal	
Hyperreflexia	
Proximal muscle weakness	
Ophthalmic	
Lid lag	Optic neuropathy
Lid retraction	
Exophthalmos and proptosis	
Eye dryness	
Chemosis	
Ophthalmoplegia	
Miscellaneous	
Thirst	
Thyroid bruit	

are making ultrasound evaluation of the thyroid increasingly informative (Fig. 70.4). In GD, there is diffuse uptake in the thyroid gland on radionuclide scanning.

Management of Graves Disease

The management of hyperthyroid GD can be divided into three broad categories: medical management, radioiodine (radioactive iodine [RAI]) treatment, and surgery.[16]

Medical Management—Antithyroid Drugs

The thionamide drugs (carbimazole, its metabolite methimazole, and propylthiouracil) compete with Tg to act as substrates for iodination by TPO. Once iodinated, they are metabolized peripherally, depleting thyroid iodine stores. When the thyroid iodine stores are depleted and the intrathyroidal thionamide concentration is high enough, thyroid hormone synthesis is abrogated. Most individuals become euthyroid following 4 to 8 weeks of treatment; however, euthyroidism may take longer to achieve in those with poor medication compliance or with a history of recent iodide exposure. Following initial treatment, thionamides may either be administered as a fixed high dose, with levothyroxine supplementation to prevent hypothyroidism (known as a "block and replace" regimen), or at progressively lower doses, titrated to allow adequate thyroid hormone generation. Following 6 to 18 months of thionamide treatment, approximately 50% of patients will remain in remission following cessation of therapy. There is no improvement in remission rate in individuals with GD who are treated for longer than 18 months. Lower TRAb levels at the time of medication cessation are associated with higher chances of remission.[17] The mechanism of the thionamide-induced remission in GD remains obscure; however, the lymphocytic infiltrate in the GD thyroid is rapidly abolished by thionamide treatment, and serum TRAb and TPO autoantibodies also decrease during treatment, suggesting an immunomodulatory effect. It is telling that several induced murine models of thyroiditis can be ameliorated by thionamide treatment, suggesting an immunomodulatory role that is distinct from the antithyroid hormone synthesis action. If the patient's disease relapses following medical treatment, definitive treatment should be considered, as a second course of thionamide treatment rarely induces prolonged remission.

FIG. 70.4 ^{99}Tc Pertechnetate Radionuclide Scan Image from an Individual with Graves Disease Showing Diffuse Uptake Throughout the Thyroid Gland.

Definitive Treatment

RAI therapy (^{131}I) and surgical thyroidectomy are both efficacious treatments for GD, effectively removing functional thyroid tissue and rendering the patient hypothyroid and dependent on levothyroxine replacement over the long term. RAI therapy takes advantage of the thyroid's ability to concentrate iodine. RAI is administered orally and concentrates in the thyroid, primarily emitting beta-radiation to produce local thyrocyte DNA damage. These cells then undergo necrotic change, and the thyroid atrophies over the subsequent year. RAI renders more than 80% of those with GD hypothyroid and needing long-term

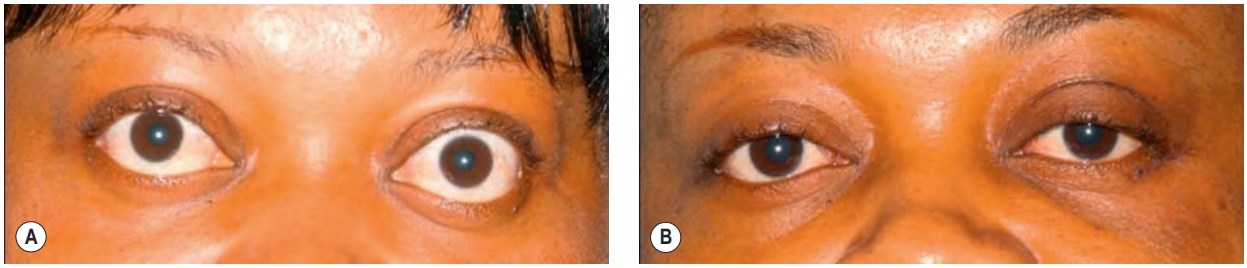


FIG. 70.5 (A) Clinical photograph of the eyes of an individual with Graves ophthalmopathy prior to treatment. (B) Clinical photograph of the same patient's eyes following surgical orbital decompression and rehabilitative surgery.

thyroid hormone replacement. RAI is generally well tolerated, and long-term follow-up studies in adults have shown reassuring results with regard to carcinogenicity. RAI is absolutely contraindicated in pregnancy and should be avoided in those with active, inflammatory GO. Total or near-total thyroidectomy is also an effective treatment for GD and is particularly suitable for individuals with a large goiter, those with severe hyperthyroidism who cannot tolerate thionamides, or in those with active GO when RAI is relatively contraindicated. The complications of thyroidectomy include change in voice because of intraoperative damage to the recurrent laryngeal nerve and hypocalcemia (often transient) caused by parathyroid gland damage.

GRAVES OPHTHALMOPATHY

Epidemiology

GO is the most common extrathyroidal manifestation of GD. It is clinically apparent in 25% to 50% of those presenting with GD, although almost all patients have radiologic changes consistent with GO to some degree. The peak age of incidence is in the fifth and seventh decades of life. GO precedes thyroid gland dysfunction in approximately 20% of patients, arises at the same time in 40%, and occurs after diagnosis of thyroid dysfunction in the remaining 40%. Males, older adults, and smokers are more likely to have severe disease. Ninety-three percent of cases of GO occur in those with hyperthyroid GD. However, 3% and 4% of GO occurs in hypothyroid and euthyroid patients, respectively. A proportion of these patients with "euthyroid Graves" will eventually become hyperthyroid; however, some will remain hypothyroid or euthyroid despite the presence of TRAbs.

Etiology

GO shares many etiologic factors with GD. Smoking is the major environmental risk factor, with smokers or ex-smokers four times more likely to develop GO compared with lifelong nonsmokers. The number of cigarettes smoked per day correlates with the risk of developing GO, suggesting a "dose-dependent" effect. RAI therapy is also known to occasionally cause flare-ups of GO. This is thought to be largely related to uncontrolled post-RAI hypothyroidism. Smokers and those with active GO seem particularly susceptible to this complication, and therefore RAI is relatively contraindicated in patients with active GO.[18]

Immunopathogenesis

The molecular mechanism that links thyroid dysfunction with GO remains incompletely understood. The TSH-R is widely believed to be the primary autoantigen linking the thyroid and the orbit. OFs have been shown to express cell-surface TSH-R, and TSH stimulation *in vitro* results in an increase in intracellular cAMP. Mechanistically, the orbital changes that occur in GO are better understood. TRAbs that bind to the TSH-R on OFs (or possibly other OF receptors, such as the insulin-like growth factor receptor; IGF-1R) are believed to activate OFs to secrete cytokines and chemokines, which attract lymphocytes and other inflammatory cells. These infiltrate the orbital tissues, augmenting further the proinflammatory cytokine environment, causing OFs to proliferate and to secrete excessive GAGs.[19] GAGs accumulate in the extraocular muscles, increasing their size. These matrix molecules are also osmotically active, resulting in edema and swelling of surrounding tissues.

In addition, the adaptive arm of the immune system is thought to interact via HLA and CD40 molecules expressed by the OFs to increase autoantigen presentation, thus perpetuating the cycle. OFs are also thought to differentiate into adipocytes, resulting in increased retroorbital fat deposition. The resulting inflammation gives rise to redness and oedema of the orbital tissues, with the cellular proliferation and GAG accumulation producing proptosis (protrusion of the eyeball from the socket), increased intraorbital pressure, and restriction of eye movements.[19]

Diagnosis and Clinical Presentation

When the clinical signs of GO are associated with thyroid dysfunction and circulating TRAb, the diagnosis is generally straightforward. If proptosis is completely unilateral or GO features occur without upper lid retraction, then the diagnosis needs to be confirmed by imaging, because the differential diagnosis for unilateral proptosis is a space-occupying retroorbital lesion. Patients with GO may complain of gritty, watery, or uncomfortable eyes, with or without a change in appearance (upper eyelid retraction, soft tissue swelling, redness of the eyes, and proptosis) (Fig. 70.5). Diplopia occurs if eye movements are restricted by stiffness of the extraocular muscles or high intraorbital pressure. Deteriorating visual acuity and color desaturation are sinister symptoms in GO, indicating incipient optic neuropathy.

GO has a predictable and generally monophasic natural history (Fig. 70.6). There is an early phase of increasing disease activity that can be targeted by medical therapy, and it is followed by a plateau phase and then gradual improvement until a stable, inactive phase is reached. GO manifestations can be classified in two ways: on the basis of severity, which indicates the extent of functional, anatomic, and cosmetic features; and on the basis of activity, which denotes the intensity of any acute inflammatory reaction. Severity of GO is assessed using the "NO SPECS" classification system (Table 70.2, A), and activity is assessed using the Clinical Activity Score (CAS) (see Table 70.2, B). A CAS

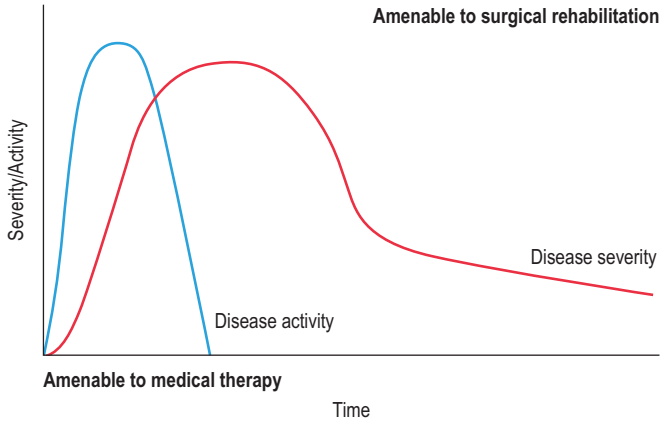


FIG. 70.6 Rundle Curve Illustrating the Natural History of Graves Ophthalmopathy.

of 3 or more indicates active GO.[18] These classifications allow determination of which patients require treatment and which therapy is most appropriate.

Treatment

Patients with GD (with and without GO) should be strongly encouraged to stop smoking and offered smoking cessation support. The general health benefits of smoking cessation are numerous; in addition, smokers are more likely to relapse after a course of medical therapy compared with nonsmokers.

In parallel to restoring and maintaining euthyroidism, the successful treatment of GO depends on staging the activity and severity of the disease. In patients with mild, active GO, an observational policy can be used with symptomatic measures, such as artificial tears and dark glasses. Selenium supplements, at a dose of 100 µg twice daily for 6 months, have been shown to significantly improve quality of life, slow disease progression, and reduce ocular involvement in these patients.[20] All patients with GO, with the exception of those with very mild disease (which is often transient) should be assessed in a joint multidisciplinary thyroid eye clinic, comprising ophthalmologists, oculoplastic surgeons, and endocrinologists. In those with moderate to severe, active, and progressive GO, a course of oral or intravenous (IV) glucocorticoids (steroids) is indicated. Orbital radiotherapy is also efficacious in people with active inflammation and diplopia, but because it has a delayed onset of action, it often needs to be used in conjunction with other therapies, such as steroids or orbital decompression surgery. In patients with sight-threatening optic neuropathy, high-dose IV steroids are used, and orbital pressure is relieved by urgent orbital decompression surgery. In patients whose eyelids do not close completely, eye ointments and protective eye pads are essential to protect the eyes against corneal damage and ulceration. Once disease activity has burned out, rehabilitative surgery can greatly improve the function and cosmetic appearance of the eyes. Orbital decompression, strabismus correction, and eyelid surgery are commonly used procedures.[21]

FUTURE DEVELOPMENTS FOR GRAVES HYPERTHYROIDISM AND OPHTHALMOPATHY

Conventional treatment of GD with surgery, RAI, or antithyroid drugs has not substantially changed over the past 50 years.

TABLE 70.2 Assessment of Graves Ophthalmopathy

A: The "NO SPECS" Classification System to Assess the Severity of GO

- Class 0—No signs or symptoms
- Class 1—Only signs (limited to upper lid retraction and stare, with or without lid lag)
- Class 2—Soft tissue involvement (edema of conjunctivae and lids, conjunctival injection, etc.)
- Class 3—Proptosis
- Class 4—Extraocular muscle involvement (usually with diplopia)
- Class 5—Corneal involvement (primarily due to lagophthalmos, the inability to completely close the eyelids)
- Class 6—Sight loss (caused by optic nerve involvement)

B: The Clinical Activity Score (CAS) to Assess Activity of GO

A single point is scored for each of the features present. Each feature has equal weighting. A higher score indicates more active disease.

- Spontaneous orbital pain
- Gaze-evoked orbital pain
- Eyelid swelling that is attributed to active GO
- Eyelid erythema
- Conjunctival redness that is considered to be due to active GO
- Chemosis
- Inflammation of caruncle or plica
- Increase of >2 mm in proptosis
- Decrease of >8 degrees in uniocular ocular excursion in any one direction
- Decrease of acuity equivalent to 1 Snellen line

GO, Graves ophthalmopathy.

There remain significant unmet needs with these strategies, resulting in the demand for new therapeutic options. Novel approaches including biologic, small-molecule, and peptide immunomodulation are currently at various stages of development. Furthermore, in light of the significant side effects associated with steroid treatment and low patient satisfaction with current therapies, steroid-sparing therapeutic agents are of particular interest for GO.

The novel immunomodulatory therapies currently being investigated, some of which have shown potential efficacy in preclinical or phase II studies, include targeting various aspects of B-cell activity, such as blocking of the costimulatory CD40 interactions, the B cell–activating factor (BAFF), and antibody recycling. However, the most widely studied B-cell therapy is rituximab, a CD20 monoclonal antibody (mAb) that depletes circulating B cells and appeared to be potentially efficacious in early studies. However, in two randomized controlled trials (RCTs) in individuals with moderate to severe active GO, results have been conflicting, and therefore further studies are needed. Other agents investigated in GO include the novel immunotherapeutics, tocilizumab (TCZ) and teprotumumab, both of which have demonstrated positive findings in GO. The anti–interleukin-6 (IL-6) receptor TCZ is a recombinant humanized IgG1 mAb that inhibits IL-6 binding to the IL-6 receptor, inhibiting the proinflammatory effects of IL-6. An RCT, which included 32 patients with moderate-to-severe corticosteroid-resistant GO, reported significant improvements in GO severity and activity (CAS <3 achieved in 86.7% vs. 35.2% in placebo). In addition, an RCT has demonstrated that targeting the orbital IGF1 receptor with teprotumumab results in a significant improvement in inflammatory eye signs, proptosis, and patient quality of life in those with GO.

TSH-R–specific immunotherapeutics including TSH-R–blocking antibodies, TSH-R–specific peptide immunotherapy, and small-molecule TSH-R antagonists are in various stages of development from preclinical to phase I trials. The key advantage of TSH-R–specific approaches is that it provides a direct, targeted approach that theoretically avoids disruption of immune system function. Novel agents have been evaluated to a limited extent and are subject to further ongoing studies.

AUTOIMMUNE HYPOTHYROIDISM

The most common cause of AH is chronic (or lymphocytic) autoimmune thyroiditis. There are two variants, atrophic and goitrous (Hashimoto thyroiditis).

Epidemiology

In populations living in iodine-sufficient areas, AH is common, affecting 1% to 10% of the population. The prevalence increases with age, with 3% to 20% of individuals older than 75 years being hypothyroid. Like GD, AH is more common in women than in men. In a UK community survey, the incidence of hypothyroidism in women was 3.5/1000/year, which increased to 13.7/1000/year in women between 75 and 80 years of age. In men, the incidence was just 0.6/1000/year.[22]

Etiology

AH, like GD, is a complex genetic condition. Familial clustering provides evidence for a genetic etiology, which, in several studies, appears stronger than that for GD. The λ_s for AH is estimated to be between 10 and 45, suggesting that AH is more heritable compared with GD. In families with autoimmunity, frequently a mixture of individuals affected by AH and GD are seen, suggesting some shared genetic factors. The differing prevalence of AH in different ethnic groups, with AH being more common in White than in Black populations, also supports a genetic background. Knowledge about the genetics of AH is limited.

The MHC class II *HLA-DR3, HLA-DR4*, and *HLA-DR5* alleles have been associated with AH in Whites only. Conflicting results have been reported for *HLA-DQ* alleles. One study reported that the *HLA-DQ* alleles *DQA1*0301* and *DQB1*0201* confer susceptibility to AH in Whites, with certain *HLA-DQ* alleles (*DQA1*0102* and *DQB1*0602*) reported to confer a protective effect.

In common with GD, the *CTLA-4* gene also appears to influence AH susceptibility. Three *CTLA-4* polymorphisms have been associated with AH in several populations. An A/G SNP located downstream of the 3′UTR (designated CT60), an A/G polymorphism at codon 17, and a 106-bp microsatellite repeat in the 3′UTR of exon 3. A locus on chromosome 8q24 containing the Tg gene was linked to AH, and several SNPs were subsequently studied in AH individuals with modest reported ORs for association of between 1.32 and 1.56. Other loci implicated in AH susceptibility include the tumor necrosis factor *(TNF)-α* gene, *PTPN22, CYP27B1*, T-cell receptor (TCR) genes, and several immunoglobulin genes and cytokine regulatory genes.

In contrast to GD, environmental factors in AH susceptibility have been challenging to identify. However, the role of iodine is widely accepted because population studies have reported an increase in the prevalence of thyroid lymphocytic infiltration and autoantibodies following public health salt iodization programs. Infectious agents have also been implicated in susceptibility to AH. Several studies have identified an increased prevalence of IgG and/or IgA antibodies to virulence-associated outer membrane proteins of *Yersinia enterocolitica* in AH patients and in relatives of individuals with AH, suggesting that susceptibility genes for *Yersinia* infection may also confer risk for AH.

The effect of radiation, either "internal" (nuclear "fallout" or from RAI treatment) or "external" (radiotherapy or direct exposure during a nuclear accident), on AH susceptibility has been extensively studied. Following the nuclear reactor accident at Chernobyl, a rise in thyroid autoantibodies was noted 15 years following exposure; however, this was not accompanied by thyroid dysfunction. Long-term follow-up studies of thyroid function in Japanese survivors of the atomic bombings of Nagasaki and Hiroshima have demonstrated a clear link between radiation exposure and thyroid cancer; however, the association with AH remains disputed. One study, at 40-year follow-up, demonstrated a significant relationship between dose of radiation exposure at Nagasaki and AH. However, a further study, at more than 50 years of follow-up, showed that radiation exposure did not correlate with either the occurrence of thyroid autoantibodies or AH.[23]

Immunopathogenesis

The mechanisms by which tolerance to thyroid antigens is lost in the first instance remain obscure. It appears that both a susceptible genetic background and a permissive environment are required before AH develops. Notably, AH is much more frequent in the autoimmune polyendocrinopathy type 1 (APECED) syndrome than GD, suggesting that central thymic T-cell selection, and therefore central tolerance, may be more important in AH than in GD. Histologically, lymphocytic infiltrates can be seen in the thyroid, consisting of both T and B cells (Fig. 70.7). These infiltrates can be diffuse or focal. Scarring and fibrosis may also be seen, with destruction of the normal thyroid architecture and an absence of colloid in thyroid follicles. An IgG4-positive histologic variant of Hashimoto thyroiditis has been proposed, although neither diagnostic criteria nor clinical significance have been clearly established.[24]

Both cell-mediated and humoral immune mechanisms are important in the continuing thyroid damage seen in AH. T cells play a pivotal role in both the initiation and perpetuation of AH. Studies in which researchers induced hypothyroidism in Rag1-deficient transgenic mice that were unable to produce autoantibodies confirm this.[25] T cells respond to antigen-presenting cells (APCs) and release cytotoxic and lytic factors, which result in thyrocyte death. Thyroid follicular cells have themselves been demonstrated to express HLA class II molecules, suggesting that they may also have a direct role in antigen presentation.

The humoral immune response is also important in AH. More than 90% of individuals with AH have detectable TPO antibodies. Autoantibodies directed against Tg and, to less degree, the TSH-R are also commonly detected. In vitro, TPO antibodies can fix complement and directly induce cell damage. Their presence within thyroid follicles in AH patients suggests they may have the same effect in vivo, although the thyrocyte destruction found in the *Rag1*-deficient mouse suggests TPO antibodies are not necessary for AH. Interestingly, the epitopes recognized by TPO antibodies in both GD and AH overlap, and there is no disease specificity for the targeted TPO domain. TRAbs found in rare patients with AH are likely to exert a blocking or antagonist effect, thus inducing hypothyroidism. As thyroid hormone secretion falls, increasing thyrocyte stimulation by elevation of serum

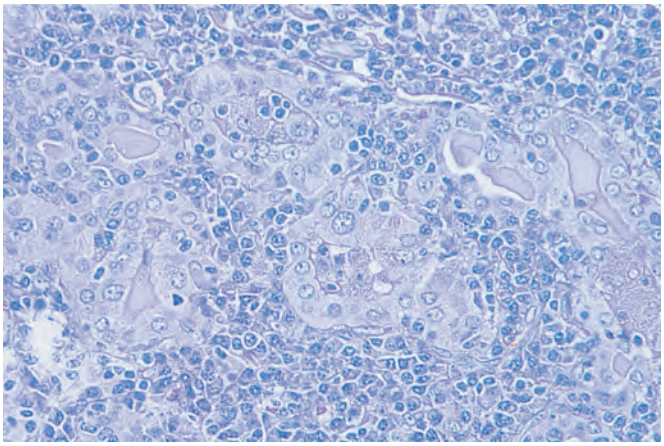


FIG. 70.7 Lymphocytic Infiltration of the Thyroid in a Patient with Autoimmune Hypothyroidism.

TSH may induce or augment thyroid autoantigen expression (*e.g.*, TPO, Tg), thereby perpetuating the autoimmune response.

Clinical Presentation

Hypothyroidism can result in changes in almost every organ system in the body (Table 70.3). Initially, the signs and symptoms may be subtle and nonspecific, including tiredness, cold sensitivity, and constipation. Hypothyroidism is frequently diagnosed incidentally following blood tests for another problem. The typical goiter palpable in Hashimoto thyroiditis is moderate in size and firm with a finely granular surface. Individuals often report a gradual increase in size over a number of years. However, rapid growth is unusual. In atrophic AH, the size of the thyroid gland is reduced.

Investigation and Diagnosis

Hypothyroidism is detected biochemically by a raised serum TSH with reduced fT_4. AH is differentiated from other forms of hypothyroidism by the presence of circulating autoantibodies, including TPO and Tg. On ultrasound scanning, the thyroid gland appears finely heterogeneous and hypoechoic, predating serum autoantibody positivity in some.

Reflecting the effects of hypothyroidism on multiple organ systems, many biochemical and hematologic abnormalities, such as mild anemia, hyponatremia, or elevated serum creatine kinase, transaminases, lactate dehydrogenase, and low-density lipoprotein (LDL) cholesterol, are also commonly detected in patients with AH.

Management

AH requires lifelong treatment with thyroid hormone replacement therapy. The most commonly prescribed is synthetic thyroxine, levothyroxine ($L–T_4$), which is widely available and inexpensive. Except for individuals with known heart disease or the very old, a full, weight-related replacement dose (≈1.6 µg/kg/day) should be started. Once the patient is on a stable dose, thyroid function should be assessed annually to ensure that the patient continues to receive the appropriate dose. Some commonly prescribed medications, such as calcium and iron supplements, and proton pump inhibitors interfere with the absorption of $L–T_4$, and patients should be advised to take these at least 4 hours before or after their $L–T_4$ to ensure maximum absorption.

TABLE 70.3 Common and Rare Clinical Manifestations of Hypothyroidism

Common	Rare
Neuropsychiatric	
Lethargy	Cerebellar ataxia
Impaired cognitive function	Deafness
Slow speech	Psychosis
Depression	
Gastrointestinal	
Anorexia	Ascites
Weight gain	
Constipation and bloating	
Abnormal liver function tests	
Cardiorespiratory	
Shortness of breath on exertion	Pericardial effusion
Reduced exercise tolerance	
Bradycardia	
Diastolic hypertension	
Cardiomegaly	
Low-voltage electrocardiogram	
Peripheral edema (nonpitting)	
Genitourinary	
Oligomenorrhea, amenorrhea, menorrhagia	
Reduced libido	
Early fetal loss	
Impotence	
Cutaneous	
Cold intolerance	
Skin dryness and thickening	
Malar flush	
Edema of the face, hands, and feet	
Change in face shape	
Pallor	
Nail abnormalities	
Alopecia	
Musculoskeletal	
Bradykinesia	
Joint and muscular pains	
Delayed relaxation of tendon reflexes	
Miscellaneous	
Goiter (in Hashimoto thyroiditis)	
Reduced basal metabolic rate	
Increased sensitivity to exogenous insulin	
Abnormal lipid metabolism	

Subclinical Hypothyroidism

Although the need to treat individuals with overt AH is universally accepted, it is unclear whether thyroid hormone replacement is beneficial in individuals with persistent subclinical hypothyroidism (increased serum TSH, but fT_4 and fT_3 within the normal reference range on at least two separate occasions).[26] Progression to overt hypothyroidism from this state occurs in approximately 2% of individuals per year who are TPO antibody negative, increasing to 5% per year if antibodies are present. Persistent subclinical hypothyroidism has been associated with a number of markers of cardiac and vascular dysfunction in observational studies, including left ventricular diastolic dysfunction, increased vascular resistance, and atherosclerosis.

A randomized clinical trial demonstrated no beneficial effects of treatment with levothyroxine in older adults (≥65 years old) with persistent subclinical hypothyroidism, although unfortunately few patients had TSH levels greater than 10 mIU/L and it did not include assessment of circulating TPO autoantibody titers which would indicate the cohort more likely to have progressive hypothyroidism and therefore possibly benefit from long-term levothyroxine therapy. Although this study was underpowered to assess cardiovascular outcomes, normalization of TSH with levothyroxine was associated with no difference in carotid intima-media thickness and carotid atherosclerosis in these patients.

CLINICAL PEARLS

Management of Persistent Subclinical Hypothyroidism

- If thyroid-stimulating hormone (TSH) is elevated, immediate treatment should be offered:
  - In pregnancy
  - Preconception, if planning a pregnancy.
- If TSH is elevated but less than 10 mIU/L, a trial of treatment can be offered:
  - To individuals with convincing symptoms of hypothyroidism.
- If TSH is greater than 10 mIU/L, treatment should be offered:
  - To individuals younger than the age of 70 years
  - To individuals older than the age of 70 years if there is a clear history of hypothyroid symptoms or there are significant risk factors for cardiovascular disease.

FUTURE DEVELOPMENTS

The genetics of AH remains understudied considering its frequency as the commonest autoimmune disease in humans. Considerable work remains to be done on whether treatment of subclinical hypothyroidism is beneficial. Given the insidious nature of the development of AH, it remains a good target for a preventive immunotherapeutic intervention, if a safe and economic treatment can be found.

OTHER FORMS OF THYROIDITIS

The term *thyroiditis* relates to conditions resulting in inflammation of the thyroid gland. A number of etiologies have been described, including infection, radiation exposure, drugs, and autoimmune factors. A common pattern to the natural history of several thyroiditides is frequently seen, involving an initial thyrotoxic phase of 1 to 3 months, followed by a rapid drop in serum thyroid hormones and a transient hypothyroid phase, often lasting another 1 to 4 months. During the thyrotoxic phase, preformed thyroid hormone stores are released from the thyroid follicles, leading to thyrotoxicosis, which may be severe. The hypothyroid phase follows when these preformed stores are exhausted, and the thyroid has become depleted of hormones. In approximately 90% of cases this hypothyroidism is transient, but in some cases, it never resolves.

POSTPARTUM THYROIDITIS

Postpartum thyroiditis (PPT) is a common endocrine condition that manifests within 1 year following pregnancy.[27] It affects between 5% and 10% of women in the general population. PPT is classically a biphasic disorder, consisting of a period of transient thyrotoxicosis (median onset 12 to 14 weeks postpartum) followed by a period of transient hypothyroidism (median onset 18 to 20 weeks postpartum); however, a monophasic (thyrotoxicosis or hypothyroidism alone) or reversed biphasic (hypothyroidism followed by thyrotoxicosis) pattern can also occur. During pregnancy, there is a state of relative immune tolerance, followed by a rebound in immune function following delivery, coinciding with the occurrence of PPT. The presence of thyroid autoantibodies and a lymphocytic infiltrate on thyroid biopsy supports an autoimmune basis for this condition.[27]

Clinically, the thyrotoxic phase of PPT is often mild, resulting in symptoms of fatigue and irritability, which can be misdiagnosed as postnatal depression. If the thyrotoxic episode is short, it may even go unnoticed. Neck pain is not a feature. Women with PPT who are thyrotoxic may benefit from a beta-blocker, such as propranolol, for symptom relief. Antithyroid drugs are not effective, as the thyrotoxicosis is caused by release of preformed thyroid hormones. Following the episode of thyroid dysfunction, 10% to 20% of women remain permanently hypothyroid.[27] In women who have had PPT and then recovered, an annual assessment of thyroid function is recommended because their risk of long-term hypothyroidism is considerable.[27] In those women who return to being euthyroid, there is a 75% risk of PPT in subsequent pregnancies and a 50% risk of permanent hypothyroidism at 7 years.

IMMUNOMODULATION AND THYROID DYSFUNCTION

Agents that modulate the immune system are increasingly being used as a treatment in various settings, either to induce, or alternatively, deplete and "reconstitute" the immune system. This disruption of immune homeostasis and costimulatory pathways lowers the threshold for the initiation of inflammation or autoimmunity. Thyroid dysfunction is one of the commonest endocrine sequelae of immunomodulatory therapies.

Alemtuzumab

ALZ, an anti-CD52 mAb licensed to treat MS, depletes both T and B lymphocytes to allow immune reconstitution with the aim of restoring a tolerogenic environment. Although used successfully to treat MS, up to 41% of those receiving ALZ will develop some form of thyroid autoimmunity, of which the vast majority is GD.[28]

The precise mechanism by which ALZ causes thyroid autoimmunity remains uncertain, but it is proposed to result from the proliferation of T cells during immune reconstitution, disrupting self-tolerance and driving the humoral autoimmune response. These patients can demonstrate an unpredictable course, and often present with fluctuating thyroid hormone levels due to the presence of both blocking and stimulating TRAbs.[28] The treatment of thyroid dysfunction, including GD, is largely the same as for those with spontaneous onset disease, although whether the prognosis is the same for ALZ-induced GD remains disputed.

Immune Checkpoint Inhibitors

Costimulatory "immune checkpoint" proteins are essential components of the normal immune system, negatively regulating the immune response to protect against uncontrolled immune destruction of healthy cells. CTLA-4 and programmed death

1/ligand 1 (PD-1/PD-L1) are inhibitory cell-surface receptors that suppress T-cell activation.

These targets have been exploited in the development of mAb for use in cancer immunotherapy, to modulate the immune system by blocking inhibitory checkpoints, resulting in T-cell activation and immune destruction of tumor cells.[29]

Unfortunately, this therapeutic approach often generates inflammatory immune-related adverse effects (irAEs), which often result in endocrinopathies, most commonly affecting the pituitary and thyroid gland.[30] Thyroid dysfunction, including hypothyroidism, thyrotoxicosis, and painless thyroiditis, is more frequently associated with anti–PD-1-antibodies. The combined use of anti–PD-1 and anti–CTLA-4 therapies has been associated with a 22% incidence of either thyroiditis or hypothyroidism, with a variable onset of thyroid dysfunction, occurring at a median of 8 weeks following the start of treatment.[30]

The natural course of thyroid dysfunction in these patients can be rapidly destructive, with an initial asymptomatic thyrotoxicosis followed by likely permanent hypothyroidism.[30] The potentially asymptomatic nature of the associated thyroid dysfunction highlights the importance of monitoring through routine laboratory surveillance.[30] However, it is important to be aware that a low TSH may be a manifestation of hypophysitis, rather than thyroid gland dysfunction.

The growing use of immunomodulatory therapeutics is influencing thyroid disease epidemiology, resulting in an expanding patient group with immune-related thyroid dysfunction that is likely to continue to grow with increasing use of these novel therapeutics.

TRANSLATIONAL RESEARCH

ON THE HORIZON

New Approaches to Therapy of Graves Disease

- Novel immunotherapeutic agents for Graves orbitopathy.
- Novel therapies targeting B-cell activity, including blocking of CD40-CD40L interactions, B cell–activating factor (BAFF) and immunoglobulin recycling.
- Thyroid-stimulating hormone receptor (TSH-R) therapies including TSH-R–blocking antibodies, TSH-R–specific peptide immunotherapy, and small-molecule TSH-R antagonists for the management of Graves hyperthyroidism.

The major challenge of the next 5 to 10 years is to take novel immunotherapeutic agents, including "biologics" that have been developed for rheumatic disorders into the clinical arena for autoimmune thyroid diseases. The primary target of these efforts should be GO, which remains a disfiguring condition, often with substantial functional impairment of vision and associated low quality of life. Phase III studies using TCZ and teprotumumab have demonstrated positive findings, leading to US Food and Drug Administration (FDA) approval for teprotumumab in the treatment of GO. Early diagnosis of GO and development of markers that predict progressive or severe disease will also be helpful in identifying patients for early intervention. Novel therapies for the treatment of Graves hyperthyroidism including those that target various aspects of B-cell activity and the TSH-R–specific immunotherapeutics remain under investigation. These agents might have a role in patients who are unlikely to gain medical remission from thionamide antithyroid drugs or in those for whom rapid control of hyperthyroidism is desirable.

REFERENCES

1. Tunbridge WM, Evered DC, Hall R, et al. The spectrum of thyroid disease in a community: the Whickham survey. *Clin Endocrinol (Oxf)*. 1977;7:481–493.
2. Laurberg P, Pedersen KM, Vestergaard H, et al. High incidence of multinodular toxic goitre in the elderly population in a low iodine intake area vs. high incidence of Graves' disease in the young in a high iodine intake area: comparative surveys of thyrotoxicosis epidemiology in East-Jutland Denmark and Iceland. *J Intern Med*. 1991;229:415–420.
3. Brix TH, Kyvik KO, Christensen K, et al. Evidence for a major role of heredity in Graves' disease: a population-based study of two Danish twin cohorts. *J Clin Endocrinol Metab*. 2001;86:930–934.
4. Vaidya B, Kendall-Taylor P, Pearce SH. The genetics of autoimmune thyroid disease. *J Clin Endocrinol Metab*. 2002;87:5385–5397.
5. Eschler DC, Hasham A, Tomer Y. Cutting edge: the etiology of autoimmune thyroid diseases. *Clin Rev Allergy Immunol*. 2011;41:190–197.
6. Ueda H, Howson JM, Esposito L, et al. Association of the T-cell regulatory gene CTLA4 with susceptibility to autoimmune disease. *Nature*. 2003;423(6939):506–511.
7. Brent GA. Environmental exposures and autoimmune thyroid disease. *Thyroid*. 2010;20:755–761.
8. Vestergaard P. Smoking and thyroid disorders—a meta-analysis. *Eur J Endocrinol*. 2002;146:153–161.
9. Cawood TJ, Moriarty P, O'Farrelly C, et al. Smoking and thyroid-associated ophthalmopathy: a novel explanation of the biological link. *J Clin Endocrinol Metab*. 2007;92:59–64.
10. Winsa B, Adami HO, Bergström R, et al. Stressful life events and Graves' disease. *Lancet*. 1991;338:1475–1479.
11. Song H, Fang F, Tomasson G, et al. Association of stress-related disorders with subsequent autoimmune disease. *JAMA*. 2018;319:2388–2400.
12. Chen F, Day SL, Metcalfe RA, et al. Characteristics of autoimmune thyroid disease occurring as a late complication of immune reconstitution in patients with advanced human immunodeficiency virus (HIV) disease. *Medicine (Baltimore)*. 2005;84:98–106.
13. Smith BR, Bolton J, Young S, et al. A new assay for thyrotropin receptor autoantibodies. *Thyroid*. 2004;14:830–835.
14. Czarnocka B. Thyroperoxidase, thyroglobulin, Na(+)/I(−) symporter, pendrin in thyroid autoimmunity. *Front Biosci*. 2011;16:783–802.
15. Costagliola S, Bonomi M, Morgenthaler NG, et al. Delineation of the discontinuous-conformational epitope of a monoclonal antibody displaying full in vitro and in vivo thyrotropin activity. *Mol Endocrinol*. 2004;18:3020–3034.
16. Hegedus L. Treatment of Graves' hyperthyroidism: evidence-based and emerging modalities. *Endocrinol Metab Clin North Am*. 2009;38:355–371. ix.
17. Barbesino G, Tomer Y. Clinical utility of TSH receptor antibodies. *J Clin Endocrinol Metab*. 2013;98:2247–2255.
18. Perros P, Dayan CM, Dickinson AJ, et al. Management of patients with Graves' orbitopathy: initial assessment, management outside specialised centres and referral pathways. *Clin Med (Lond)*. 2015;15:173–178.
19. Bahn RS. Graves' ophthalmopathy. *N Engl J Med*. 2010;362:726–738.
20. Marcocci C, Kahaly GJ, Krassas GE, et al. Selenium and the course of mild Graves' orbitopathy. *N Engl J Med*. 2011;364:1920–1931.
21. Bartalena L, Baldeschi L, Boboridis K, et al. The 2016 European Thyroid Association/European Group on Graves' Orbitopathy guidelines for the management of Graves' orbitopathy. *Eur Thyroid J*. 2016(5):9–26.
22. Vanderpump MP, Tunbridge WM, French JM, et al. The incidence of thyroid disorders in the community: a twenty-year follow-up of the Whickham Survey. *Clin Endocrinol (Oxf)*. 1995;43:55–68.
23. Imaizumi M, Usa T, Tominaga T, et al. Radiation dose-response relationships for thyroid nodules and autoimmune thyroid diseases in Hiroshima and Nagasaki atomic bomb survivors 55–58 years after radiation exposure. *JAMA*. 2006;295:1011–1022.
24. Rotondi M, Carbone A, Coperchini F, et al. Diagnosis of endocrine disease: IgG4-related thyroid autoimmune disease. *Eur J Endocrinol*. 2019;180:R175–R183.

25. Quaratino S, Badami E, Pang YY, et al. Degenerate self-reactive human T-cell receptor causes spontaneous autoimmune disease in mice. *Nat Med*. 2004;10:920–926.
26. Pearce SHS, Brabant G, Duntas LH, et al. ETA guideline: management of subclinical hypothyroidism. *Eur Thyroid J*. 2013;2013(2):215–228.
27. Alexander EK, Pearce EN, Brent GA, et al. Guidelines of the American Thyroid Association for the diagnosis and management of thyroid disease during pregnancy and the postpartum. *Thyroid*. 2017;2017(27):315–389.
28. Pariani N, Willis M, Muller I, et al. Alemtuzumab-induced thyroid dysfunction exhibits distinctive clinical and immunological features. *J Clin Endocrinol Metab*. 2018;103:3010–3018.
29. Chang LS, Barroso-Sousa R, Tolaney SM, et al. Endocrine toxicity of cancer immunotherapy targeting immune checkpoints. *Endocr Rev*. 2019;40:17–65.
30. Ferrari SM, Fallahi P, Galetta F, et al. Thyroid disorders induced by checkpoint inhibitors. *Rev Endocr Metab Disord*. 2018;19:325–333.

Type 1 Diabetes

Leonard C. Harrison

Diabetes mellitus is not a single disease but a metabolic syndrome in which different mechanisms lead to deficiency of insulin and/or impaired insulin action and persistent hyperglycemia. The American Diabetes Association classified diabetes into four categories based on etiology rather than age of onset (juvenile-onset versus adult-onset) or requirement for insulin therapy (insulin-dependent versus noninsulin-dependent).[1] The vast majority of cases of diabetes—approximately 10% and 90%—are classified as type 1 and type 2 diabetes, respectively. This chapter focuses on type 1 diabetes, which results from an absolute deficiency of insulin secondary to the loss of pancreatic β cells. Type 1 diabetes is classified as 1 A (immune-mediated) or 1B (idiopathic), primarily depending on the presence or absence, respectively, of pancreatic islet autoantibodies. However, as discussed below, type 1 A diabetes (T1DA) and type 1B diabetes (T1DB) also differ in their natural history and clinical features.

Diabetes is diagnosed on the basis of the following criteria[1]: symptoms in association with a casual plasma glucose ≥200 mg/dL (11.1 mmol/L) OR a fasting plasma glucose ≥126 mg/dL (7.0 mmol/L) OR 2-hour plasma glucose ≥200 mg/dL (11.1 mmol/L) in an oral glucose tolerance test (OGTT; 75 g glucose in water). Without symptoms, the diagnosis of diabetes must rest on confirmation of a raised plasma glucose concentration. A further diagnostic criterion introduced by the World Health Organization (WHO) in 2011, in particular for the diagnosis of type 2 diabetes, is a confirmed blood glycated hemoglobin (HbA_{1c}) ≥48 mmol/mol (6.5%). Because HbA_{1c} is an integrated measure of glycemia over many weeks, it is not suitable for diagnosing children or in the following circumstances: suspected type 1 diabetes; symptoms of diabetes for <2 months; acute illness; medication that may increase blood glucose (*e.g.*, steroids, anti-psychotic drugs); pregnancy. The criteria for diagnosing gestational diabetes (in pregnancy) are stricter: fasting plasma glucose ≥101 mg/dL (5.6 mmol/L) or 2-hour plasma glucose in an OGTT ≥140 mg/dL (7.8 mmol/L).

The classic symptoms and signs of T1DA, secondary to high concentrations of blood glucose, are polyuria, polydipsia and unexplained weight loss; others include fatigue, increased hunger, impaired visual acuity due to changes in the refractive index of the vitreous humor, tingling or numbness in the hands or feet resulting from sensory nerve changes, and vaginal irritation caused by *Candida* infection. If diabetes is undiagnosed or untreated, failure to metabolize glucose may lead to the breakdown of fat, resulting in ketonemia and ketoacidosis, with nausea and hyperventilation preceding life-threatening ketoacidotic coma. In children who present with the classic symptoms, T1D can be diagnosed clinically. However, in Caucasians the diagnosis can be confirmed, and in older individuals and in less clear-cut cases clearly established, by detecting circulating autoantibodies to islet antigens. Autoantibodies to insulin (IAA), glutamic acid decarboxylase 65,000 mol. wt. isoform (GADA), insulinoma-like antigen-2 (IA-2A) and zinc transporter-8 (ZTA) are markers of β-cell autoimmunity that usually appear many months to years before symptoms and therefore denote high risk for clinical disease.[2–5] Islet autoantibodies are detected in the majority of Caucasian children with T1DA (compared with ≤1% of the background population), but in only ≈50% of Hispanic and African American children with diabetes, an increasing number of whom have T2D and, in some cases, T1DB. Negative results for islet autoantibodies in children with diabetes should also flag the possibility of relatively rare genetic disorders of β cells, viz. monogenic maturity-onset diabetes of the young (MODY) and sulfonylurea receptor syndromes. That the presence of autoantibodies to two or more islet antigen specificities is so predictive of clinical T1DA has led to a paradigm shift in defining T1DA primarily as an autoimmune β-cell disorder ("ABCD")[6] and to its staging: autoantibodies to two or more islet antigen specificities define Stage 1; loss of β-cell mass/function leading to abnormal plasma glucose concentrations defines Stage 2; and symptomatic clinical disease defines Stage 3.[7]

The hallmark of T1DA is progression to absolute insulin deficiency within several years after clinical diagnosis. The connecting peptide in proinsulin (C-peptide), secreted in equimolar amounts to insulin, is used as a surrogate for insulin to evaluate residual β-cell function, because measurement of plasma insulin may be inaccurate in the face of treatment with exogenous insulin or the presence of IAA or insulin antibodies induced by exogenous insulin. The area under the curve of plasma C-peptide after a mixed meal tolerance test (MMTT) is the gold standard for assessing β-cell function and has been used as the primary outcome in clinical trials of immune therapy in recent-onset Stage 3 type 1 diabetes. However, the MMT requires repeated venous blood sampling for up to four hours and is not convenient in routine clinical settings, in which β-cell function is of increasing interest with the introduction of immune therapies for T1DA. To facilitate assessment of β-cell function, an algorithm incorporating clinical variables and biochemical measures at a single time-point has been shown to be an accurate substitute for the full MMTT, including for the identification of treatment effects in Stage 3 T1DA.[8]

Although children with T1DA have lost most of their β-cell function at diagnosis, the measurement of plasma C-peptide is not a reliable way of distinguishing T1D, especially at diagnosis. Hyperglycemia impairs β-cell function and—when corrected by rehydration and insulin replacement—can be followed by a

"honeymoon phase" of partial recovery of β-cell function and a decreased requirement for exogenous insulin that may last many months. Several years after diagnosis, most young children display little residual β-cell function; however, in older children and adults, residual C-peptide secretion has been observed for many years and may be associated with better glycemic control.

Classically, T1DA was considered a disease of "juvenile onset" in normal-weight individuals, in contrast to type 2 diabetes in middle-aged, overweight individuals. However, this view requires reappraisal.[9] First, up to 10% of adults who present with diabetes diagnosed initially as T2D have evidence of low-grade islet autoimmunity manifest by the presence of GADA, and occasionally IA-2A or ZTA. As they appear to have a slowly progressive form of autoimmune β-cell destruction, termed *latent autoimmune diabetes in adults* (LADA), their number doubles the prevalence of T1DA. Second, the "obesity epidemic" has impacted on T1D and T2D stereotypes. "Hybrid" types with β-cell autoimmunity and insulin resistance ("double diabetes") are becoming increasingly common, characterized by weaker immune (lower avidity autoantibodies, predominantly GADA) and lower risk genetic (human leukocyte antigen [HLA] alleles) markers. Obesity is associated with low-grade immune inflammation and insulin resistance that may promote and uncover latent β-cell autoimmunity in individuals with lower-risk genes for T1DA.

T1DB excludes known specific causes of β-cell dysfunction, such as monogenic diabetes. It encompasses forms of ketosis-prone diabetes initially described in West Africans and African Americans[10] and subsequently in other ethnic groups, but rarely seen in Caucasians. Obesity and relatively well-preserved residual β-cell function were features of the African cases. T1DB also includes "fulminant diabetes" initially described in Japan, where it may account for 15% to 20% of new-onset T1D.[11,12] Fulminant diabetes is associated with widespread mononuclear cell infiltration of both the exocrine and endocrine pancreas and elevated concentrations of serum pancreatic amylase, elastase, and lipase in individuals with HLA susceptibility genes for T1D.[12] Case reports have associated it with acute viral infection and drug hypersensitivity reactions, but its etiology remains unclear. In Caucasians, T1D in which islet autoantibodies are undetectable has been called T1DB, but this may reflect assay insensitivity or the waxing and waning of autoantibodies over time. A classification of diabetes that hinges on the presence or absence of islet autoantibodies is not ideal. β Cell–specific T cells, not islet autoantibodies, are the effectors of β-cell damage, and autoantibodies and autoreactive T cells to islet antigens can be reciprocally related.[13] In addition, the more recent discovery of autoantibodies to ZT8 suggests the possibility that other target islet antigens remain to be discovered and might be markers of particular subtypes of T1D. For example, in fulminant diabetes, the clinical picture suggests that if autoimmunity is involved it also includes the exocrine pancreas. Furthermore, in contrast to T1DA, in which there is an extended preclinical history of islet autoimmunity, autoantibodies may not be present at the time when fulminant diabetes presents acutely.

CLINICAL PEARLS

- In Caucasian children, the diagnosis of T1D can be made on the basis of classic clinical features.
- Islet autoantibodies are present in more than 90% of Caucasian children presenting with diabetes but in only about 50% of Hispanic or African American children, some of whom have T2D.
- The presence of islet autoantibodies confirms the diagnosis of type 1A diabetes (T1DA).
- Islet autoantibodies are detected in the first 3 years of life in the majority of Caucasian children who go on to develop T1DA.
- The risk for T1DA increases with the number and titer of islet autoantibodies.
- The T1D stereotype of the thin juvenile is beginning to overlap with the T2D stereotype of the obese, insulin-resistant adult.
- T1DB has been characterized by an absence of islet autoantibodies and often a more fulminant natural history, but T1DA and T1DB have overlapping pathologic features and diagnostic separation based on islet autoantibodies is not ideal.

EPIDEMIOLOGY AND NATURAL HISTORY OF T1DA

The incidence of T1DA varies widely across the world, being highest among Caucasians of Northwestern Europe and countries to which they have emigrated (Fig. 71.1). This, in part, reflects the population distribution of HLA risk genes, which account for up to half of the lifetime risk for T1DA. However, the incidence of T1DA is rising in many countries on a background of lower-risk HLA alleles. In Western societies, the incidence in childhood has more than doubled since the 1980s and has been rising at ≈3% annually, particularly among younger children.[14] The same trend is also occurring in other countries, such as Kuwait and Saudi Arabia and some regions of India and China, where Western lifestyles have been adopted but where the prevalence of high-risk HLA haplotypes for T1DA is much lower. As in T2D, environmental factors may increase the penetrance of risk genes for T1DA. In the case of HLA genes in Caucasians, the increasing incidence of T1DA is accounted for by children with intermediate-risk (DR 4,4 or DR 3,3) or low-risk (DR 4,X or DR 3,X) phenotypes, not the highest-risk HLA phenotypes (DR 3,4; DQ 2,8).[15] These lower-risk phenotypes are also seen in non-Caucasians and adults with T1DA. The greatest number of children with diabetes will soon be found in the most populous regions of the world, India and China.

T1DA affects both sexes equally in childhood, with a slight excess of males in early adult life. Of newly diagnosed cases, no more than 10% to 15% have a family history of T1DA. However, studies of affected families have yielded major insights into the genetics and natural history of T1DA. In T1DA relatives, the rate of progression to clinical diabetes is positively associated with the number and titer of islet autoantibodies,[2,3,5] the number of HLA class 2 risk alleles (*DR3, DR4*) and specific HLA class I alleles (*A24*),[16] and the degree of insulin resistance[17] and is negatively correlated with age. The specificity of islet autoantibodies is also important. IAAs are more often the first sign of islet autoimmunity in children who are followed from birth and, of all the autoantibodies, has the highest predictive value.[2–5] In Europe, North America, and Australia, birth cohort studies of children with a T1DA relative have shown that the development of diabetes by the age of 18 years is almost always associated with the appearance of islet autoantibodies in the first few years of life.[5] Of children with ≥2 islet autoantibodies before the age of 3 years, 57% (95% confidence interval [CI] 51.7% to 62.3%) and 74.8% (95% CI 69.7% to 79.9%) progressed to diabetes by 6 and 10 years of age, respectively. With a single islet autoantibody, 14.5% progressed to diabetes by 10 years of age. [5]

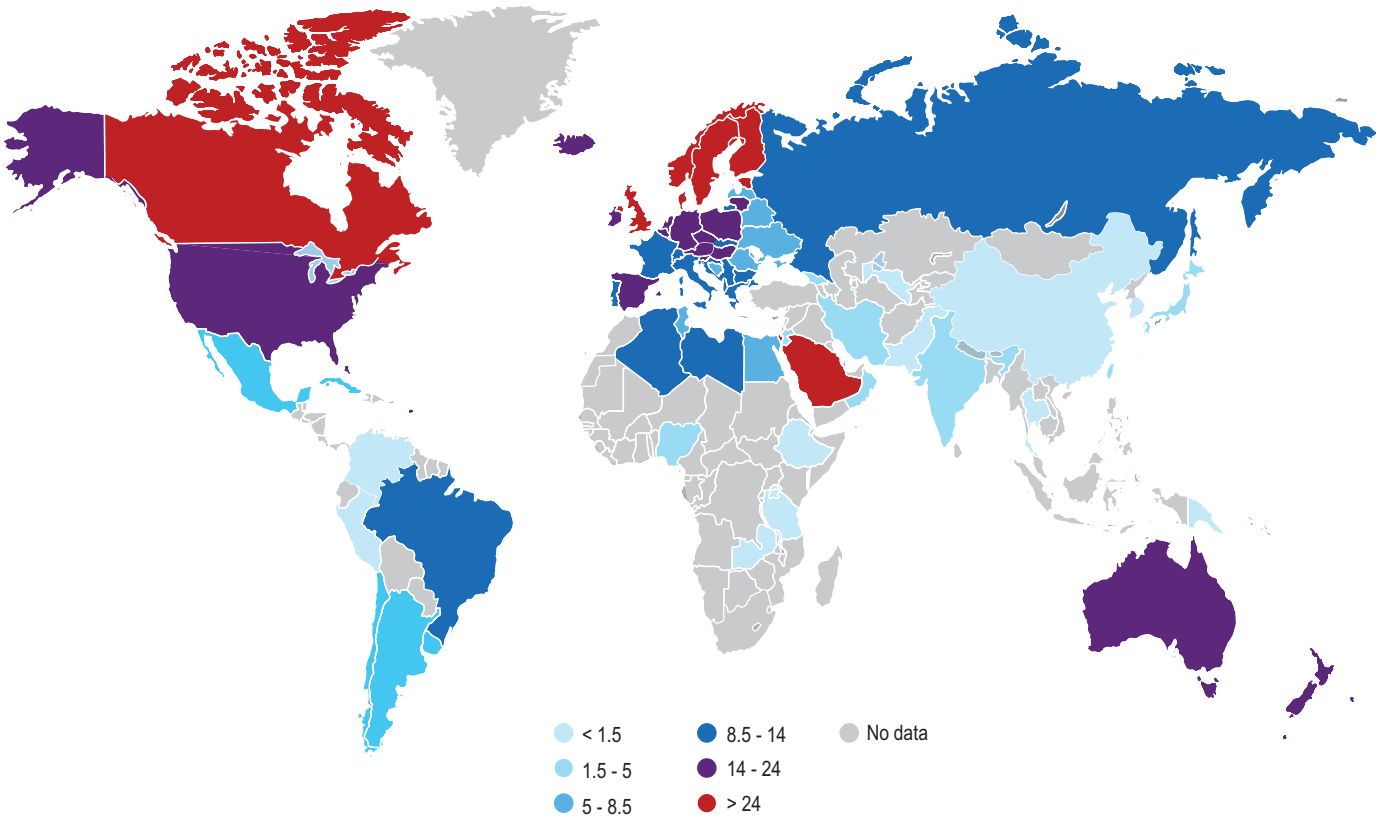


FIG. 71.1 Global Incidence of type 1A diabetes (T1DA) in 100,000 children ages 0 to 14 years per year. (From *IDF Diabetes Atlas*, 7th ed. Brussels, Belgium: International Diabetes Federation; 2015.)

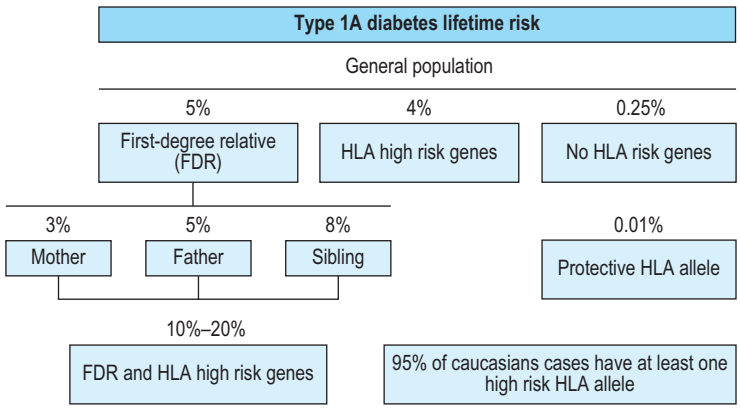


FIG. 71.2 Lifetime risk of type 1A diabetes. *HLA*, Human leukocyte antigen.

Men with early-onset T1DA are more likely to transmit diabetes to their offspring compared with women (Fig. 71.2), but it is not clear whether this results from genomic imprinting of a gene expressed only from a paternally inherited allele, protection from the mother including via her microbiome, or other possible reasons. Risk for T1DA is also greater in later-born offspring. The peak age of incidence in children is at the cusp of puberty, a time when the body's requirement for insulin increases along with an increase in insulin resistance. Like the seasonal autumn–winter peak in diagnosis attributed to viral infections, this is likely to be "the straw that breaks the camel's back" on a background of longer-standing autoimmune β-cell disorder.

PATHOGENESIS: NATURE AND NURTURE

In the contemporary model for the staged natural history of T1DA (Fig. 71.3), the pattern and rate of β-cell loss is depicted as linear, but it is more likely to be episodic—reflecting direct (*e.g.*, virus, chemical toxin) or indirect (*e.g.*, diet, pollutants,

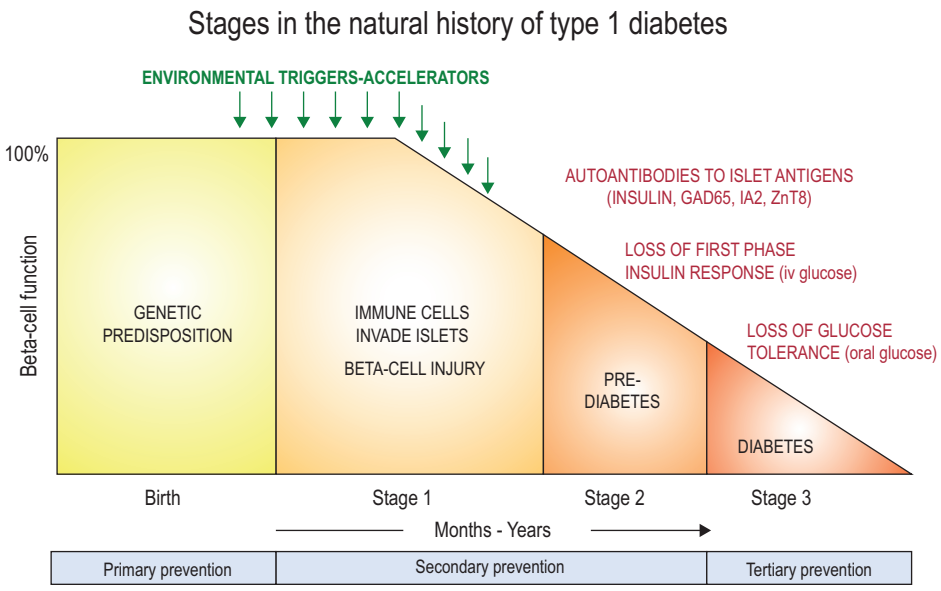


FIG. 71.3 Natural history of type 1A diabetes.

microbiome) interactions between the environment and β cells. Furthermore, there is evidence that immune activation and β-cell destruction may be accelerated in the late preclinical stage. The appearance of predictive islet autoantibodies in the first years of life[5] means that the stage for developing T1DA is set very early, even before birth, on a background of genetic susceptibility. These early years provide clues to environment–gene interactions that lead to immune dysfunction and disease, now being sought in pregnancy-birth cohort studies.[18]

It is generally accepted that β-cell destruction in T1DA is mediated by autoreactive T cells, the ultimate effectors being CD8 cytotoxic T cells (Fig. 71.4). The evidence for this is unequivocal in the inbred nonobese diabetic (NOD) mouse model of T1DA, which shares several key features with T1DA in outbred humans (Table 71.1) despite 65 million years of evolutionary distance. The molecular mechanisms of β-cell death, gleaned mostly from the NOD mouse, encompass a combination of both apoptosis induced by activation of extrinsic (*e.g.*, tumor necrosis factor [TNF] receptor or Fas ligation) or intrinsic (*e.g.*, endoplasmic reticulum [ER] stress) pathways and necroptosis induced by cytotoxic CD8 T-cell granule components (granzymes and perforin), reactive oxygen species (ROS), or ischemia. However, these findings cannot be simply extrapolated to humans in whom access to the pancreas is limited. Studies of pancreas biopsies and organ-donor pancreas—more recently from the Network for Pancreatic Organ Donors with Diabetes (nPOD; www.jdrfnpod.org)—have not revealed the florid immune cell infiltration of islets (insulitis) seen in NOD mice but rather patchy insulitis predominantly from CD8 T cells and macrophages. Confirming earlier findings,[19] HLA class I is hyperexpressed by islets together with immunoreactive interferon-α (IFN-α), even in atrophic islets lacking insulin, which is evidence for the presence of viral nucleic acid. A key question is: why are β cells specifically targeted? The answer may lie in insulin itself. Central to the concept of autoimmune disease is the notion that the pathology is driven by loss of immune tolerance to self-antigens. In the case of T1DA, considerable evidence, direct in the NOD mouse, identifies (pro)insulin as a key disease-initiating self-antigen (Table 71.2).

TABLE 71.1 The Nonobese Diabetic Mouse as a Model of Human Type 1A Diabetes

Feature	NOD Mouse	Human
Preclinical stage	Yes	Yes
Gender	F > M	M > F after puberty
Genetic susceptibility		
MHC class II 57 non-Asp	Yes (I-Ag7)	Yes (HLA-DQ8)
Polygenic non-MHC	Yes	Yes
Environmental influence on gene penetrance	Yes	Yes
Disease transmission via bone marrow	Yes	Yes
Mononuclear cell infiltration of islets (insulitis)	Marked	Moderate
Other organs	Yes	Sometimes
Impaired immune regulation	Yes	Yes
Autoantigens:		
(Pro)insulin	Yes	Yes
Glutamic acid decarboxylase	Yes	Yes
Clinical response to autoantigen-specific therapy	Yes	Not yet shown

NOD, Nonobese diabetic.

GENES

The concordance for T1D in monozygotic twins, who are almost genetically identical, approaches 50%. This indicates that both genetic and environmental–epigenetic mechanisms contribute to disease. Large case/control studies and genome-wide association studies (GWAS) have identified over 40 chromosomal loci associated with risk for T1DA (Fig. 71.5).[20] Many are weak associations defined by single-nucleotide polymorphisms (SNPs) and the genes with which the SNPs are in linkage and/or their functional contribution to pathogenesis are not yet known. Different combinations of genes/susceptibility loci in different environmental contexts most likely lead by different routes to the final outcome of β-cell destruction, consistent with phenotypic–clinical heterogeneity in T1D. Nevertheless, what is

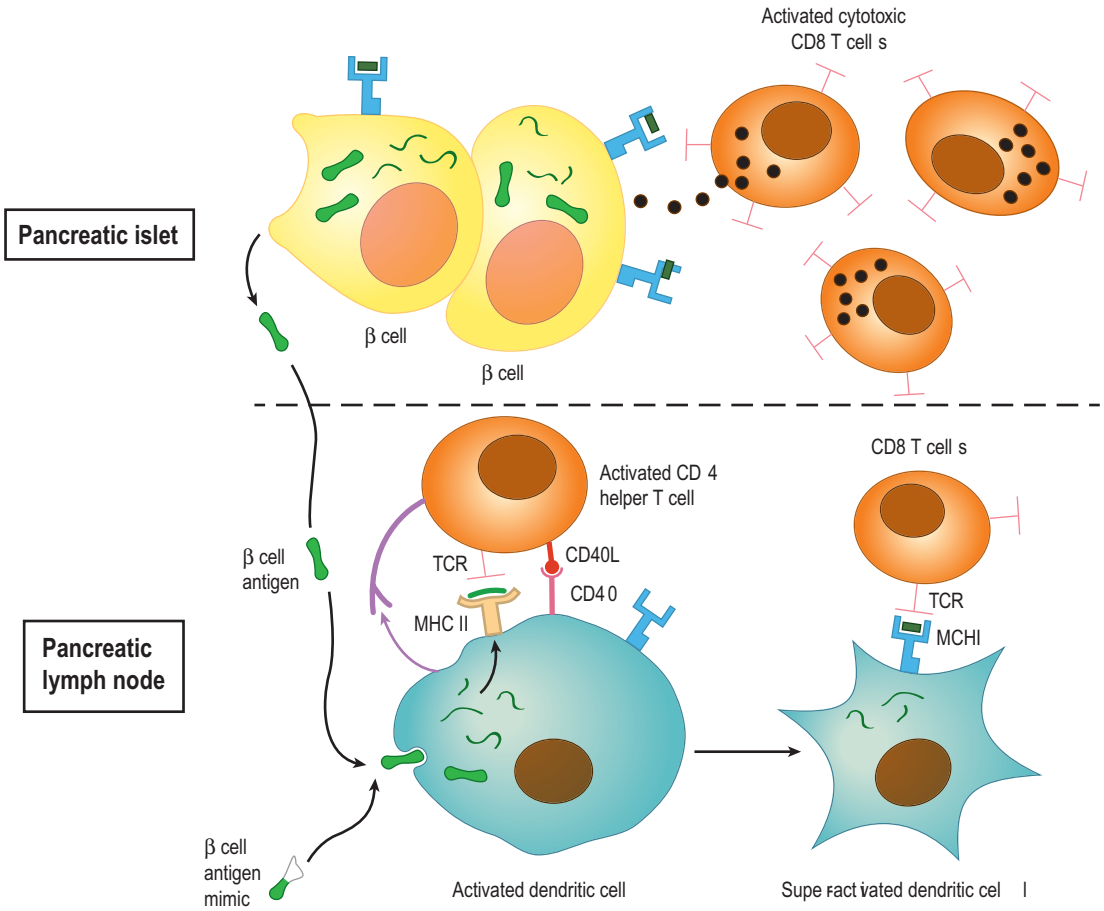


FIG. 71.4 Immune mechanisms of β-cell destruction in type 1A diabetes.

TABLE 71.2 Evidence for the Key Role of Proinsulin as Primary Autoantigen in Type 1A Diabetes

- β Cell–specific (except for thymus)
- Second strongest genetic locus in humans (*IDDM2*)=VNTR 5′ of INS allelism correlates with proinsulin transcription in the thymus
- Early target of autoimmunity in humans and nonobese diabetic (NOD) mice
- Main target of T cells isolated from islets of NOD mice and humans with T1DA

Genetic manipulation in the NOD mouse:

- Transgenic expression of proinsulin (but not glutamic acid decarboxylase) in antigen-presenting cells (APCs) prevents insulitis/diabetes
- Transfer of hematopoietic stem cells (HSCs) or myeloid progenitors encoding proinsulin in APC progeny prevents diabetes
- Knockout of proinsulin II (expressed in thymus) accelerates diabetes
- Induction of mucosal tolerance to (pro)insulin prevents diabetes

clear across populations globally is the dominance of the HLA locus (IDDM1) as the single most important genetic contributor to T1D risk (see Table 71.1), accounting for about half the lifetime risk. The highest risk HLA haplotypes in Europeans are DR3 (*DRB1*03:01-DQA1*05:01-DQB1*02:01*) and DR4 (usually *DRB1*04:01* or **04:04* with *DQA1*03:01-DQB1*03:02*). On the other hand, the HLA DQ6 haplotype *DQA1*01:02-DQB1*06:02*, in linkage with DR15 (*DRB1*15:01*), is dominantly protective for T1DA. After the HLA loci, the next most important contribution is from the insulin gene (*INS*) locus (IDDM2), which maps to a variable number of tandem repeats (VNTR) 5′ of the coding sequence. Apart from the HLA loci, IDDM2 is still the only locus for which genome-wide association is reflected by linkage, which might be explained by disease heterogeneity. Long (class III) and short (class I) VNTR alleles are associated, respectively, with higher and lower transcription of proinsulin messenger RNA (mRNA) in medullary thymic epithelial cells (mTECs) under the control of the autoimmune regulator gene (*AIRE*)[21] and with lower and higher T1DA risk.[22] We infer, therefore, that IDDM2 controls the extent of deletion of proinsulin-specific T cells during their intrathymic development, which would predict that the long VNTR should be associated with fewer proinsulin-specific T cells in the periphery; the evidence for this is equivocal, perhaps because of the challenge of measuring islet antigen-specific T cells in human blood. The *INS* polymorphism is unique to humans. Mice, instead, have two insulin genes: *INSI* is expressed in the β cell and *INSII* in the thymus, and the latter is decreased in the NOD mouse. In summary, IDDM2 provides a genetic mechanism for autoimmune targeting of proinsulin and the β cell.

Most of the other candidate genes (see Fig. 71.5) are involved in immune function and are associated with other autoimmune diseases. For example, a nonsynonymous gain-of-function polymorphism in *PTPN22* (lymphoid tyrosine phosphatase) enhances T-cell suppression, which may impair negative selection of autoreactive T cells in the thymus. Polymorphisms around *IL2RA* (interleukin-2 [IL-2] receptor α chain; CD25) and *IL2* itself are associated with impaired IL-2 signaling, which impairs

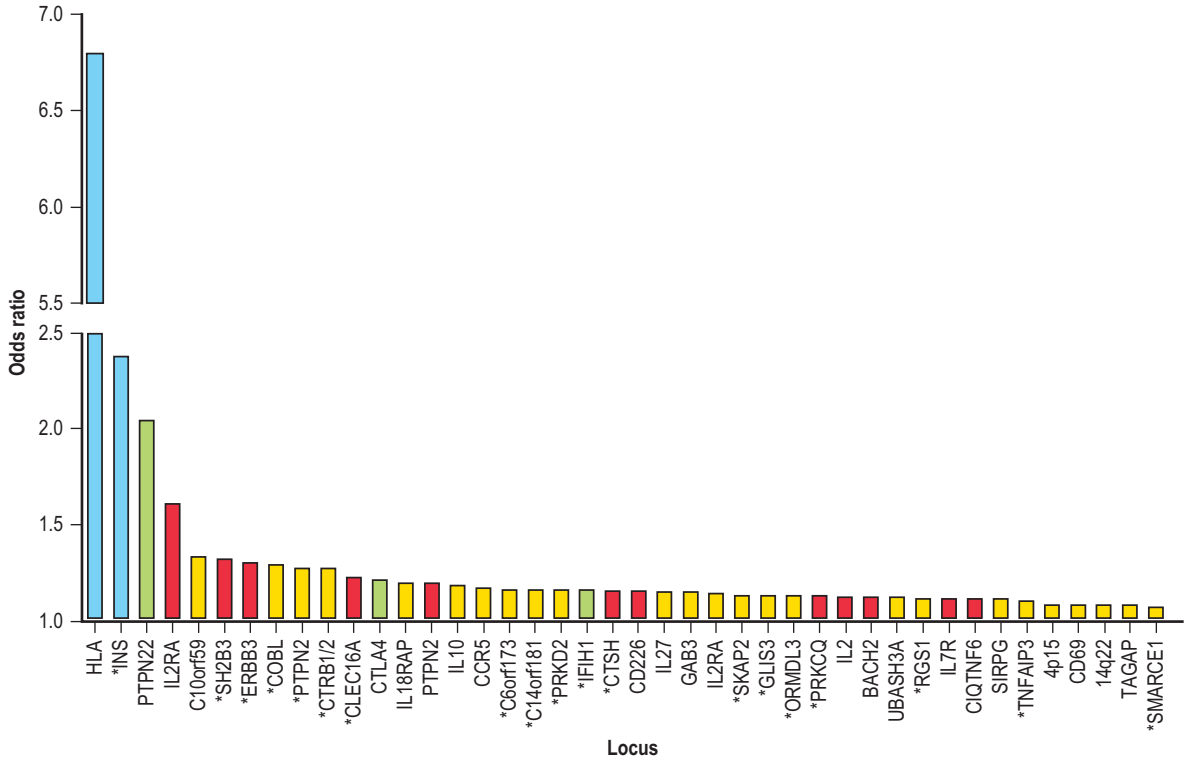


FIG. 71.5 Candidate genes in type 1A diabetes (T1DA). The *y*-axis shows the odds ratio for risk alleles at each of the loci on the x-axis.[20] The majority of the known genes are involved in immune responses, and many (marked *) are expressed in islets. Colors indicate era of identification: *blue* 1970–2000; *green* 2001–2006; *red* 2007–2008; *yellow* 2009–2010.

the generation and maintenance of regulatory T cells (Tregs) in both humans and NOD mice. Polymorphisms involving the *VDR* (vitamin D receptor) and *CYP27B1* (25-hydroxyvitamin vitamin D3 1-α-hydroxylase) and *IFIH1* (IFN-induced with helicase C domain 1, which enhances the type I IFN response to virus infection) are clues to gene–environment interactions in T1DA.

β CELLS

Why are β cells selectively destroyed by immune inflammation that involves the whole islet that also contains glucagon-secreting α cells and other endocrine cells? In the case of T1DB, the whole pancreas appears to be targeted. First, cytotoxic CD8 T cells recognize autoepitope peptides (*e.g.*, from insulin) presented by hyperexpressed HLA class I proteins on β cells. Second, because of their unique metabolic wiring, β cells may contribute to their own death at the hands of the immune system; evidence for "assisted suicide" is compelling.[23] Studies of rodent islets indicate that β cells lack efficient antioxidant and free radical scavenging mechanisms and are especially sensitive to mitochondrial oxidative and ER stress in response to cytokines and granzymes, but whether this applies equally to human islets is unresolved. Finally, other than insulin, many candidate T1D genes identified in GWAS are transcribed in β cells and encode proteins that interact with the immune system. Moreover, in response to inflammation, human islets generate hundreds of RNA splice variants for proteins, which, if translated as neoantigens, might not be subject to immune tolerance.

Although it is still unclear whether human β cells are more sensitive than other autoimmune target cells to immune effectors, what is clear is that having undergone apoptotic or necrotic death, β cells are not restorable in the face of autoimmune memory.

KEY CONCEPTS

Evidence that Type 1A Diabetes Is an Autoimmune Disease

- Association with other autoimmune diseases, including autoimmune polyendocrinopathy syndromes (*e.g.*, APS-1 due to mutations in *AIRE*)
- Presence of autoantibodies and T cells reactive with islet-cell antigens
- Strong association with specific human leukocyte antigen alleles and haplotypes and immune-response genes
- Transfer of disease by bone marrow or development of disease in the healthy pancreas transplanted from an identical twin without diabetes to the twin with T1DA
- Neonatal onset associated with loss of natural Tregs due to mutations in *FOXP3* (immune dysfunction/polyendocrinopathy/enteropathy/X-linked syndrome)

ENVIRONMENT—OUTSIDE AND INSIDE

The environment in its protean manifestations may impact β-cell function in multiple ways (*e.g.*, by activating innate immunity to, in turn, drive adaptive immunity to β cells in people at genetic risk for T1DA) by specifically activating innate immune cells (macrophages) in the islets to release inflammatory cytokines (IL-1β, TNF). This elicits oxidative ER stress or induces posttranslational modifications in β-cell proteins that

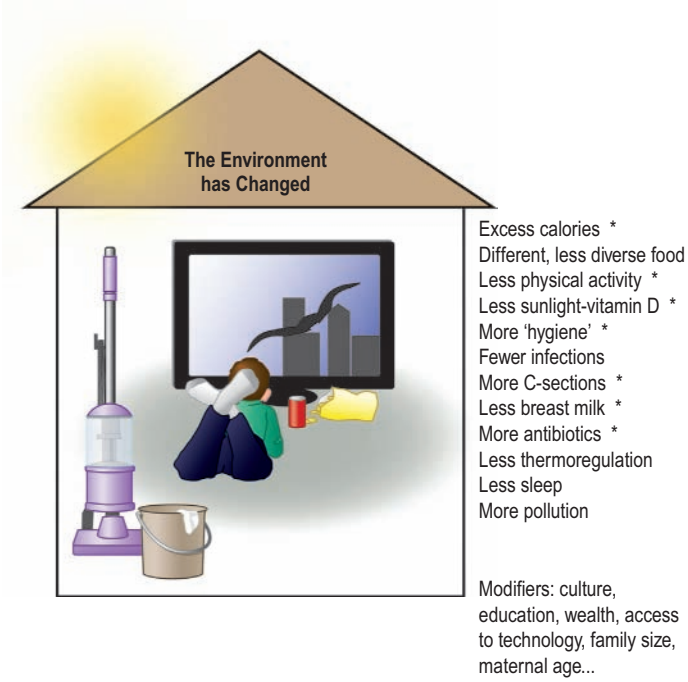


FIG. 71.6 Changes in the modern environment. Those shown to be associated with an increased risk for type 1A diabetes in humans or nonobese diabetic mice are indicated (*).

render them immunogenic, and promotes obesity and insulin resistance, thereby increasing the workload of the β cell.

The environment in Western societies has changed dramatically in many ways during the last century (Fig. 71.6), and some of these changes have been associated epidemiologically or in animal studies with development of T1DA.[9] An index of the contemporary "exposome" is the ongoing rise in the prevalence of obesity. When children at increased genetic risk for T1DA (with a T1DA first-degree relative) were monitored from birth, weight gain in the first 2 to 3 years of life was shown to be a risk factor for islet autoimmunity.[24] When such at-risk children developed islet autoantibodies and were then followed up over time, those with insulin resistance progressed most rapidly to clinical diabetes.[17] Whether insulin resistance is a risk factor for the development of islet autoimmunity is an open question that should be answered by ongoing pregnancy–birth cohort studies.[18] If insulin resistance synergizes with impaired insulin secretion to promote T1DA, then lifestyle factors must be considered in the efforts to forestall or prevent T1DA.

Excessive energy consumption leading to obesity is part of a complex interplay of related factors that promote low-grade inflammation and insulin resistance. These include a poor-quality diet and lack of physical exercise, sunlight (vitamin D), and sleep. The highly processed fast-food "Western" diet lacks diversity of components, plant-derived prebiotics, and complex carbohydrates (starches and fiber), is high in saturated fats and in sucrose and fructose sugars, and contains added artificial preservatives, emulsifiers, and sweeteners. All of these alter the composition of the gut microbiome and reduce its diversity, which is a feature found in children at risk for T1DA.[25] Diets containing a diverse range of plant products (cereals, fresh fruits, and vegetables) provide complex carbohydrates for fermentation by colonic bacteria to short-chain fatty acids (SCFAs), which are anti-inflammatory.[26] The influence of the environment on T1DA is evidenced by comparing highly developed Finland and geographically adjacent and ethnically similar but less modernized Russian Karelia. Finland has the highest incidence of T1DA in the world, currently around 60 cases/100,000 population ≤14 years of age, six-fold higher than Karelia. This difference could reflect a relatively lower rate of childhood infections (in Finland), in keeping with the "hygiene hypothesis."[27] It has also been associated with differences in gut microbiome composition[25] (see below).

Virus infection has long been considered a trigger of T1DA, consistent with increased expression of HLA class I and IFN-α observed in islets of humans with T1DA.[19] Enteroviruses are the leading candidates,[28] based mainly on the temporal association with the onset of clinical T1DA, but the significance of this has to be weighed against the long pre-symptomatic prodrome of T1DA. Viral mechanisms in T1DA could be direct or indirect (*e.g.*, infection of β cells, infection of the exocrine pancreas with bystander death of β cells, mimicry between T-cell epitopes in a viral protein and β-cell autoantigens, or activation of endogenous retroviruses in β cells by environmental agents). If an exogenous virus were clearly identified, then protective vaccination early in life would be the logical approach to primary prevention. However, even if a particular enterovirus strain were shown directly to be diabetogenic, vaccination may be a challenge because of the many thousands of strain variants, the only one for which a vaccine exists being poliovirus. There is no scientific evidence that any form of vaccination triggers T1DA.

Rotavirus is the most common cause of gastroenteritis in young children. The discovery of strong sequence similarities between T-cell epitopes in the VP7 protein of rotavirus and the IA2 and GAD islet antigens in autoantibody-positive children led to speculation that molecular mimicry with rotavirus might contribute to islet autoimmunity (reviewed in Harrison, *et al.*[29]). Subsequently, in the Australian BabyDiab Study, rotavirus infection was associated over time with the first appearance of, or an increase in, islet autoantibodies in children before they developed diabetes.[29] Moreover, rotavirus has been shown to infect β cells in islets from mice, pigs, and monkeys and causes transient involution of the pancreas and hyperglycemia in a Toll-like receptor 3 (TLR3)–dependent manner in mice.[29] Cross-reactive immunity between ubiquitous rotaviruses and islet autoantigens may not be directly diabetogenic but could complement and sustain the immune response to direct infection of β cells. Evidence for rotavirus was strengthened by the finding that the incidence of T1DA in young children had decreased following the introduction of routine rotavirus vaccination in Australia in 2007.[29] This was confirmed in a larger study in the USA,[30] but requires further validation in other locations.

The microbiome—the trillions of microorganisms (bacteria, fungi, archaea, protozoa, and viruses; Chapter 22) and their millions of genes and proteins that reside within the human mucosae, skin, and secretions—has gained increasing attention as a bellwether of health and disease ("dysbiosis"). Most microbiome analyses are based on DNA extraction and polymerase chain reaction (PCR) amplification of regions of the 16S ribosomal RNA (rRNA) gene that are conserved among bacteria. Although other internal regions of the 16S gene are variable and enable taxonomic classification from the phylum to genus level, 16S sequencing does not have the sensitivity to distinguish individual species and their strains. Direct metagenomic sequencing

of all species, as well as analysis of their transcriptomes, epigenomes, and functions, will be required to fully comprehend the role of the microbiome in T1DA and other chronic immune–inflammatory diseases. The marked difference in the incidence of T1DA between Finland and neighboring Karelia was associated in Finnish children with a decrease in gut bacterial diversity, dominance of the phylum Bacteroidetes, and relative deficiency of butyrate-producing bacteria.[25] These changes were seen after the appearance of autoantibodies, suggesting that they followed, rather than preceded, the disease process. However, in a further small study in Finnish children, metagenomic sequencing identified an increase in the abundance of *Bacteroides dorei*, which peaked around 7 to 8 months of age with the introduction of solids and preceded the appearance of islet autoantibodies. Gut *Bacteroides* species are abundant in Finnish children, including *B. dorei* that produces a lipopolysaccharide (LPS) endotoxin, which inhibits the immunostimulatory activity of *Escherichia coli* LPS and protects against diabetes in NOD mice. These findings suggest that an approach to lowering the environmental contribution to T1DA may be through intervention with prebiotics, probiotics, or other means that favorably alter the composition of the gut microbiome.

KEY CONCEPTS

Type 1A Diabetes: Epidemiology, Environment, and Genes

- Incidence is increasing due to the impact of a changing environment.
- Environment factors enable increasing penetrance of lower-risk human leukocyte antigen alleles.
- Environmental factors are multifactorial but generally "pro-inflammatory" and impact via microbiome dysbiosis and epigenetic modifications on gene expression to cause immune dysregulation in early life.

TREATMENT AND PREVENTION

Treatment of the metabolic syndrome of T1D is focused on optimizing blood glucose control with various modes of insulin delivery to prevent the short- and long-term complications of hyperglycemia. Over the past 40 years, the quality of life and prognosis of individuals with T1DA have greatly improved due to the introduction of pure, recombinant forms of human insulin injected by syringe or continuously via pump, blood glucose self-monitoring, and better blood pressure control. However, individuals with T1DA have not been rescued from a life sentence, and their blood glucose control remains less than physiological. The "cure" of T1DA requires transplantation of insulin-secreting cells or their progenitors or the regeneration of β cells *in situ*, in conjunction with approaches to prevent allograft rejection and/or recurrence of autoimmunity. The cure for T1DA also requires prevention. Currently, in some economically developed societies, allografts of whole pancreas or isolated islets are offered on a limited scale and at considerable cost to individuals with T1DA who suffer life-threatening and uncontrollable complications such as hypoglycemia "unawareness". However, shortage of donor tissue, cost, and the need for life-long immune suppressive agents militate against this approach. Longer-term, genetically engineered pig islet xenografts may overcome the tissue supply barrier. Stem cells, in particular autologous induced pluripotent stem cells (iPSCs), remain the great hope—but scale-up from proof-of-concept in rodents to translation in humans remains a challenge.

The incidence of spontaneous autoimmune diabetes in inbred NOD mice is decreased by many immune and other (environmental) interventions early in the disease course, although most are only effective in a proportion of mice and delay disease onset. Nevertheless, the findings in the NOD mouse suggest that T1DA may be preventable in outbred humans, especially with intervention before the onset of islet autoimmunity or very soon thereafter. Prevention of T1DA can be classified as primary (before onset of islet autoimmunity), secondary (after onset of islet autoimmunity), and tertiary (after onset of clinical disease) (see Fig. 71.3). Neonatal screening based on HLA genotyping to identify genetically at-risk individuals is a basis for primary prevention; but even in countries with a high incidence of T1DA, this still would have only modest predictive value and would need to be at least practical and safe, if not efficacious, to justify an intervention. Immune modifying agents that are effective when used at a relatively earlier stage in other autoimmune diseases are less likely to be effective in Stage 3 T1DA, after clinical onset, when most β cells have been destroyed. Indeed, most clinical trials of tertiary prevention, with more than 70 different agents since the early 1980s, have failed to demonstrate a sustained effect to preserve residual β-cell function, with several notable recent exceptions (Table 71.3).[31,32] Even so, the decline in C-peptide secretion was not slowed indefinitely in individuals with stage 3 disease. However, the more recent finding that the onset of clinical disease in individuals with stage 2 disease treated with the anti-CD3 monoclonal antibody, teplizumab, was delayed by 2 years (48.4 months compared to 24.4 months in placebo controls) indicates that immunotherapeutic intervention is likely to succeed if administered early in pre-clinical T1DA.[32] Accepting the paradigm shift that T1DA is primarily an autoimmune β-cell disorder could profoundly change the approach to its prevention. A further benefit of diagnosing T1DA early in the asymptomatic stage, based on evidence of underlying pathology, is that it markedly reduces the risk of life-threatening ketoacidosis associated with the classic symptomatic presentation.

Based on the key role of proinsulin in driving β-cell autoimmunity, the NOD mouse has provided "proof-of-concept" for antigen-specific vaccination strategies.[33] In health, immune responses to autoantigens are regulated to prevent development of autoimmune diseases. Autoantigen-specific immunotherapy aims to boost or restore autoantigen-specific immunoregulatory mechanisms. Allergen-specific immunotherapy has been shown in randomized trials to be effective in ameliorating allergic asthma and rhinitis. Such "negative vaccination" can be achieved in several ways: by administering antigen via a "tolerogenic" route (mucosal, dermal), cell type (resting dendritic cell), mode (with blockade of costimulation molecules), or form (as an "altered peptide

TABLE 71.3 Agents Shown in Randomized Trials to Preserve β-Cell Function in Stage 3 T1DA*

Anti-CD3 monoclonal antibody (OKT3)—teplizumab or otelixizumab
Anti-CD20 monoclonal antibody—rituximab
CTLA-4 Ig—abatacept
Anti-thymocyte globulin
Anti-CD2 monoclonal antibody—alefacept

*Following demonstration of its effect to preserve C-peptide secretion in individuals with stage 3 T1D, teplizumab was shown to delay the onset of clinical disease by a median of 2 years in individuals with stage 2 disease.

ligand"). Mechanisms of antigen-induced tolerance include deletion and/or anergy of effector T cells and induction of regulatory T cells (iTregs). Of clinical importance is the ability of iTregs to exert antigen-nonspecific "bystander" suppression. Thus, by direct cell contact and/or the release of soluble immunosuppressive factors, such as IL-10, iTregs impair the function of antigen-presenting DCs to elicit effector T-cell responses to the same or other antigens presented locally in the lesion or draining lymph nodes. Bystander suppression does not require that the "tolerizing" antigen is necessarily the primary driver of pathology. Its clinical importance is that it obviates specificity restrictions imposed by polymorphisms in the HLA and human T-cell receptors.

In NOD mice, administration of insulin, proinsulin peptides, or proinsulin DNA via oral or naso-respiratory routes, acting locally on the mucosal immune system, induces Tregs and decreases the incidence of diabetes. Results of randomized trials of insulin or GAD in relatives at risk for T1DA (ClinicalTrials.gov) have been summarized elsewhere.[31–33] In the DPT-1 oral insulin trial, islet autoantibody-positive relatives with a 5-year diabetes risk of 25% to 50% received 7.5 mg human insulin or placebo daily for a median of 4.3 years. There was no effect overall, but posttrial hypothesis testing revealed a 4-year delay in diabetes onset in participants with significant IAA at entry. This outcome led to a follow-up international trial by TrialNet in islet autoantibody-positive relatives using the same dose of 7.5 mg daily, which is not optimal because on a body-weight basis it is very small compared with the dose required to induce protective iTregs in the NOD mouse. Again, there was no effect of oral insulin overall, but a significant delay in clinical diabetes onset in a sub-group of participants with more marked loss of insulin secretory function.

Apart from dose, other variables have not been systematically tested in humans, in part because of the expense and duration of prevention trials. These variables include route of administration, form of the antigen, combinations of antigen with antigen-nonspecific agents, and the nature of induced immune responses. Unfortunately, surrogate markers of a potential therapeutic response have not been included in most trials. Oral delivery might not be optimal because proteins are degraded after ingestion, and the concentration or form of peptide reaching the upper small intestine is variable and unpredictable. T-cell responses observed after naso-respiratory administration of a peptide are not observed after oral administration. In a randomized trial of nasal insulin in individuals with recent-onset T1D not initially requiring insulin treatment, participants in the nasal insulin arm had markedly blunted insulin antibody responses after subsequent subcutaneous insulin.[34] This was the first demonstration, in humans, of immune regulation induced by mucosally administered exogenous autoantigen. Although this result is a rationale for further disease endpoint trials, a Finnish study[35] found no evidence that nasal insulin (1 unit/kg daily) altered the rate of progression to diabetes in islet-autoantibody-positive children <3 years of age. However, these children were at very high risk, and many appeared to have had borderline β-cell function. As in the oral insulin trial, markers of an insulin bioeffect and evidence of immune tolerance were not reported. In the Intranasal Insulin Trial II (INIT II) in Australia, New Zealand, and Germany (unpublished), higher doses of nasal insulin (44 or 440 IU) were administered weekly for a year to T1DA relatives who had autoantibodies to at least two islet antigens (≈40% risk of diabetes over 5 years). Nasal insulin was associated with evidence of immune tolerance to insulin, but this did not translate into protection from clinical T1DA. The promise of antigen-specific therapy, therefore, has yet to be realized in humans. If a balance between pathogenic and protective T cells is deterministic for disease development, then antigen-specific therapy is most likely to be effective as a primary preventive strategy. Once disease has been initiated, insulin- or other islet-antigen-specific approaches may be applicable as complementary therapy with immune-modifying agents that inactivate or delete the burden of pathogenic effector cells.

ON THE HORIZON

- Closed-loop insulin delivery—blood glucose monitoring devices will become more refined, cheaper, and widely available.
- "Linking the -omes" (exposome–microbiome–metabolome–epigenome) across time in early life will provide new insights into how environment–gene interactions lead to immune dysregulation in type 1A diabetes (T1DA).
- Manipulation of the gut microbiome via scientifically formulated probiotics and prebiotics will be explored for primary prevention of T1DA.
- Antigen-specific vaccination, *e.g.*, with insulin, will be applied for the primary prevention of T1DA.
- Success in delaying the onset of stage 3 clinical T1DA by treatment of high-risk stage 2 individuals with anti-CD3 monoclonal antibody heralds a new era of T1DA prevention trials.
- Recognition that T1DA is primarily an immune disease, an autoimmune β-cell disorder, and only secondarily a metabolic disorder will accelerate the application of immune modifying therapies to the prevention of clinical T1DA.

REFERENCES

1. The Expert Committee on the Diagnosis and Classification of Diabetes Mellitus Report of the expert committee on the diagnosis and classification of diabetes mellitus. *Diabetes Care*. 2003;26:S5–S20.
2. Verge CF, Gianani R, Kawasaki E, et al. Prediction of type I diabetes in first-degree relatives using a combination of insulin, GAD, and ICA512bdc/IA-2 autoantibodies. *Diabetes*. 1996;45:926–933.
3. Colman PG, Steele C, Couper JJ, et al. Islet autoimmunity in infants with a type I diabetic relative is common but is frequently restricted to one autoantibody. *Diabetologia*. 2000;43:203–209.
4. Bingley PJ, Williams AJ, Gale EA, et al. Optimized autoantibody-based risk assessment in family members. Implications for future intervention trials. *Diabetes Care*. 1999;22:1796–1801.
5. Ziegler AG, Rewers M, Simell O, et al. Seroconversion to multiple islet antibodies and risk of progression to diabetes in children. *JAMA*. 2013;309:2473–2479.
6. Bonifacio E, Mathieu C, Nepom G, et al. Rebranding type 1 diabetes: the case for autoimmune beta cell disorder as a pathologic and diagnostic entity. A consensus statement from the type 1 diabetes Iceland Summit. *Diabetologia*. 2016;60:35–38.
7. Insel RA, Dunne JL, Atkinson MA, et al. Staging presymptomatic type 1 diabetes: a scientific statement of JDRF, the Endocrine Society and the American Diabetes Association. *Diabetes Care*. 2015;38:1964–1974.
8. Wentworth JM, Bediaga NG, Gitelman SE, et al. Clinical trial data validate the C-peptide estimate model in type 1 diabetes. *Diabetologia*. 2020;63:885–886.
9. Wentworth JM, Fourlanos S, Harrison LC. Deconstructing the stereotypes of diabetes within the modern diabetogenic environment. *Nat Rev Endocrinol*. 2009;5:483–489.
10. Balasubramanyam A, Nalini R, Hampe CS, et al. Syndromes of ketosis-prone diabetes mellitus. *Endocr Rev*. 2008;29:292–302.
11. Imagawa A, Hanafusa T, Miyagawa J, et al. A novel subtype of type 1 diabetes mellitus characterized by a rapid onset and an absence of diabetes-related antibodies. Osaka IDDM Study Group. *N Engl J Med*. 2000;342:301–307.
12. Imagawa A, Hanafusa T. Pathogenesis of fulminant type 1 diabetes. *Rev Diabet Stud*. 2006;3:169–177.

13. Harrison L, Honeyman M, DeAizpurua H, et al. Inverse relationship between humoral and cellular immunity to glutamic acid decarboxylase in humans at-risk for insulin-dependent diabetes. *Lancet*. 1993;341:1365–1369.
14. Gale EAM. The rise of childhood type 1 diabetes in the 20th century. *Diabetes*. 2002;51:3353–3361.
15. Fourlanos S, Varney MD, Tait BD, et al. The rising incidence of type 1 diabetes is accounted for by cases with lower risk human leukocyte antigen genotypes. *Diabetes Care*. 2008;31:1546–1549.
16. Tait BD, Colman PG, Morahan G, et al. HLA genes associated with autoimmunity and progression to disease in type 1 diabetes. *Tissue Antigens*. 2003;61:146–153.
17. Fourlanos S, Narendran P, Byrnes GB, et al. Insulin resistance is a risk factor for progression to type 1 diabetes. *Diabetologia*. 2004;47:1661–1667.
18. Penno MAS, Couper JJ, Craig ME, et al. Environmental determinants of islet autoimmunity (ENDIA): a pregnancy to early life cohort study in children at-risk of type 1 diabetes. *BMC Pediatr*. 2013;13:124.
19. Willcox A, Richardson SJ, Bone AJ, et al. Analysis of islet inflammation in human type 1 diabetes. *Clin Exp Immunol*. 2009;155:173–181.
20. Barrett JC, Clayton DG, Concannon P, et al. Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. *Nat Genet*. 2009;41:703–707.
21. Pugliese A, Zeller M, Fernandez Jr A, et al. The insulin gene is transcribed in the human thymus and transcription levels correlated with allelic variation at the INS VNTR-IDDM2 susceptibility locus for type 1 diabetes. *Nat Genet*. 1997;15:293–297.
22. Bennett ST, Lucassen AM, Gough SC, et al. Susceptibility to human type 1 diabetes at IDDM2 is determined by tandem repeat variation at the insulin gene minisatellite locus. *Nat Genet*. 1995;9:284–292.
23. Atkinson MA, Bluestone JA, Eisenbarth GS, et al. How does type 1 diabetes develop?: the notion of homicide or beta-cell suicide revisited. *Diabetes*. 2011;60:1370–1379.
24. Couper JJ, Beresford S, Hirte C, et al. Weight gain in early life predicts risk of islet autoimmunity in children with a first degree relative with type 1 diabetes. *Diabetes Care*. 2009;32:94–99.
25. Knip M, Siljander H. The role of the intestinal microbiota in type 1 diabetes mellitus. *Nat Rev Endocrinol*. 2016;12:154–167.
26. Thorburn AN, Macia L, Mackay CR. Diet, metabolites, and "Western lifestyle" inflammatory diseases. *Immunity*. 2014;40:833–842.
27. Kondrashova A, Seiskari T, Ilonen J, et al. The 'hygiene hypothesis' and the sharp gradient in the incidence of autoimmune and allergic diseases between Russian Karelia and Finland. *APMIS*. 2013;121:478–493.
28. Op de Beeck A, Eizirik D. Viral infections in type 1 diabetes mellitus--why the β cells? *Nat Rev Endocrinol*. 2016;12:263–273.
29. Harrison LC, Perrett KP, Jachno K, et al. Does rotavirus turn on type 1 diabetes? *PLoS Pathog*. 2019;15(10):e1007965. https://doi.org/10.1371/journal.ppat.1007965.
30. Rogers MA, Basu T, Kim C. Lower incidence rate of type 1 diabetes after receipt of the rotavirus vaccine in the United States, 2001–2017. *Sci Rep*. 2019;9:7727.
31. Harrison LC, Wentworth JM. Prevention of autoimmune disease: the type 1 diabetes paradigm. In: Rose N, Mackay ID, eds. *The Autoimmune Diseases*. 6th ed. London, UK: Elsevier; 2018. chap 70.
32. Lord S, Greenbaum CJ. Insulin is necessary but not sufficient: changing the therapeutic paradigm in type 1 diabetes. *F1000Research*. 2020;9:827. https://doi.org/10.12688/f1000research.21801.1.
33. Harrison LC, Wentworth JM, Zhang Y, et al. Antigen-based vaccination and prevention of type 1 diabetes. *Curr Diab Rep*. 2013;13:616–623.
34. Fourlanos S, Perry C, Gellert SA, et al. Evidence that nasal insulin induces immune tolerance to insulin in adults with autoimmune diabetes. *Diabetes*. 2011;60:1237–1245.
35. Nanto-Salonen K, Kupila A, Simell S, et al. Nasal insulin to prevent type 1 diabetes in children with HLA genotypes and autoantibodies conferring increased risk of disease: a double-blind, randomized controlled trial. *Lancet*. 2008;372:1746–1755.

Immunologic Lung Diseases

Scott M. Matson, Joyce S. Lee, and Andrew P. Fontenot

The lung serves as an interface between the environment and the sanctuary of the body. The defense systems of the upper airways clear the majority of inhaled particulates. Those particulates that evade the upper-airway defenses are combated by the innate and acquired immune responses. Essentially all autoimmune diseases arise from the inappropriate activation of autoreactive CD4 T cells, as well as autoreactive B cells responsible for pathogenic autoantibodies. Immunologic lung diseases develop when the normal mechanisms of immune self-tolerance fail. This chapter considers the pulmonary manifestations of these disorders.

INFLAMMATION OF INTERSTITIAL LUNG DISEASES

In the normal host, the resident alveolar macrophage is the predominant cell type in fluid from the bronchoalveolar lavage (BAL). Resident alveolar macrophages function to ingest and degrade the inhaled antigenic load, clear pulmonary surfactant, and tonically suppress the development of inappropriate immune responses. Relatively few lymphocytes are present in the normal lung parenchyma. However, after stimulation by the relevant antigen in the lung-draining lymph nodes, antigen-specific lymphocytes migrate to the lung and participate in inflammatory responses. In addition to lymphocytes, other inflammatory and immune cells accumulate in the lung of patients with immunologic lung disease, including neutrophils, eosinophils, and other mononuclear cells, depending on the underlying disease. Within normal alveoli, the major cellular constituents are alveolar macrophages and epithelial cells. Resident dendritic cells also project dendrites through the tight junctions of the alveolar epithelium to sample the alveolar space. An initial insult typically involves the type I alveolar epithelial cell, which results in release of chemokines that recruit and activate inflammatory cells, allowing for resolution of inflammation and repair of injured tissue (Fig. 72.1). With either prolonged exposure or failure to adequately clear an inhaled antigen, persistent inflammation results in extracellular matrix deposition, culminating in tissue remodeling, progressive collagen deposition, and pulmonary fibrosis. Impaired ventilation and gas exchange occur as a consequence of lung fibrosis, resulting in patient morbidity and mortality. With destruction of type I alveolar epithelial cells, exposure of the underlying basement membrane can cause further inflammation. Proper restoration of the epithelial barrier is critical for resolution of lung inflammation. Type II alveolar epithelial cells can serve as progenitor cells that migrate and differentiate into type I alveolar epithelial cells to re-establish an intact lung epithelium.

Macrophages and lymphocytes have also been localized to areas of pulmonary fibrosis in patients with immunologic lung diseases including idiopathic pulmonary fibrosis (IPF), systemic sclerosis, and rheumatoid arthritis.[1–3] In the lung, macrophages can be divided by their location into alveolar or interstitial macrophages. Resident alveolar macrophages are dependent on several cytokines for their development (such as interleukin-34 [IL-34], colony stimulating factor 1, and the transcription factor PU.1).[1–3] These yolk-sac-derived macrophages are self-renewing,[2,3] maintained at a stable number throughout life, and are localized on epithelial surfaces and lining fluid of the alveoli and airways. With relatively poor antigen-presenting ability, they remove inhaled particles and bacteria. Conversely, interstitial macrophages are derived from hematopoietic stem cells and are located in the tissue spaces between the alveoli.[4] In the context of lung injury, interstitial macrophages greatly increase in number. Although interstitial macrophages have less phagocytic activity than alveolar macrophages, they have increased ability to present antigens to T cells. Upon activation, these macrophages express a variety of cytokines, including tumor necrosis factor (TNF) and monocyte chemotactic protein-1 (MCP-1). In addition to antigen presentation to T cells, macrophages are important for lung fibrosis and tissue remodeling in immunologic lung diseases through secretion of specific growth factors such as transforming growth factor-β (TGF-β), insulin-like growth factor-1 (IGF-1), platelet-derived growth factor (PDGF), and fibroblast growth factor (FGF).

T cells are also associated with lung fibrosis. For example, T cells have been located in areas of interstitial fibrosis and honeycombing with relative sparing in areas of normal lung in patients with IPF.[5] In connective tissue disease-associated interstitial lung disease (ILD), T cells are diffusely distributed throughout the lung and within focal lymphoid aggregates.[6] In animal models, depending on their phenotype, T cells can be either pro-fibrotic or anti-fibrotic. CD8 T cells have been found in high proportion in BAL and surgical lung biopsy samples from patients with IPF, as well as patients with systemic sclerosis, but their role in these immunologic lung diseases is not well understood.

Different subsets of CD4 T cells (Chapter 11) have been implicated in the pathogenesis of immunologic lung diseases. T-helper cell-1 (Th1) cells express interferon-gamma (IFN-γ). Although IFN-γ is a potent pro-inflammatory cytokine, it has anti-fibrotic [t.o] effects through inhibiting fibroblast proliferation and collagen expression. Th2 cells are defined by expression of IL-4, IL-5, and IL-13. In contrast to IFN-γ, Th2 cytokines have been shown to promote lung fibrosis. Therefore, the

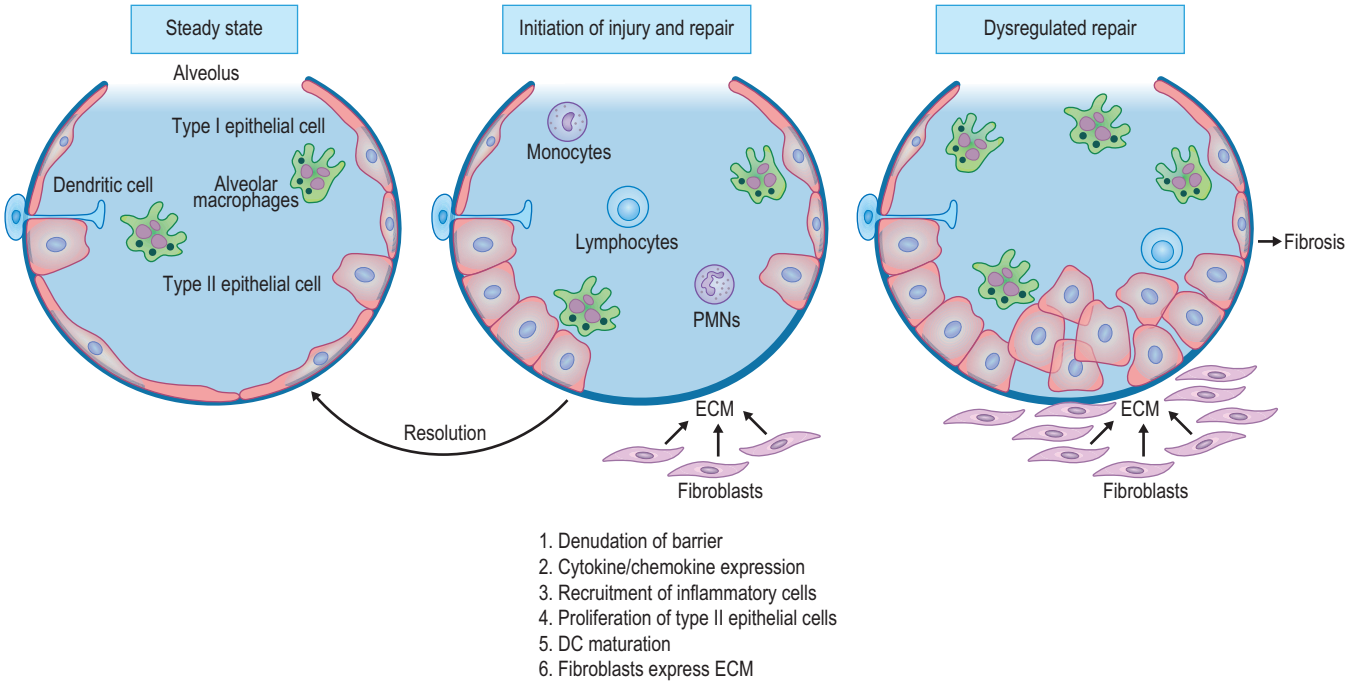


FIG. 72.1 Pathogenesis of Interstitial Lung Disease. In the healthy lung, the alveoli maintain an anti-inflammatory state to prevent unwanted inflammation. The predominant cells in the healthy alveolus are macrophages and types I and II epithelial cells. Upon injury, there is denudation of the epithelial barrier, resulting in the expression of cytokines/chemokines, recruitment of inflammatory cells, proliferation of type II epithelial cells, dendritic cell *(DC)* maturation, and expression of extracellular matrix *(ECM)* by fibroblasts. This inflammatory milieu leads to the resolution of inflammation and repair of lung injury. Conversely, in the presence of either prolonged antigen exposure or an inability to clear antigen, persistent inflammation ensues, resulting in extracellular matrix deposition with subsequent tissue remodeling, fibrosis, and permanent lung dysfunction. *PMNs*, Polymorphonuclear neutrophils.

balance between Th1/Th2 T cells through expression of different cytokines affects the development of pulmonary fibrosis.[7] Th17 cells express IL-17A and -17F, which are potent inflammatory cytokines important for the recruitment of neutrophils to areas of inflammation. Th17 cells have been implicated in the development of lung fibrosis in murine models.[8] Conversely, regulatory T cells (Tregs) suppress pathogenic T-cell responses that promote inflammation (Chapter 13). In patients with IPF, Tregs may be less able to suppress expression of cytokines by Th1 and Th2 cells, suggesting that Tregs are important in regulating inflammatory lung disease and pulmonary fibrosis.[9]

IDIOPATHIC INTERSTITIAL PNEUMONIAS

The idiopathic interstitial pneumonias (IIPs) are a group of diffuse inflammatory and/or fibrotic lung disorders that include idiopathic pulmonary fibrosis (IPF), acute interstitial pneumonitis (AIP), desquamative interstitial pneumonitis (DIP), respiratory bronchiolitis-associated interstitial lung disease (RB-ILD), nonspecific interstitial pneumonitis (NSIP), cryptogenic organizing pneumonia (COP), lymphocytic interstitial pneumonia (LIP), and idiopathic pleuroparenchymal fibroelastosis (IPPFE).[10] The diagnosis of IIPs requires the exclusion of connective tissue diseases (CTDs), drug toxicity, and environmental exposures, and a thoracoscopic lung biopsy may be required. In addition, some patients with IIPs have clinical features suggesting an underlying autoimmune disorder. However, these subjects do not meet established criteria for a CTD. A joint European Respiratory Society and American Thoracic Society task force has proposed the term "interstitial pneumonia with autoimmune features" (IPAF), as well as clinical classification criteria for affected subjects.[11] Importantly, the distinction between these various disorders is of clinical relevance since response to treatment and outcomes may differ.

Idiopathic Pulmonary Fibrosis

IPF is the most common diffuse idiopathic parenchymal lung disease. Despite the nomenclature, idiopathic pulmonary fibrosis has known genetic factors that increase the risk for the development of IPF.[12,13] IPF is characterized by progressive clinical deterioration despite available therapy. Although IPF has characteristic clinical, radiographic, and histologic appearances, other interstitial lung diseases, including the CTDs, drug reactions, and environmental exposures, can mimic these findings.

Clinical Manifestations

CLINICAL PEARLS

Idiopathic Pulmonary Fibrosis

- IPF is one of the most common causes of diffuse parenchymal lung disease, of unknown etiology, and is characterized by insidious onset of cough and dyspnea.
- The histopathologic pattern of IPF is usual interstitial pneumonitis.
- A confident diagnosis of IPF based on high-resolution computed tomography alone can only be made in one-third of cases.
- IPF is generally a fatal disorder, characterized by relentless progression and a 5-year survival of 30%–50%.
- Pirfenidone and nintedanib appear to slow disease progression in IPF.

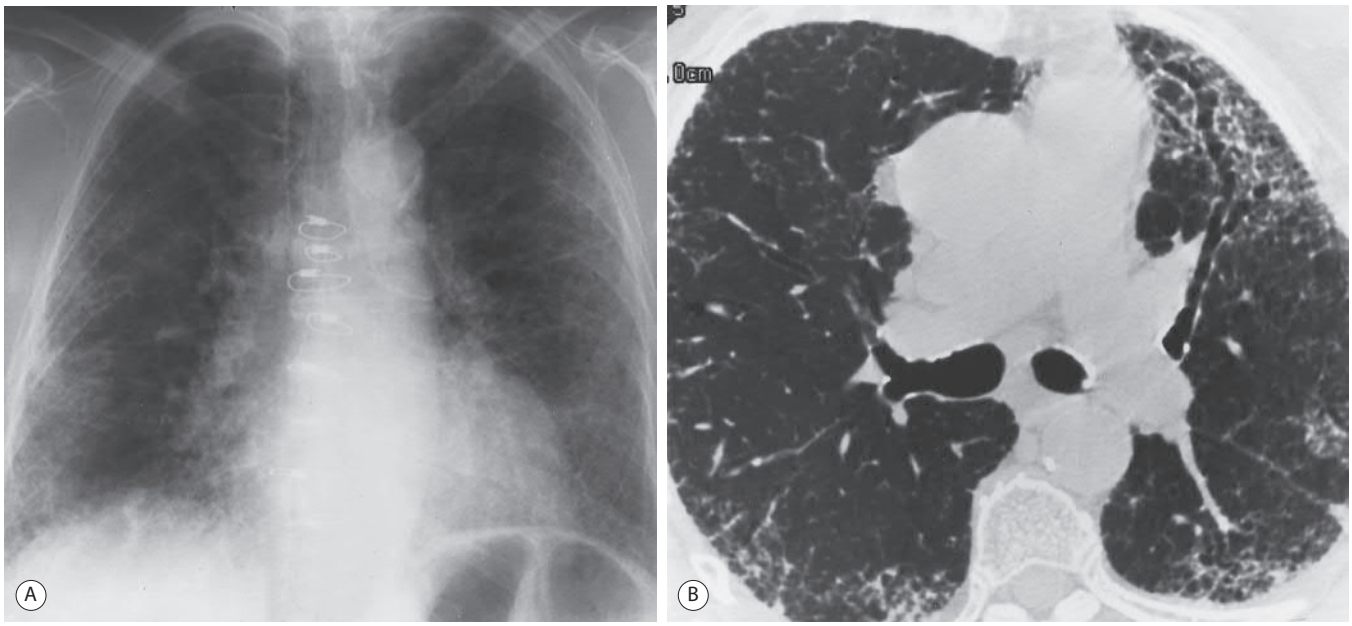


FIG. 72.2 Radiographic Manifestations of Idiopathic Pulmonary Fibrosis. (A) Chest radiograph in a patient with idiopathic pulmonary fibrosis showing diffuse, coarse reticular opacities with a lower lung zone predominance. Cystic radiolucencies, consistent with honeycombing, are evident. (B) High-resolution computed tomography shows peripheral reticular opacities, honeycombing, and traction bronchiectasis.

The exact incidence and prevalence of IPF are uncertain, although the incidence in North America has been reported between 2.8 and 19 cases per 100,000 people per year.[14] Both the incidence and prevalence of IPF increase with age, with most patients presenting between 50 and 70 years of age. Most patients present with the insidious onset of exertional dyspnea and a dry, nonproductive cough. Physical examination reveals dry, end-inspiratory crackles, with digital clubbing being common.

The chest radiograph typically shows diffuse reticular opacities, predominantly in the peripheral lower lung zones. High-resolution computed tomography (HRCT) findings include peripheral and basilar predominant and subpleural reticular abnormalities, cysts in the periphery of the lung (honeycombing), and traction bronchiectasis in addition to a paucity of ground-glass opacities (Fig. 72.2, A and B).[15] These radiographic changes often precede the onset of symptoms, and serial chest imaging usually reveals progressive loss of lung volume.

The typical physiologic abnormalities in IPF are those of a restrictive lung disease with a low diffusing capacity for carbon monoxide (DLCO) and severe gas exchange abnormalities exacerbated by exercise.

Histopathology

The gross appearance of the lungs in IPF shows a nodular pleural surface while histopathologic examination reveals usual interstitial pneumonitis (UIP).[16] UIP is characterized by nonuniform and variable distribution of the interstitial changes. At low magnification, alternating zones of interstitial fibrosis, inflammation, honeycombing, and normal lung can be seen (Fig. 72.3, A). At higher magnification, derangement of alveolar walls with fibroblast proliferation and fibrosis occur. Honeycomb change refers to enlarged airspaces lined by metaplastic bronchial epithelium and surrounded by walls thickened with collagen (see Fig. 72.3, B). The earliest finding in UIP is the fibroblast focus, a lesion consisting of distinct clusters of fibroblasts and myofibroblasts in a loose connective tissue matrix within the alveolar wall, with minimal interstitial inflammation or intra-alveolar macrophage accumulation (see Fig. 72.3, C).[16] Tables 72.1 and 72.2 compare the clinical and pathologic features of UIP, DIP, RB-ILD, and NSIP.

Pathogenesis

The pathogenesis of IPF is poorly understood, but current evidence suggests that fibrosis results from aberrant wound healing in response to repetitive injury and is mediated through an interplay between immunologic, genetic, and environmental factors (Fig. 72.4).[12] Some cases of IPF are familial, inherited as an autosomal dominant trait with variable penetrance. Dysregulated expression of a mucin gene, *MUC5B*, has been associated with development of familial forms of IPF and confers the greatest genetic risk for the development of IPF.[13] Mutations in the telomerase ribonucleoprotein complex associated with telomere shortening have also been linked with familial interstitial pneumonia.[12]

In the normal lung, the interstitium is thin and delicate with few lymphoid cells and fibroblasts. Following initiation of the inflammatory process, damage to the alveolar epithelial cell occurs, followed by vascular leak, fibroblast activation and proliferation, extracellular matrix synthesis, and activation of the innate immune system.[12] The release of danger-associated molecular patterns from dead or dying cells results in macrophage activation. Following activation, alveolar macrophages secrete IL-1, IL-8, TNF, PDGF, and IGF-1. This cytokine milieu promotes the activation and recruitment of neutrophils and lymphocytes to the area of alveolitis.

T lymphocytes, which accumulate in the alveolar space and interstitium, express an activated phenotype, including the

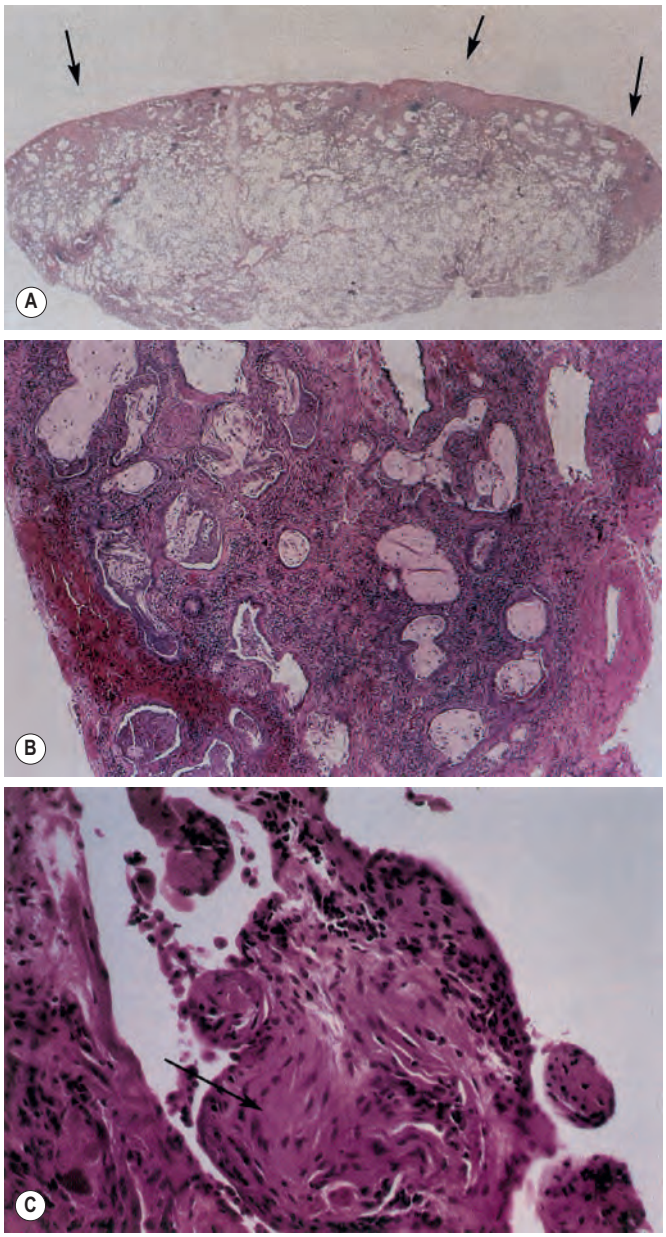


FIG. 72.3 Histopathology of Usual Interstitial Pneumonitis (UIP). (A) Low-magnification photomicrograph of UIP showing the variegated appearance from one field of view to the next, with areas of dense subpleural fibrosis *(arrows)* separated from other areas of normal lung. (B) High-magnification photomicrograph of UIP showing honeycomb change characterized by enlarged airspaces filled with mucin and separated by fibrosis. (C) Fibroblast focus in UIP is characterized by clusters of spindle-shaped fibroblasts *(arrow)* in a loose connective tissue matrix within the alveolar wall.

expression of human leukocyte antigen (HLA)-DR and IL-2 receptors. Following activation, CD4 T cells evolve into three major subsets distinguished by the cytokines produced (Chapters 11 and 14). In IPF, T cells expressing a Th2-type phenotype predominate, producing IL-4, -5, and -13. In addition, the Th17 cytokine, IL-17A, has been linked to the development of bleomycin-induced lung injury and collagen deposition.[17,18] Evidence also suggests that IPF patients have oligoclonal CD4 T-cell expansions that proliferate in response to antigens present in diseased tissue. Regulatory T-cell function may be impaired in patients with IPF.[9] In addition, immune complexes have been identified in the serum and lungs of IPF patients.[19]

In addition to their role as scavengers, alveolar macrophages are vital in the repair phase of inflammation. However, the distinguishing feature between a self-resolving inflammatory process and a fibrotic response, as seen in IPF, is the accumulation of collagen. Evidence suggests that the fibrotic process in IPF is a consequence of dysregulation of both collagen synthesis and degradation. Macrophage-derived growth factors, including TGF-β, PDGF, and IGF-1, stimulate fibroblast proliferation and collagen deposition. Adequate resolution of an inflammatory process requires matrix degradation. Matrix metalloproteases produced by macrophages and fibroblasts are involved in matrix degradation, and control of metalloprotease production involves substances known as tissue inhibitors of metalloproteases (TIMPs). TIMPs are elevated in the lungs of patients with IPF. In addition, TGF-β can markedly augment TIMP production.[20] Thus, there appears to be a loss of balance between the events mediating resolution and those mediating perpetuation of the inflammatory response, setting the stage for lung injury, tissue remodeling, and the development of irreversible pulmonary fibrosis.

KEY CONCEPTS

Pathogenesis of the Idiopathic Interstitial Pneumonias

- Although the inciting event(s) is unknown in the different diseases, a common result is a dysregulated fibroproliferative response (similar to wound healing), which leads to excessive extracellular matrix production and lung remodeling.
- A genetically determined inability to repair and re-epithelialize the denuded basement membranes adequately may be a contributing factor and may relate to the familial occurrence of some cases of idiopathic pulmonary fibrosis.
- The presence of a chronic stimulus (autoantigen), as is seen in the pneumoconioses, may result in a persistent inflammatory and immune response and lead to a failure in the normal healing process.
- The release of transforming growth factor-β following epithelial injury stimulates collagen synthesis and the prevention of apoptosis of proliferating fibroblasts in the lung, and may impair collagen degradation by inhibiting the production of metalloproteases.
- A predominant Th2 response in the lung and the absence of interferon-γ favor the development of a fibrotic response.

Diagnosis

The diagnostic evaluation of a patient with diffuse parenchymal lung disease includes a thorough history and physical examination, with particular attention to symptoms and signs that could indicate a CTD, occupational and environmental exposures, or medication and drug usage. A careful family history is also important.

The history and physical findings in IPF are nonspecific. However, extrapulmonary involvement does not occur: the presence of fever, arthralgias, myalgias, or pleuritis should suggest an alternative diagnosis. Circulating autoantibodies associated with CTD (anti-cyclic citrullinated peptide [CCP], rheumatoid factor [RF], and antinuclear antibodies [ANAs])

TABLE 72.1 Clinical Features of Selected Idiopathic Interstitial Pneumonias

	IPF	DIP	RB-ILD	AIP	NSIP
Mean age (years)	70	42	36	49	49
Childhood	No	Rare	No	Rare	Occasionally
Onset	Insidious	Insidious	Insidious	Acute	Subacute, insidious
Mortality (mean survival)	68% (5–6 years)	27% (12 years)	0%	62% (1–2 months)	11% (17 months)
Response to steroids	Poor	Good	Good	Poor	Good
Recovery possible	No	Yes	Yes	Yes	Yes

AIP, Acute interstitial pneumonitis; *DIP*, desquamative interstitial pneumonitis; *IPF*, idiopathic pulmonary fibrosis; *NSIP*, nonspecific interstitial pneumonitis; *RB-ILD*, respiratory bronchiolitis-associated interstitial lung disease.
Adapted from Katzenstein AL, Myers JL. Idiopathic pulmonary fibrosis: clinical relevance of pathologic classification. *Am J Respir Crit Care Med*, 1998;157:1301.

TABLE 72.2 Histopathologic Features of Selected Idiopathic Interstitial Pneumonias

	IPF/UIP	DIP/ RB-ILD	AIP	NSIP
Temporal appearance	Variegated	Uniform	Uniform	Uniform
Interstitial inflammation	Scant	Scant	No	Prominent
Collagen/ fibrosis	Patchy	Diffuse (DIP) Focal (RB-ILD)	No	Diffuse
Fibroblast proliferation	Prominent	No	Diffuse	Rare
Organizing pneumonia	No	No	No	Focal
Honeycomb change	Yes	No	No	Rare
Intra-alveolar macrophages	Focal	Diffuse (DIP) Focal (RB-ILD)	No	Patchy
Hyaline membranes	No	No	Focal	No

AIP, Acute interstitial pneumonitis; *DIP*, desquamative interstitial pneumonitis; *IPF*, idiopathic pulmonary fibrosis; *NSIP*, nonspecific interstitial pneumonitis; *RB-ILD*, respiratory bronchiolitis-associated interstitial lung disease; *UIP*, usual interstitial pneumonitis.
Adapted from Katzenstein AL, Myers JL. Idiopathic pulmonary fibrosis: clinical relevance of pathologic classification. *Am J Respir Crit Care Med.* 1998;157:1301.

are present in 20% of IPF patients; however, higher autoantibody titers and any systemic signs of an underlying CTD should prompt further evaluation for underlying CTD or IPAF.[11,19]

The majority of patients with IPF have an abnormal chest radiograph at the time of presentation. Basal peripheral reticular opacities are the characteristic radiographic findings. A confident diagnosis of IPF from HRCT of the lung requires the presence of patchy, peripheral bibasilar reticular abnormalities with honeycombing.[15] The presence of extensive ground-glass opacities on HRCT should suggest an alternative diagnosis, such as DIP, hypersensitivity pneumonitis, COP, or NSIP.

A surgical lung biopsy is recommended in suspected IPF patients without a definitive HRCT appearance and who do not have contraindications to the procedure. This is especially important in patients with atypical clinical or radiographic findings, which could suggest the possibility of one of the other histologic patterns of the idiopathic interstitial pneumonias and an improved prognosis or possible treatment divergence. Biopsy may be omitted in elderly patients with cardiovascular disease,

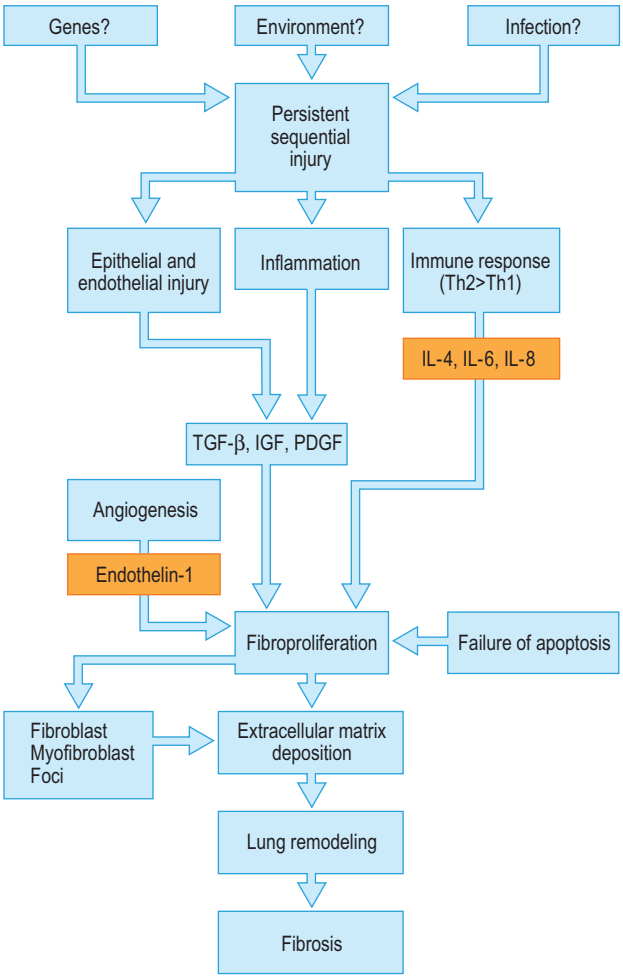


FIG. 72.4 Events Hypothesized to be Involved in the Pathogenesis of Idiopathic Pulmonary Fibrosis. The initiating event(s) leading to persistent lung injury remains poorly understood. The interaction between genetic factors, environmental exposures, and infectious agents leads to epithelial and endothelial injury, resulting in the secretion of macrophage-derived growth factors, including transforming growth factor-β *(TGF-β)*, insulin-like growth factor-1 *(IGF-1)*, and platelet-derived growth factor *(PDGF)*. This cytokine milieu stimulates fibroblast proliferation and collagen deposition. In addition, the resulting T-helper cell-2 (*Th2*) immune response stimulates extracellular matrix production and fibroblast proliferation, resulting in lung remodeling and, eventually, lung fibrosis. *IL*, Interleukin

or those with evidence of extensive honeycomb change. Video-assisted thoracoscopic (VATS) biopsy is the preferred surgical technique, as it is associated with less morbidity and a decreased hospital stay compared with open lung biopsy.

Treatment and Outcome

The usual course of IPF is relentless progression without spontaneous remission, commonly with a fatal outcome. The most common cause of death in patients with IPF is progression of the underlying disease, with two-thirds of deaths due to respiratory failure or cardiovascular complications. Other causes of death in IPF include bronchogenic carcinoma, infection, and pulmonary embolism. Recent studies in patients with biopsy-proven IPF indicate a poor prognosis (30% to 50% 5-year survival).[21] Previously, there was no evidence to support the use of any specific therapy in the management of IPF. However, clinical trials have shown decreased decline in forced vital capacity after use of either pirfenidone[22] or nintedanib.[23] Thus, for the first time, there are two Food and Drug Administration (FDA)-approved drugs for the treatment of IPF, with multiple other phase II and III clinical trials scheduled for completion in the near future. Finally, lung transplantation should be considered in patients with progressive clinical and physiologic deterioration who meet established criteria.

Acute Interstitial Pneumonia

Acute interstitial pneumonia (AIP) is a fulminant form of idiopathic interstitial pneumonia. Although it was previously thought to represent an acute phase of UIP, studies suggest it is a distinct entity.[16] However, patients with documented UIP/IPF experiencing acute exacerbations can have the pathology of AIP superimposed on UIP.[24]

Clinical Manifestations

AIP presents with the abrupt onset of dyspnea, followed by rapid progression to respiratory failure. The clinical, radiographic, physiologic, and histologic features are identical to those of the acute respiratory distress syndrome (ARDS) but without any identifiable cause. Most patients are previously healthy individuals over 40 years of age. Men and women are equally affected. A viral prodrome is common, with symptoms including fever, nonproductive cough, and dyspnea. Laboratory studies are nonspecific. Chest radiographs and HRCT show diffuse airspace opacities and ground-glass attenuation, respectively. A similar presentation may occur as the initial manifestation of a CTD.

Histopathology

AIP is characterized by diffuse interstitial fibrosis that is temporally uniform (Fig. 72.5).[25] The changes are identical to the organizing phases of diffuse alveolar damage, as seen in ARDS. Within the thickened interstitial space, there is active, diffuse fibroblast proliferation similar to the focal fibroblast foci seen in UIP. If this is progressive, honeycomb change occurs. Other features of acute lung injury, which are frequently seen in AIP, are intra-alveolar hyaline membranes.

Diagnosis

The diagnosis of AIP is based on a clinical syndrome of idiopathic ARDS and the presence of organizing diffuse alveolar damage on lung biopsy. Lung biopsy is occasionally performed to establish the diagnosis and exclude other causes of acute interstitial lung disease.

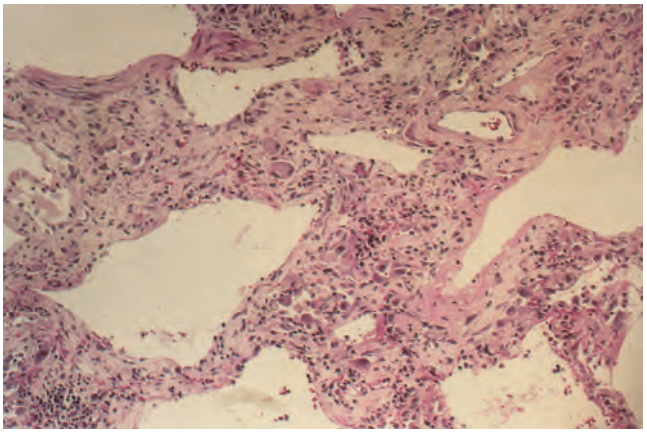


FIG. 72.5 Histopathology of Acute Interstitial Pneumonitis. Diffuse thickening of the alveolar septum with an infiltration of mononuclear cells is the characteristic abnormality. The temporal uniformity of this process is also apparent.

Treatment and Outcome

No effective therapy exists for patients with AIP. Glucocorticoids are utilized in most cases, but no survival benefit has been shown. Overall, the prognosis of patients with AIP is poor, with mortality rates ranging from 50% to 88%. Half of patients die within 6 months of disease onset. However, those who survive may have complete recovery of lung function, and AIP rarely recurs in survivors.

Desquamative Interstitial Pneumonitis

DIP represents fewer than 3% of all cases of interstitial lung disease.[26] However, it is a distinct clinicopathologic entity that differs substantially from UIP.

Clinical Manifestations

DIP affects individuals in their fourth to fifth decades of life with a male predominance. It predominantly occurs in cigarette smokers. Clinically, most individuals present with subacute onset of a dry, nonproductive cough and dyspnea. Clubbing is present in approximately 50% of DIP patients. Laboratory evaluation is usually nonspecific.

While the chest radiograph can be normal in up to 20% of symptomatic individuals, it typically shows nonspecific bibasilar ground-glass opacities. Reticulonodular interstitial infiltrates have also been reported. HRCT confirms the presence of ground-glass attenuation in the periphery of the lower lung zones (Fig. 72.6). Pulmonary function testing shows a restrictive defect with hypoxemia and a decrease in the DLCO.

Histopathology

DIP is a misnomer. It was initially thought that the intra-alveolar cells represented sloughed or desquamated alveolar epithelial cells. However, DIP is pathologically characterized by uniform, diffuse accumulation of macrophages in the alveolar space (Fig. 72.7). At low magnification, the overall appearance is one of uniformity from one field of view to the next, as opposed to the variegated appearance of UIP. In addition, there is scant interstitial inflammation, with varying degrees of fibrosis of the alveolar septum.

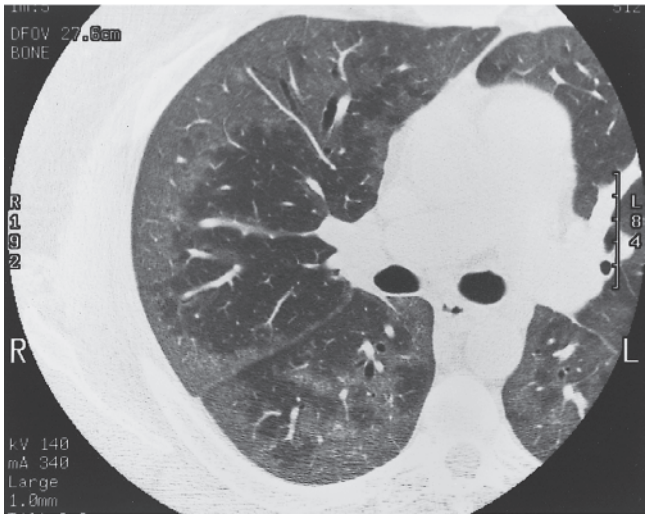


FIG. 72.6 Radiographic Manifestations in Desquamative Interstitial pneumonitis. High-resolution computed tomography in a patient with desquamative interstitial pneumonitis shows ground-glass attenuation in the periphery of the upper and lower lung fields.

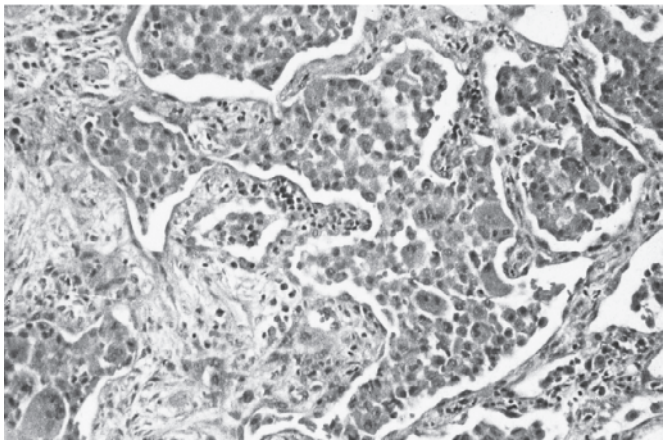


FIG. 72.7 Histopathology of Desquamative Interstitial Pneumonitis. A high-magnification photomicrograph of desquamative interstitial pneumonitis shows the uniform, diffuse accumulation of macrophages within the alveolar space, with associated thickening of the alveolar septum. These aggregates of macrophages almost completely fill the alveolar spaces.

Diagnosis

The diagnosis of DIP requires tissue confirmation of the pathologic lesion. This is important since DIP has a better prognosis and response to therapeutic intervention compared to IPF. A DIP-like pattern is frequently seen in other IIPs, as well as in pulmonary Langerhans cell histiocytosis, CTDs, and drug reactions. Thus, the diagnosis of DIP requires careful correlation of pathologic findings with clinical and radiologic findings.

Treatment and Outcome

The primary intervention in DIP is smoking cessation. Since this is a rare condition with relatively few published cases, it is unclear whether glucocorticoids alter the natural history of this disease. A mortality rate of 28% with a mean survival of 12 years has been reported, compared to a 30% to 50% 5-year survival in UIP.[21] Of note, 22% of patients improved spontaneously and 60% responded to glucocorticoid therapy. This picture is dramatically different from IPF, in which spontaneous improvement does not occur. There are, however, a significant minority of DIP patients who fail to respond to treatment and progress to respiratory failure secondary to advanced fibrosis.

Respiratory Bronchiolitis-Associated Interstitial Lung Disease

RB-ILD is a distinct clinical entity that occurs in current or former cigarette smokers. It is unclear whether RB-ILD and DIP represent different diseases or different ends on the spectrum of the same disease process.[27] DIP occurs predominantly and RB-ILD occurs exclusively in cigarette smokers, suggesting a common pathogenesis related to cigarette smoke exposure.

Clinical Manifestations

The mean age at presentation with RB-ILD is 36 years. Males are more often affected, and all individuals with RB-ILD are cigarette smokers. Symptoms include a dry, nonproductive cough and dyspnea. Clubbing is absent in RB-ILD, whereas it is frequently present in DIP. Laboratory evaluation is nonspecific.

The chest radiograph typically shows diffuse, fine reticular or nodular interstitial opacities with normal lung volumes. Additional findings include bronchial wall thickening and a prominent peribronchovascular interstitium. HRCT may reveal ground-glass opacification and emphysema.

Pulmonary function tests commonly reveal a mixed restrictive-obstructive pattern with a reduced diffusing capacity and mild hypoxemia. The residual volume may be increased with no change in other spirometry.

Histopathology

The pathology of RB-ILD is similar to DIP. However, in RB-ILD, the intra-alveolar macrophages accumulate primarily within the peribronchiolar airspaces and are associated with thickening of the alveolar septum in these areas (Fig. 72.8).

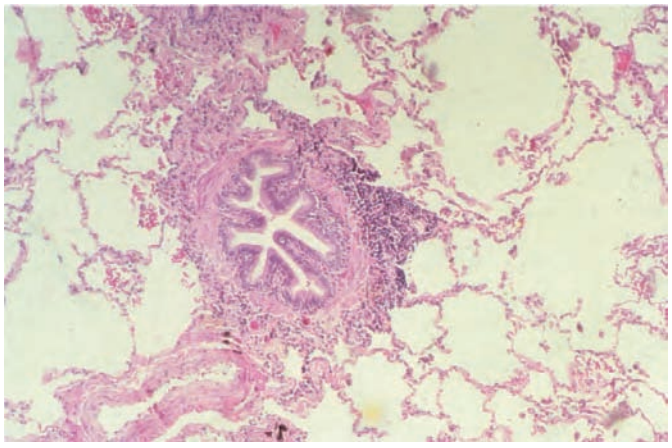


FIG. 72.8 Histopathology of Respiratory Bronchiolitis-Interstitial Lung Disease. An ectatic bronchiole with a thickened wall is shown, with a mononuclear infiltrate extending into the immediately surrounding alveoli.

The differentiation of this lesion from DIP requires sparing of distal airspaces, with the lesion confined to the peribronchiolar airspaces in RB-ILD.

Diagnosis

RB-ILD should be suspected in young individuals with a history of cigarette use who complain of cough and dyspnea with a chest radiograph or HRCT showing nodular and/or reticular interstitial opacities. The diagnosis requires tissue confirmation of the pathologic findings noted above.

Treatment and Outcome

The key therapeutic intervention in RB-ILD is cessation of smoking. The use of glucocorticoids has been associated with favorable results. At present, the clinical course and prognosis of patients with RB-ILD are unknown. In most clinical series, patients either improved or stabilized, and mortality is uncommon.[26,27]

Nonspecific Interstitial Pneumonitis

The term NSIP was first used to describe cases of interstitial pneumonia that did not demonstrate a pattern of UIP, AIP, or DIP. Currently, the term NSIP is applied to an IIP or to a similar histologic pattern that occurs in CTD, hypersensitivity pneumonitis, infection, or drug-induced lung disease. Thus, the diagnosis of NSIP should prompt investigation for a causative agent. In fact, 16% of patients in the original description of NSIP had one of the CTDs.[28]

Clinical Manifestations

Idiopathic NSIP occurs in middle-aged individuals, with a slight female predominance. A dry, nonproductive cough and exertional dyspnea are the most common symptoms, although fever is present in 25% of patients. Symptoms are usually present for 6 to 10 months prior to diagnosis. As in other IIPs, the laboratory evaluation is nonspecific.

The chest radiograph usually shows bilateral interstitial infiltrates and sometimes can be normal in a symptomatic patient. HRCT characteristically shows bilateral, patchy ground-glass attenuation indistinguishable from DIP or RB-ILD.[15]

Histopathology

NSIP is characterized by temporally uniform degrees of fibrosis and inflammation of the alveolar septum, without histopathologic features indicative of UIP, AIP, or DIP (Fig. 72.9). NSIP has been divided into three groups, depending on the presence or absence of interstitial fibrosis: interstitial lymphoplasmacytic inflammation (48% of cases); inflammation and fibrosis (38%); and fibrosis (14%). Although the changes are temporally uniform, they may be patchy with intervening areas of normal lung.

This temporal uniformity is in contrast to the variegated pattern seen in UIP. Fibroblast foci, the earliest lesion seen in UIP, are found in 20% of patients with NSIP, making the differentiation of fibrotic NSIP from UIP difficult. The key feature in this circumstance is the temporal uniformity of the lesion in NSIP.

Treatment and Outcome

Unlike patients with UIP, individuals with NSIP have a favorable prognosis. In the original description of the disease, 45% of subjects completely recovered, while another 42% remained stable or improved.[28] Only 11% of patients died, with a mean survival of 16 months. All of the individuals with an aggressive course were in the fibrotic group. Ten-year survival in the cellular group was 90%, compared to 35% in patients with the fibrotic pattern. Despite the worse prognosis of NSIP with a fibrosing pattern, this is still significantly better than the 15% survival rate at 10 years for patients with UIP.[29]

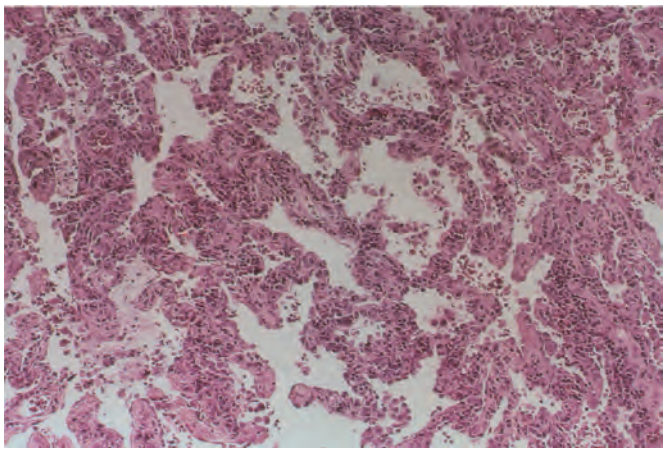


FIG. 72.9 Histopathology of Nonspecific Interstitial Pneumonitis. Low-magnification photomicrograph of cellular nonspecific interstitial pneumonitis shows diffuse uniform thickening of the alveolar septum due to the presence of a lymphoplasmacytic infiltrate.

Cryptogenic Organizing Pneumonia

Cryptogenic organizing pneumonia (COP), which was previously named idiopathic bronchiolitis obliterans organizing pneumonia (BOOP), is a specific clinicopathologic disorder of unknown etiology characterized by excessive proliferation of granulation tissue within the lumen of distal airspaces.[30] The term COP is reserved for cases demonstrating organizing pneumonia without an obvious cause, since this histologic appearance occurs in a variety of inflammatory lung disorders, including CTDs, malignancy, infections, and those caused by medications.

Clinical Manifestations

The onset of disease is usually in the fifth to sixth decades of life; men and women are equally affected. Most individuals have symptoms for less than 2 months prior to diagnosis. The initial presentation is usually with a dry, nonproductive cough and flu-like symptoms, including fever, sore throat, and malaise. This is followed by progressive dyspnea, and routine laboratory evaluation is nonspecific.

The chest radiograph shows diffuse, often patchy alveolar opacities in the setting of normal lung volumes (Fig. 72.10, A). These opacities can be migratory and usually have a peripheral distribution similar to those seen in chronic eosinophilic pneumonia. Rarer radiographic manifestations include linear or nodular interstitial opacities and honeycombing. The presence of a pleural effusion or pleural thickening should suggest an associated CTD. HRCT shows patchy airspace consolidation, especially in the lung periphery, with a lower-lung zone predominance (see Fig. 72.10, B). Other findings include ground-glass attenuation, small nodular opacities, and bronchial wall thickening.

As in other interstitial lung diseases, a restrictive ventilatory defect is the most common pulmonary function abnormality.

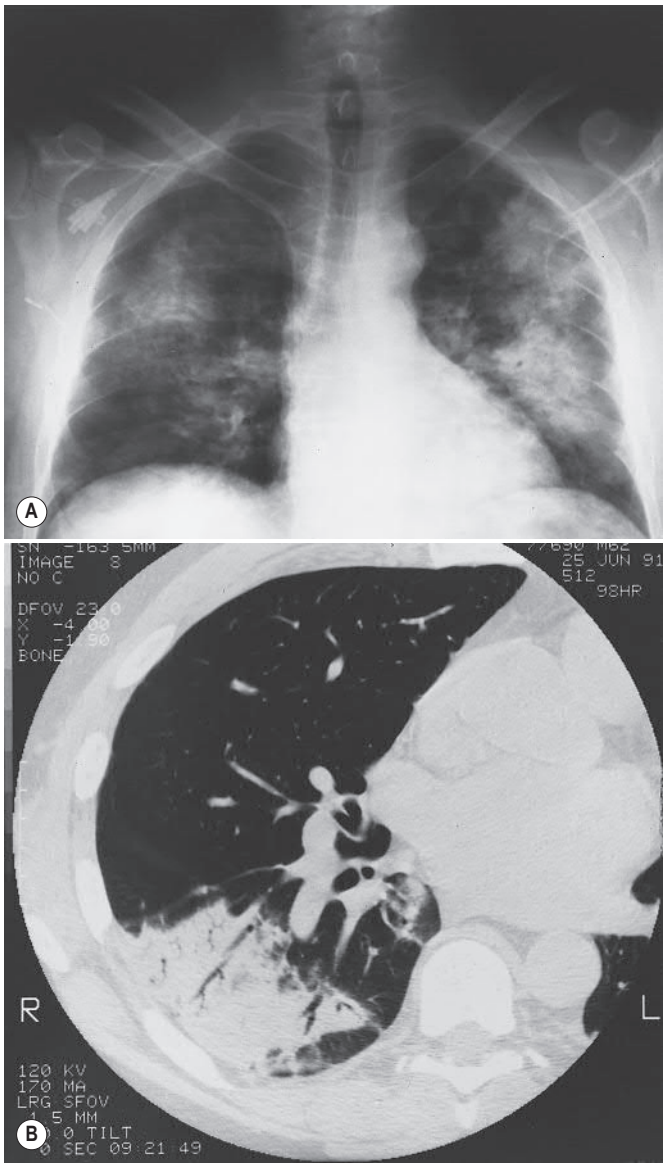


FIG. 72.10 Radiographic Findings in Cryptogenic Organizing Pneumonia. (A) Chest radiograph in a patient with cryptogenic organizing pneumonia shows bilateral patchy alveolar opacities with a peripheral distribution in the setting of normal lung volumes. (B) Chest computed tomography shows a dense right lower lung consolidation with the presence of air bronchograms.

Gas exchange abnormalities are common, and are accompanied by decreased diffusing capacity, widening of the alveolar–arterial gradient, and exercise-induced hypoxemia.

Histopathology

The histopathology of COP is characterized by excessive proliferation of granulation tissue in the small airways and alveolar ducts, with associated chronic inflammation in the alveolar walls (Fig. 72.11).[31] The intraluminal fibrotic buds (Masson bodies) consist of loose collagen-embedding fibroblasts and myofibroblasts and have a tendency to extend from one alveolus to the next, giving a characteristic "butterfly" pattern. The lesions are patchy in nature and have a uniform temporal appearance at low magnification with preservation of the underlying lung

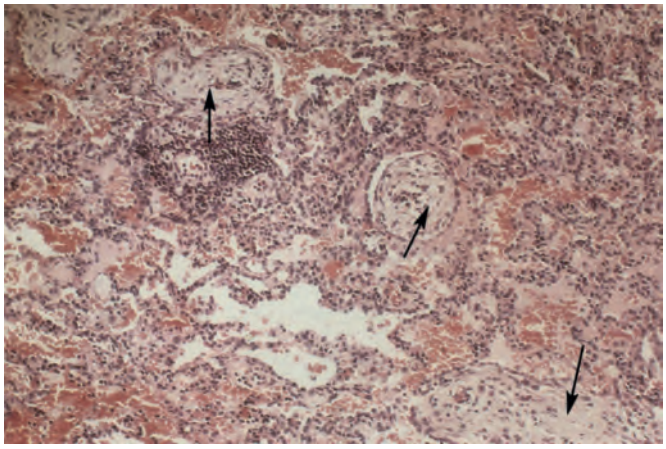


FIG. 72.11 Histopathology of Cryptogenic Organizing Pneumonia. A photomicrograph of cryptogenic organizing pneumonia shows intra-alveolar fibroblast proliferation *(arrows)* and early collagen production. In addition, thickening of the alveolar septa with a lymphoplasmacytic infiltrate consistent with cellular nonspecific interstitial pneumonitis is present.

parenchyma. COP has been described as the prototypical healing response of the lung to a variety of insults.

Diagnosis

The presence of organizing pneumonia in a lung biopsy does not necessarily represent COP since this is a diagnosis of exclusion. Organizing pneumonia is a nonspecific response to many lung injuries, and may occur in conjunction with another pathologic process or as a component of other primary pulmonary disorders, such as infections, irradiation, CTD, hypersensitivity pneumonitis, granulomatosis with polyangiitis, or chronic eosinophilic pneumonia (Table 72.3).

TABLE 72.3 Disorders Associated With Organizing Pneumonia

Secondary Organizing Pneumonia
Connective Tissue Diseases
- Systemic lupus erythematosus
- Rheumatoid arthritis
- Polymyositis/dermatomyositis
- Sjögren syndrome

Hypersensitivity Pneumonitis
Chronic Eosinophilic Pneumonia
Drug-Induced
- Gold
- Penicillamine
- Amiodarone
- Bleomycin
- Sulfa drugs

Wegener Granulomatosis
Bone marrow transplantation
Lung transplantation/rejection
Inhalational injury
Neoplasms
Lung irradiation
Virus-associated:
- Human immunodeficiency virus (HIV)
- Influenza
- Adenovirus

Treatment and Outcome

Treatment with glucocorticoids usually offers dramatic clinical and radiographic improvement within days to weeks.[30] Complete clinical, physiologic, and radiographic recovery occurs in two-thirds of cases. In the remainder, persistent disease with progression to fibrosis occurs. It is common for relapses to occur with glucocorticoid tapering, followed by improvement with reintroduction of treatment; consequently, at least 6 months of therapy is recommended. The 5-year survival in COP is 73%, compared to 5-year survival rates of 44% in patients with organizing pneumonia due to other causes (e.g., CTD), or 30% for IPF.

LUNG INVOLVEMENT IN CONNECTIVE TISSUE DISEASES

CTDs are a heterogeneous group of systemic autoimmune diseases that frequently involve the lungs. The pleuropulmonary manifestations of these diseases are diverse, affecting all parts of the respiratory tract (i.e., airways, alveoli, blood vessels, and pleura) (Table 72.4). Although pulmonary complications generally occur in patients with well-established disease, occasionally lung involvement is the first manifestation of the autoimmune disorder. This section discusses the pleuropulmonary manifestations of systemic lupus erythematosus (SLE), rheumatoid arthritis (RA), and systemic sclerosis (SSc) (for a discussion of other manifestations in these diseases, see Chapters 52, 53, and 56).

TABLE 72.4 Pleuropulmonary Manifestations of Connective Tissue Diseases

	SLE	RA	SSc
Pulmonary hypertension	+	+	+++
Vasculitis	+	±	±
Pleural disease	+++	+++	+
Bronchiolitis obliterans	±	++	+
Aspiration pneumonia	–	–	++
Diaphragmatic dysfunction	++	–	–
Lung nodules	–	++	–
Diffuse alveolar damage	+	±	±
Organizing pneumonia	±	+	±
UIP	+	+++	+
Capillaritis	++	+	±
LIP	+	+	+
NSIP	+	++	+

LIP, Lymphocytic interstitial pneumonitis; *NSIP*, nonspecific interstitial pneumonitis; *RA*, rheumatoid arthritis; *SLE*, systemic lupus erythematosus; *SSc*, systemic sclerosis; *UIP*, usual interstitial pneumonitis.

Systemic Lupus Erythematosus

SLE is a disease of unknown etiology, characterized by the presence of autoantibodies directed against various nuclear antigens. These autoantibodies and the resultant immune complexes mediate many of the manifestations of SLE (Chapter 52). This disease primarily affects young women (female-to-male ratio > 8:1) and may involve virtually every organ system. Pleuropulmonary involvement occurs at some point in the disease course in 38% to 89% of cases.[31] Thus, the respiratory system is affected more commonly in SLE than in any other CTD. However, infectious pneumonia remains the most common cause of pulmonary disease and death in these patients. Thus, in SLE patients presenting with a febrile illness and pulmonary infiltrates, a community-acquired or opportunistic infection must be promptly excluded.

Acute Lupus Pneumonitis

Acute lupus pneumonitis is an uncommon pulmonary manifestation of SLE, occurring in fewer than 5% of cases.[31] The clinical presentation mimics that of an infectious pneumonia with the abrupt onset of fever, cough, and dyspnea. Serum complement levels are often low, and the chest radiograph typically shows diffuse alveolar opacities. It can be accompanied by pericarditis, and often pleuritis and pleural effusion.

It can be difficult to distinguish acute lupus pneumonitis from an infectious pneumonia. BAL is recommended prior to instituting corticosteroid therapy. The histopathology varies, and includes diffuse alveolar damage, organizing pneumonia, NSIP, or a combination of these.

There are no controlled trials of therapy for acute lupus pneumonitis. Treatment includes high-dose glucocorticoids (1 to 2 mg/kg/day) with or without accompanying cytotoxic drugs, such as cyclophosphamide. Mortality rates as high as 50% have been reported in untreated cases. In patients who fail to respond to treatment, respiratory failure is the usual cause of death.

Diffuse Alveolar Hemorrhage

Diffuse alveolar hemorrhage (DAH) occurs in fewer than 5% of patients with SLE, and it represents the initial manifestation of disease in 11% to 20% of those cases.[32] However, most cases develop in individuals with well-established diagnoses of SLE, usually with pre-existing lupus nephritis.

Hemoptysis is not present in every case of DAH at presentation. Therefore, the absence of hemoptysis does not exclude the diagnosis, particularly in the setting of a falling hematocrit, diffuse pulmonary infiltrates, and progressively bloody BAL fluid. DAH in SLE is most often due to pulmonary capillaritis, but it can also be caused by diffuse alveolar damage. Immunofluorescence studies show granular deposits of immunoglobulin G (IgG) and C3 along alveolar walls, interstitium, and capillary endothelial cells.

There are no controlled trials for the treatment of alveolar hemorrhage in SLE. Glucocorticoids, cytotoxic drugs, plasmapheresis, and extracorporeal membrane oxygenation (ECMO) have been used in various combinations. The mortality rate associated with DAH and concomitant respiratory failure approaches 100% without treatment; therefore early recognition and aggressive treatment should be undertaken.

Lupus Pleuritis

The pleura are the most common site of respiratory involvement in SLE, with pleurisy and pleural effusions occurring in 50% to 80% of patients. Lupus pleuritis can be the presenting manifestation of disease, but more commonly develops in patients with established SLE and is often recurrent. The clinical manifestations include chest pain, fever, and dyspnea, and the chest radiograph typically shows bilateral pleural effusions. The pleural fluid is serous or serosanguineous and exudative in nature. Compared to effusions in RA, the glucose is higher, and the lactate dehydrogenase level is lower. The most helpful measurement is a pleural fluid ANA titer greater than 1:160. Examination of the pleura reveals infiltration with plasma cells and lymphocytes, accompanied by pleural thickening and fibrosis. Treatment with nonsteroidal anti-inflammatory drugs

and/or glucocorticoids is usually effective for relief of pleural discomfort.

Interstitial Lung Disease

The presence of ILD in SLE is uncommon, especially when compared to SSc or RA. However, minor interstitial abnormalities can be found on HRCT in approximately one-third of SLE patients despite normal physiologic testing. The significance and natural history of these subclinical findings are uncertain. The presence of anti-SSA (Ro) has been noted in approximately 80% of lupus patients with interstitial changes, and ILD is more common in those SLE patients with sclerodermatous skin changes.

The diagnosis of SLE is usually well-established in patients who develop the insidious form of ILD. The disease course is characterized by progressive dyspnea and cough; the chest radiograph shows reduced lung volumes and reticular interstitial infiltrates. A restrictive lung function pattern with reduced diffusing capacity and exercise-induced hypoxemia are typical. The histopathology of chronic interstitial disease in SLE resembles NSIP, although cases of organizing pneumonia, lymphocytic interstitial pneumonitis (LIP), and UIP have been described. Response to therapy depends on the underlying histopathology, with the UIP-like form being least responsive.

Pulmonary Vascular Disease

Although previously thought to be unusual, the development of pulmonary hypertension has been increasingly noted in SLE, with an incidence ranging from 0.5% to 17%. Pulmonary hypertension in SLE has been associated with the presence of Raynaud syndrome, serositis, digital vasculitis, and antiphospholipid antibodies.[33] Dyspnea and fatigue, despite a normal chest radiograph, is the most common presentation. The majority of SLE patients with pulmonary hypertension are female, with 3- and 5-year survival rates of 45% and 17%, respectively, which represents a worse prognosis than patients with idiopathic pulmonary hypertension. The vascular changes of SLE-associated pulmonary hypertension are similar to those seen in idiopathic pulmonary arterial hypertension (PAH) with intimal hyperplasia, smooth muscle hypertrophy, and medial thickening. Several pathologic mechanisms have been proposed for the development of pulmonary hypertension, including vasoconstriction in addition to vasculitis and thrombosis with an association with antiphospholipid and anticardiolipin antibodies. Serum endothelin levels are elevated in patients with SLE-associated pulmonary hypertension and correlate with pulmonary arterial pressures.

Pulmonary function testing shows an isolated decrease in the DLCO. Patients with SLE-associated pulmonary hypertension may respond to immunosuppressive therapy, such as azathioprine or mycophenylate mofetil, based on small studies indicating hemodynamic improvement with this approach. SLE patients treated with bosentan did not have clinical worsening and showed an improvement in 6-minute walk distance.[34] Unfortunately, despite these modest improvements with pharmacotherapy, the long-term survival of patients with SLE-associated pulmonary hypertension is poor.

Respiratory Muscle Dysfunction

The shrinking lung syndrome is due to diaphragmatic weakness as well as weakness of other respiratory muscles. This entity accounts for the findings of dyspnea without evidence of interstitial infiltrates or pulmonary vascular disease. It occurs in 25% of patients with SLE. The chest radiograph typically shows elevated diaphragms and basilar atelectasis. The pathogenesis of respiratory muscle dysfunction is unknown; however, weakness remains localized to the respiratory compartment without generalized muscle involvement, indicating specific involvement of the muscles of respiration. Glucocorticoids are generally ineffective in the treatment of this syndrome. Improvement has been noted with inhaled β-agonist and theophylline therapy. Despite a variable response to therapy, it is unusual for this manifestation of SLE to be progressive.

Rheumatoid Arthritis

RA is an autoimmune disease associated with autoantibodies directed against citrullinated antigens and characterized by the presence of a symmetric, inflammatory polyarthritis (Chapter 53). It occurs more frequently in women, with a 2:1 female-to-male ratio. Disease onset is most commonly in the fourth to fifth decades of life. Pleuropulmonary complications of RA occur more commonly in individuals with subcutaneous nodules, high titers of rheumatoid factor, anti-cyclic citrullinated proteins (anti-CCP), and more severe chronic articular involvement. Although RA itself is more common in women, the pleuropulmonary manifestations occur more commonly in men. The pleuropulmonary complications of RA are numerous, but there are also treatment-related lung toxicity and pulmonary infections related to immunosuppression, which can complicate the diagnosis of RA-related lung manifestations.

Pleuritis and Pleural Effusions

Pleural effusion in RA is common and can occur prior to the development of arthritis. Pleural disease is often discovered as an incidental finding on routine chest radiographs, but nonspecific chest pain, dyspnea, and fever are not unusual. The effusion can be unilateral or bilateral and can coexist with interstitial lung disease. Typically, the effusion is an exudate, with a glucose level usually less than 30 mg/mL. The mechanism underlying the low pleural fluid glucose is impaired membrane transport of glucose. A low pleural fluid pH is thought to occur secondary to impaired carbon dioxide exit from the pleural space. If the effusion is chronic, the cholesterol concentration can be increased, and the pleural fluid can have a milky appearance (pseudochylothorax). Cytologic examination reveals multinucleated giant cells, spindle-shaped macrophages, and necrotic debris.

Most rheumatoid effusions are small and asymptomatic, requiring no treatment. They resolve over several months without complications. The use of glucocorticoids for active articular disease hastens the resolution of the pleural process.

CLINICAL PEARLS

Lung Involvement in Rheumatoid Arthritis

- RA is more common in women, but pleuropulmonary complications occur more frequently in men.
- Factors associated with pleuropulmonary complications of RA include more severe articular involvement, subcutaneous nodules, and high levels of rheumatoid factor and antibodies to citrullinated peptides.
- Pleural effusions are the most common complication, characterized by an exudate, low glucose, and low pH.
- The differentiation of rheumatoid nodules from malignant lesions can be difficult.
- The rapid growth of a nodule should prompt aggressive investigation for a malignant cause.

Rheumatoid Nodules

Rheumatoid or necrobiotic nodules are most commonly seen in men with active articular disease, high rheumatoid factor titers, and subcutaneous nodules. Most individuals are asymptomatic and are diagnosed on routine chest radiograph. Radiographically, these nodules can be singular or multiple with an upper to mid-lung zone predominance. Cavitation occurs in approximately 50% of cases. HRCT indicates a higher frequency of nodules than previously thought. Rarely, subpleural necrobiotic nodules can erode into the pleural space, resulting in a pneumothorax with a complicating bronchopleural fistula. It can be difficult to differentiate these nodules from malignant lesions, making thoracoscopic lung biopsy necessary. Evidence on chest radiograph of rapid growth should prompt an aggressive diagnostic evaluation.

Airway Disease

Airflow limitation is a common finding in patients with RA, being present in approximately one-third of patients. The mechanism(s) responsible for airway disease is poorly understood. The interplay of cigarette smoking and RA may play a role.

A life-threatening complication of RA is upper-airway obstruction, resulting from synovitis of the cricoarytenoid joint. Common presenting complaints include a sore throat, hoarseness, and fullness in the throat. It can progress to inspiratory stridor and upper-airway obstruction. This complication occurs more commonly in women, particularly in those with advanced RA. Seventy-five percent of patients were found to have cricoarytenoid abnormalities when screening with direct or indirect laryngoscopy and computed tomography was utilized. The treatment of cricoarytenoid arthritis includes anti-inflammatory medications.

Constrictive bronchiolitis is a progressive form of obstructive lung disease that is being increasingly recognized as a complication of RA.[35] The histopathologic lesion of constrictive bronchiolitis is concentric submucosal and peribronchiolar fibrosis resulting in extrinsic compression and obliteration of the bronchiolar lumen. The typical clinical presentation is with insidious onset of cough and dyspnea, with a normal or hyperinflated chest radiograph. This complication occurs more commonly in women than in men. Pulmonary function studies show airflow limitation with hyperinflation and a reduced diffusing capacity. Inspiratory imaging on HRCT demonstrates mosaic attenuation, while expiratory imaging demonstrates corresponding areas of air trapping. Some individuals respond to high-dose glucocorticoids and cytotoxic drugs, but most patients with bronchiolitis obliterans progress to respiratory failure and require lung transplantation.

Interstitial Lung Disease

Although ILD is a common complication of RA, the incidence is difficult to determine, since different methods of detection have been employed and dissimilar populations of patients have been studied; however, recent estimates of RA-ILD are 3.2 to 6.0 cases per 100,000 people.[36] The development of ILD in relation to the onset of arthritis is variable. Most often, the ILD develops subsequent to arthritis, but, in approximately 20% of patients, the lung disease precedes the onset of arthritis. ILD in RA is associated with cigarette smoking, presence of the shared HLA-DR4 epitope, and RA-specific anti-citrullinated protein antibodies. The most important genetic risk factor for the development of ILD in RA patients appears to be the presence of the same *MUC5B* promoter polymorphism, similar to what has been observed in IPF.[13,37]

The most common histopathologic pattern of ILD in RA is UIP, followed by NSIP, LIP, and organizing pneumonia. The clinical manifestations of ILD in RA resemble those seen in idiopathic disease and include a dry, nonproductive cough and dyspnea on exertion. The chest radiograph and HRCT show increased reticular markings with a predilection for the peripheral lower-lung zones. LIP usually occurs in cases of RA complicated by Sjögren syndrome; the presence of keratoconjunctivitis sicca and xerostomia in a patient with RA and ILD should suggest this histologic subtype.

Additionally, some patterns of ILD in RA patients are believed to be related to medications used to treat the condition. The most common of these is methotrexate-induced lung disease. Methotrexate lung toxicity presents with the subacute onset of fever, cough, and dyspnea occurring 1 to 5 months after initiation of the drug. The chest radiograph shows mixed interstitial–alveolar infiltrates. Nonspecific laboratory abnormalities include leukocytosis, sometimes with mild eosinophilia, and an elevated erythrocyte sedimentation rate. In most cases, BAL reveals a lymphocytosis. Histologically, cellular NSIP is seen with areas of organizing pneumonia. Non-caseating granulomatous inflammation similar to that seen in hypersensitivity pneumonitis may also be present. The primary treatment of methotrexate-induced pneumonitis is withdrawal of methotrexate in addition to supportive care.

Unfortunately, RA-ILD typically confers a poor prognosis with high morbidity and mortality, especially in the presence of a UIP pattern. Current therapeutic strategies for RA-ILD are similar regardless of the underlying histopathologic pattern and are based on steroid-sparing therapies such as azathioprine or mycophenolate mofetil. Additionally, a recent study evaluated the use of nintedanib in a non-IPF population of progressive fibrotic ILD which included RA-ILD patients and had similar results to previous IPF trials, indicating a possible role for nintedanib in this condition.[38]

Systemic Sclerosis (Scleroderma)

SSc is characterized by excessive deposition of extracellular matrix in the skin and internal organs, and vascular involvement (Chapter 56). The degree of visceral organ involvement determines morbidity and mortality. Pulmonary involvement occurs in 70% to 100% of patients with SSc and does not correlate with the degree of extrapulmonary disease. Interstitial lung disease is the most common pulmonary manifestation of SSc. Of note, with the improved mortality associated with renal involvement in SSc, lung disease has become the most important cause of morbidity and mortality.

Interstitial Lung Disease

The incidence of ILD in SSc depends on the method of detection. Autopsy studies have reported an ILD incidence of 60% to 100% of cases, whereas studies based on chest radiographs have noted interstitial changes in 14% to 66% of cases. Cough and dyspnea on exertion are the most common symptoms. Physical examination reveals bibasilar rales. Radiographic findings include basal reticulonodular infiltrates, enlargement of pulmonary arteries, and progressive volume loss. Pulmonary function testing reveals restrictive lung disease, preservation of flow rates, and decreased diffusing capacity. A disproportionate

decrease in diffusing capacity compared to lung volume changes should suggest pulmonary hypertension, especially in individuals with limited scleroderma (calcinosis, Raynaud phenomenon, esophageal dysmotility, sclerodactyly, and telangiectasia [CREST] syndrome). The predominant histopathologic abnormality is NSIP. Rarely, LIP may complicate cases of SSc associated with Sjögren syndrome. Although the 5-year survival for SSc patients with ILD is 38% to 45%, it is better than that of patients with IPF.

Pulmonary Vascular Disease

Pulmonary hypertension is a frequent complication of SSc, occurring in approximately 30% of patients with diffuse scleroderma and in 10% to 50% of those with limited scleroderma (Chapter 56). Because it is a major cause of morbidity and mortality in systemic sclerosis, it has also become part of the diagnostic criteria for the disease.[39] Pulmonary hypertension can either be associated with interstitial fibrosis or result from involvement of small and medium-sized arteries and arterioles with smooth-muscle hyperplasia, medial hypertrophy, and intimal proliferation (plexogenic). Direct involvement of the pulmonary circulation is more common with limited scleroderma, while pulmonary hypertension in patients with diffuse scleroderma is more likely associated with ILD.

The clinical presentation is characterized by the insidious onset of fatigue and dyspnea on exertion. Physical examination and chest radiographs show signs typical of pulmonary hypertension, while a decreased diffusing capacity is seen on pulmonary function testing. Risk factors for developing SSc-associated pulmonary hypertension include limited skin involvement, duration of disease greater than 10 years, onset of SSc at older age, and severity and duration of Raynaud phenomenon.

The pathogenesis of SSc-associated pulmonary hypertension is poorly understood. Vascular changes occur at an early stage in SSc, including apoptosis, endothelial cell activation with increased expression of cell adhesion molecules, inflammatory cell recruitment, intimal proliferation, and adventitial fibrosis leading to vessel obliteration. Endothelial injury is reflected by increased levels of soluble cell adhesion molecules, disturbances of angiogenesis with increased levels of circulating vascular endothelial growth factor, and presence of angiostatic factors. It remains unclear to what extent dysregulated angiogenesis in SSc-associated pulmonary hypertension is driven by an inflammatory process or other as-yet unidentified mechanisms.

Treatment of SSc-associated pulmonary hypertension has been disappointing, with no therapy showing a significant survival benefit. Calcium channel blockers are not usually indicated for patients with SSc-associated pulmonary hypertension, although often used at lower doses for Raynaud phenomenon. Continuous intravenous epoprostenol improves exercise capacity and hemodynamics.[40] Randomized clinical trials with phosphodiesterase inhibitors, including sildenafil, showed a modest effect on exercise capacity, hemodynamic parameters, and functional class after 12 weeks of treatment. Carefully selected patients may be considered for heart–lung transplantation but are often excluded due to postoperative complications arising for SSc-related gastrointestinal reflux disease and renal dysfunction.

Sarcoidosis

Sarcoidosis is a disease of unknown etiology that occurs in individuals of all ages, sex, and ethnic backgrounds.[41] The disease can affect every compartment in the body; however, in over 90% of cases the lungs are involved. Most cases of sarcoidosis are self-limited and typically responsive to glucocorticoid-based regimens, but some patients develop a chronic form of the condition which typically necessitates a steroid-sparing treatment-based strategy.

Epidemiology

The incidence of sarcoidosis is heavily influenced by geographical factors, as well as the race of the population being studied. For instance, within an American population, African American women had a higher incidence of sarcoidosis when compared to Caucasian women (17.8 per 100,000 per year vs. 8.1 per 100,000 per year).[42] Furthermore, the incidence of sarcoidosis varies widely based on the region of the world being studied (ranging from 2.3 per 100,000 per year in Guadalupe to 11.5 per 100,000 per year in Sweden). Sarcoidosis typically affects adults in a bimodal distribution, with common peaks of diagnosis in the 20–29 and 60–69 age groups.

Pathogenesis

Despite being recognized as a distinct clinical entity for nearly 150 years, the cause of sarcoidosis remains unknown. Similar inflammatory granulomatous disorders, such as berylliosis, result in a nearly identical clinical disorder. This suggests that the underlying pathophysiology of sarcoidosis is similar to that of conditions like berylliosis, but the antigen (or antigens) leading to the phenotype of sarcoidosis remains unknown. While the environmental antigen that causes sarcoidosis is not yet known, it is clear that an environmental factor plays a causal role in the development of sarcoidosis, based on regional variation and penetrance of the disease. Environmental antigens trigger an immunologic reaction, leading to granuloma formation. Several potential causal antigens have been explored in sarcoidosis. Suspected antigens have included organic and environmental airborne exposures. For instance, since 2001, World Trade Center first responders have developed a pulmonary granulomatous disorder, similar to sarcoidosis, likely related to the inorganic and organic respiratory particle inhalation from their exposure to the debris from the site.[43] Other potential antigens in sarcoidosis have included infectious organisms such as Mycobacteria or Cutibacteria, given evidence of genetic elements from these organisms in the granulomas from patients with sarcoidosis.[44]

The hallmark immunologic feature of sarcoidosis is the accumulation of CD4 T cells in lung tissue; the BAL, CD4 T cell, and B-cell lymphopenia are common in peripheral blood. Th1 effector phenotype has expression of IFN-γ, TNF, and IL-2. CD4 T cells with a Th17 phenotype and T cells expressing both IL-17 and IFN-γ (i.e., Th17.1 cells) have been found in lung tissue and BAL fluid from patients with sarcoidosis, but whether these cells play a role in disease progression is currently unknown.[43] This inflammatory cytokine milieu ultimately drives the formation of the sarcoidosis granuloma and the clinical phenotype of this syndrome.

Genetic studies show that sarcoidosis is most strongly associated with the HLA region on chromosome 6, with *HLA-DRB1*11:01* increasing risk in both African Americans and Whites.[44] In Europeans, *DRB1*03:01* has a strong association with increased disease risk. However, this allele was found to be protective against sarcoidosis in an African American cohort.[44] Genome-wide association studies have identified associations with *ANXA11*, *NOTCH4*, and *BTNL2* (reviewed in Moller

et al.[44]). However, the role of these single-nucleotide polymorphisms in the pathogenesis of sarcoidosis remains unknown.

Clinical Manifestations

Sarcoidosis is a multisystem inflammatory disorder that can affect every organ with wide-ranging clinical impacts based on the severity of involvement and the organ system involved (Table 72.5). While every organ system is at risk for involvement with sarcoidosis, the lung is the site most commonly impacted, with over 90% of sarcoidosis cases exhibiting some degree of pulmonary involvement. The following discussion will focus primarily on the pulmonary manifestations of sarcoidosis.

Symptoms of pulmonary sarcoidosis can vary from incidental radiographic findings of bilateral hilar lymphadenopathy in asymptomatic individuals (see Fig. 72.1) to marked dyspnea leading to significant morbidity and mortality. Pulmonary function testing can vary in sarcoidosis. In asymptomatic cases or cases with only nodal involvement of sarcoidosis, there can be little to no abnormality of pulmonary physiology. However, in cases with advanced parenchymal fibrosis, there will typically be pulmonary restriction and a reduction in DLCO. Additionally, sarcoidosis can lead to pulmonary vascular disease through two mechanisms: (1) primary involvement of the pulmonary vasculature bed and (2) secondary pulmonary hypertension as a result of parenchymal fibrosis and pulmonary restriction. Therefore, pulmonary sarcoidosis can also result in out-of-proportion reductions in DLCO and morbidity due to pulmonary vascular disease. Pulmonary sarcoidosis can also display an obstructive ventilatory defect, which is often related to endobronchial involvement of sarcoidosis.

Chest radiographs in sarcoidosis have been characterized by the Scadding staging system (Fig. 72.12). This system describes five stages of lung parenchymal involvement in sarcoidosis, from Stage 0 (absence of pulmonary findings) to Stage IV (pulmonary fibrosis without nodal involvement). While the Scadding staging system remains a useful tool for prognosis, advances in imaging such as HRCT have improved our understanding of lung parenchymal involvement in sarcoidosis (Fig. 72.13, A–C).[45]

Histopathology

The pathologic hallmark of sarcoidosis is the non-caseating granuloma (Fig. 72.14).[46] Sarcoidosis granulomas are discreet and compact, consisting of a central core of epithelioid cells, macrophages, and CD4 T cells, with a peripheral localization of CD4 and CD8 T cells along with B cells. Multinucleated giant cells are scattered throughout the granuloma (see Fig. 72.14). Similar to berylliosis, granulomatous inflammation in sarcoidosis occurs along the peribronchovascular bundle.

The histologic finding of non-caseating granulomas requires a search for other potential causes of granulomatous inflammation, such as infection (e.g., mycobacteria, fungi, and parasites) or foreign material (talc), by meticulous analysis of the tissue using special stains and cultures. Typically, there is no necrosis in the granulomas of sarcoidosis; however, the finding of necrosis does not exclude the diagnosis of sarcoidosis. While sarcoidosis is a multisystem granulomatous disorder, it is rarely necessary to biopsy more than one site, as findings of granulomatous inflammation in one affected organ with clinical end-organ damage elsewhere is typically enough to secure the diagnosis in the right clinical scenario.

TABLE 72.5 Clinical Manifestations of Sarcoidosis

Organ System (% Involvement)	Clinical Features
Pulmonary (90%)	Hilar adenopathy, peribronchovascular nodularity, bronchiectasis
Ocular (5%–10%)	Anterior and posterior uveitis, optic neuritis, glaucoma, chorioretinitis
Skin (20%–30%)	Erythema nodosum, nodules and plaques, lupus pernio
Hepatic (10%)	Hepatomegaly, cirrhosis, jaundice
Cardiac (10%–15%)	Heart block, arrhythmias, cardiomegaly, sudden death
Nervous system (5%–10%)	Cranial neuropathies, mass, meningitis, seizures, obstructing hydrocephalus, small fiber neuropathy
Hematologic (30%–50%)	Anemia, lymphopenia, thrombocytopenia, splenomegaly, lymphadenopathy
Joint/muscle (10%–20%)	Arthritis, bone cysts, myopathy
Endocrine (<10%))	Hypercalcemia, hypercalciuria, diabetes insipidus, hypopituitarism
Renal (<5%)	Renal calculi, nephrocalcinosis, renal failure

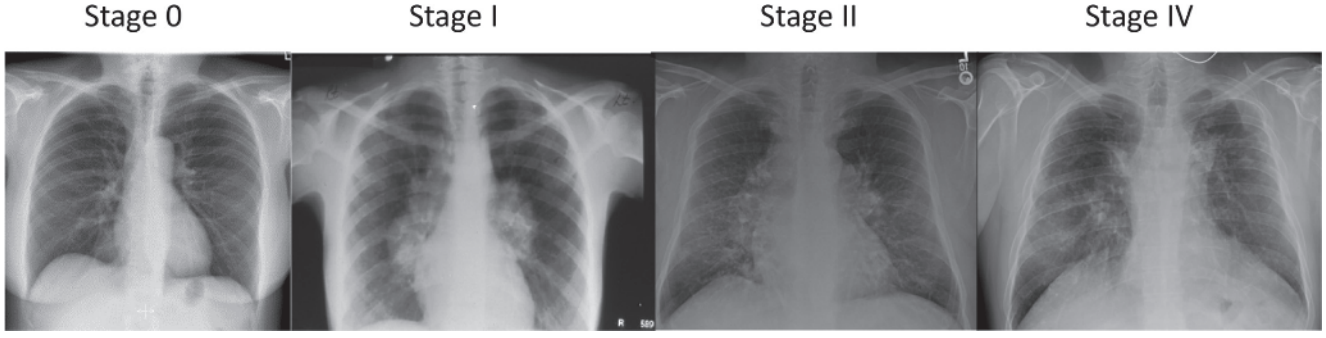


FIG. 72.12 Radiographic Staging of Sarcoidosis via the Scadding Stages. Stage 0 is a normal-appearing chest x-ray. Stage 1 has hilar lymphadenopathy in the absence of lung parenchymal infiltrates. Stage II has hilar lymphadenopathy as well as pulmonary infiltrates. Stage III has pulmonary infiltrates alone. Stage IV is notable for pulmonary fibrosis. Fifty percent of patients with sarcoidosis present at Stage I, and 75% to 80% of those have spontaneous remission.

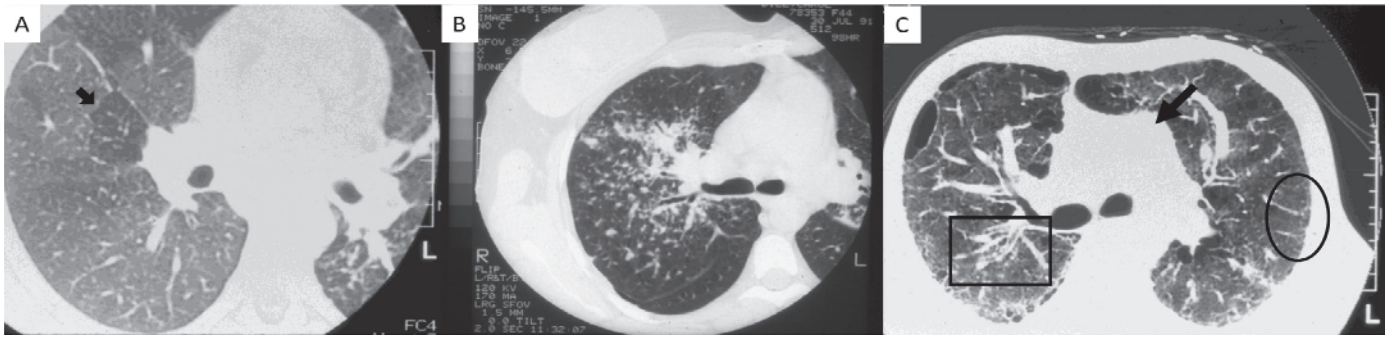


FIG. 72.13 High-Resolution Computed Tomography (HRCT) Images of Lung Involvement in Sarcoidosis. (A) HRCT showing lobular mosaicism *(arrow)* with differential air-trapping in a lobule of the lung likely reflecting airways disease from granulomatous inflammation. Tiny nodules are also present throughout the lung parenchyma in a peribronchovascular distribution. (B) HRCT image of the lung shows nodules with confluence around right hilar lymphadenopathy. These nodules and the alveolar infiltrate are perilymphatic in distribution. (C) HRCT image showing evidence of pulmonary hypertension as depicted by enlarged peripheral pulmonary arteries *(black arrow)*, peripheral reticular abnormality *(black circle)*, and areas of traction bronchiolectasis *(black square)*.

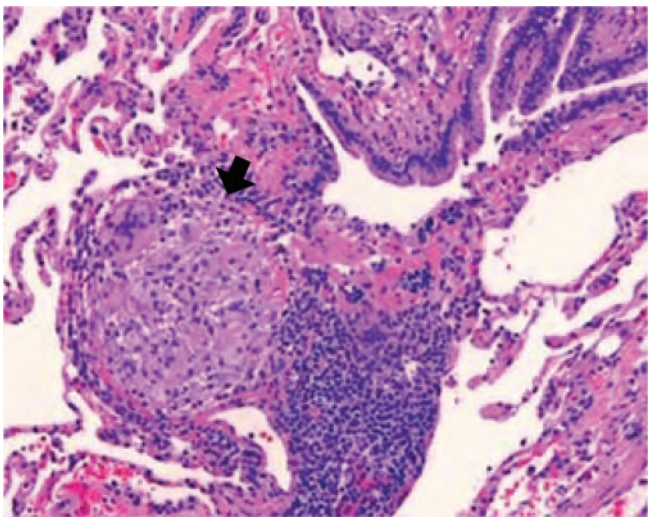


FIG. 72.14 Lung Biopsy Showing Non-Caseating Granulomatous Inflammation *(arrow)* in a Sarcoidosis Patient. The granulomas consist of a tight conglomerate of epithelioid and multinucleated-giant cells encircled by CD4 T and B lymphocytes.

Diagnosis

The diagnosis of sarcoidosis is typically made through a combination of clinical reasoning to exclude other diagnoses combined with radiographic and histologic findings consistent with sarcoidosis. While the diagnosis typically requires histopathologic confirmation of granulomatous inflammation, there are clinical scenarios where a biopsy may not be necessary. Certain clinical syndromes of sarcoidosis, such as Löfgren syndrome (e.g., erythema nodosum, arthritis, and bilateral hilar lymphadenopathy) or Heerfordt syndrome (uveoparotid fever and facial nerve palsy), rarely require a biopsy. With these exceptions, the suspected diagnosis of sarcoidosis requires tissue evaluation. In most cases, selection of the tissue site for biopsy is guided by evidence of end-organ involvement, as well as considerations for patient safety and ease of biopsy. [^{18}F]-fluorodeoxyglucose positron emission scanning (PET) can detect active sites of granulomatous inflammation that can be useful to identify alternative sites for biopsy if certain sites are deemed high risk (e.g., brain or myocardium). Given the high rate of lung involvement and advances in bronchoscopic biopsy techniques, such as endobronchial ultrasound (EBUS) and transbronchial needle aspiration (TBNA), the most common method for obtaining tissue diagnosis for sarcoidosis is via bronchoscopy. This also allows for evaluation of BAL fluid, which can aid in the evaluation of alternative diagnoses (e.g., infection or hypersensitivity pneumonitis). BAL fluid also allows for evaluation of lymphocyte populations from the distal pulmonary compartment, which can aid in the diagnosis of sarcoidosis. For instance, a high CD4:CD8 ratio (>3.5) is indicative of a CD4 T-cell alveolitis, consistent with sarcoidosis.

Serum biomarkers are not useful in isolation to diagnose sarcoidosis. Angiotensin-converting enzyme (ACE) is elevated in 40% to 50% of sarcoidosis patients and is mentioned in international guidelines as a diagnostic biomarker for sarcoidosis. However, it does not have high sensitivity or specificity, and many conditions that mimic sarcoidosis also have high ACE levels (e.g., tuberculosis, fungal infections, and thyroid disease).

Given the multisystem nature of the disease, other organs should be screened after a diagnosis of sarcoidosis is made. Screening typically includes pulmonary function testing, electrocardiogram (ECG), screening ophthalmologic exam, comprehensive metabolic panel, complete blood count, measurement of vitamin D levels, and a 24-hour urine for calcium excretion. In the presence of cardiac symptoms or an abnormal ECG, an echocardiogram and Holter monitor should be performed. If abnormalities are detected on the ECG, echocardiogram or Holter, cardiac magnetic resonance imaging (MRI), or cardiac PET may be indicated to assess for cardiac sarcoidosis. Cardiac MRI and PET have greater sensitivity for detecting patchy inflammation or scarring. Since endomyocardial biopsy has a lower sensitivity for detecting endomyocardial granulomas, the diagnosis of cardiac sarcoidosis hinges on histologic confirmation at other sites, along with cardiac imaging studies that are consistent with a diagnosis of sarcoidosis.

Treatment and Outcome

Seventy percent of cases with acute sarcoidosis experience spontaneous and lasting remission without evidence of persistent

CLINICAL PEARLS

Initial Evaluation of Patients Diagnosed With Sarcoidosis

- Chest radiograph or computed tomography
- Pulmonary function tests (spirometry, lung volumes, and diffusing capacity)
- Electrocardiogram
- Ophthalmologic evaluation
- Blood work including complete blood count, comprehensive metabolic profile, and vitamin D
- 24-hour urine for calcium excretion

disease activity. Additionally, many cases of sarcoidosis are incidental discoveries in asymptomatic patients. For most cases, careful clinical monitoring is sufficient for initial management. However, for those patients with symptomatic pulmonary sarcoidosis or evidence of end-organ damage (progressive pulmonary infiltrates, progressive loss of pulmonary function, development of pulmonary arterial hypertension), glucocorticoids should be initiated. In many cases of pulmonary sarcoidosis, glucocorticoids alone are enough to lead to remission. However, for cases of progressive, chronic sarcoidosis where long-term treatment is required, a steroid-sparing regimen should be initiated.[41]

Hypersensitivity Pneumonitis

Hypersensitivity pneumonitis (HP) is a non-IgE-mediated inflammatory lung disease that results from recurrent exposure to any of a wide variety of inhaled antigenic aerosols containing organic and possibly infectious matter. Pulmonary involvement is typically diffuse and includes predominantly mononuclear inflammation of the terminal bronchioles, interstitium, and alveoli, with little involvement of the larger airways. Despite the term "extrinsic allergic alveolitis," another commonly used name for this disorder, low-grade eosinophilia, is only seen in 1% to 3% of bronchoalveolar lavage specimens and is not a consistent or characteristic feature of disease. Over time, this pattern of inflammation leads to destruction of alveoli and to irreversible pulmonary fibrosis that may be fatal.

The etiology of HP is frequently idiopathic, but many cases can be traced to exposure to organic aerosols that contain thermophilic bacteria (e.g., *Saccharopolyspora rectivirgula*) commonly found in heated water reservoirs such as room humidifiers and hay; filamentous fungi (e.g., *Aspergillus* spp.); animal proteins and fecal matter (e.g., pigeon breeder's disease); and industrial chemicals such as isocyanates. Symptoms including chest tightness, chest pain, dyspnea, and fever appear 4 to 6 hours after exposure. Removal of the offending antigen may prevent the progression to chronic, irreversible disease. Other than oxygen therapy in the setting of profound hypoxemia, there is no defined medical therapy for HP; specifically, there is no role for glucocorticoids.

The immunopathogenesis of HP is complex and not well understood. During the acute presentation, HP is thought to be initiated by an immune-complex-mediated hypersensitivity (type III), with in situ immune-complex deposition in the lung interstitium occurring as a result of the interaction of the inhaled antigen and pre-existing IgG antibodies in the alveolar spaces. Complement activation occurs and most likely contributes to the alveolitis and neutrophilia. However, a putative question remains as to why many people with precipitating antibodies to putative HP antigens (precipitins) do not develop actual parenchymal or symptomatic disease. Therefore, according to a two-hit model, antigen exposure associated with genetic or environmental promoting factors provokes an immunopathologic response, mediated by immune complexes in the acute form and by Th1 and likely Th17 T cells in subacute/chronic cases.[47] Subsequent disease stages evolve into predominant mononuclear inflammatory infiltrates consisting of lymphocytes, plasma cells, and foamy macrophages, followed by granulomas. In advanced disease, fibrosis replaces the inflammatory infiltrates.

THERAPEUTIC PRINCIPLES

Indications for Corticosteroids in Sarcoidosis

- Pulmonary involvement:
  - Moderate or severe symptomatic pulmonary disease
  - Progressive pulmonary disease
- Extrapulmonary involvement:
  - Severe ocular, cardiac, hematologic, or central nervous system disease
  - Persistent hypercalcemia
  - Posterior or anterior uveitis that is unresponsive to topical corticosteroids
  - Persistent renal dysfunction
  - Disfiguring skin lesions

CONCLUSIONS

The immunologic lung diseases comprise a diverse group of disorders ranging from idiopathic etiologies to those related to an underlying autoimmune condition. There is likely a complex interplay between the innate and adaptive arms of the immune system and the pro-fibrotic pathways that lead to the development of these various disease states. Future work should focus on better understanding this relationship and, more importantly, translating these findings to the clinical setting. The role of standard immunosuppressive approaches with the ongoing discovery of novel approaches to targeting the immune system, as well as novel anti-fibrotic therapies, will need to be determined.

ON THE HORIZON

- Our understanding of the idiopathic interstitial pneumonias and autoimmune-related interstitial lung diseases has evolved over time. With increased emphasis being placed on diagnosis due to the implications for treatment and prognosis, we are more accurately phenotyping the diseases. This has led to a better understanding of genetic associations and biologic pathways, as well as more accurate identification of risk factors for disease development and progression.
- However, there are still many areas in need of further understanding and research. In the future, determining if risk factor identification and modification can lead to primary and/or secondary prevention strategies in the treatment of immunologic lung diseases will be essential.
- Other therapies that target the immune system also need to be more carefully studied in diseases outside of idiopathic pulmonary fibrosis, including sarcoidosis.

REFERENCES

1. Desai O, Winkler J, Minasyan M, et al. The role of immune and inflammatory cells in idiopathic pulmonary fibrosis. *Front Med (Lausanne)*. 2018;5:43.
2. Misharin AV, Morales-Nebreda L, Reyfman PA, et al. Monocyte-derived alveolar macrophages drive lung fibrosis and persist in the lung over the life span. *J Exp Med*. 2017;214(8):2387–2404.

3. Zhang L, Wang Y, Wu G, et al. Macrophages: friend or foe in idiopathic pulmonary fibrosis? *Respir Res*. 2018;19(1):170.
4. Schulz C, Gomez Perdiguero E, Chorro L, et al. A lineage of myeloid cells independent of Myb and hematopoietic stem cells. *Science*. 2012;336(6077): 86–90.
5. Parra ER, Kairalla RA, Ribeiro de Carvalho CR, et al. Inflammatory cell phenotyping of the pulmonary interstitium in idiopathic interstitial pneumonia. *Respiration*. 2007;74(2):159–169.
6. Wells AU, Lorimer S, Majumdar S, et al. Fibrosing alveolitis in systemic sclerosis: increase in memory T-cells in lung interstitium. *Eur Respir J*. 1995;8(2):266–271.
7. Wynn TA. Integrating mechanisms of pulmonary fibrosis. *J Exp Med*. 2011;208(7):1339–1350.
8. Kolahian S, Fernandez IE, Eickelberg O, et al. Immune mechanisms in pulmonary fibrosis. *Am J Respir Cell Mol Biol*. 2016;55(3):309–322.
9. Kotsianidis I, Nakou E, Bouchliou I, et al. Global impairment of $CD4^+CD25^+FOXP3^+$ regulatory T cells in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med*. 2009;179(12):1121–1130.
10. Raghu G, Remy-Jardin M, Myers JL, et al. Diagnosis of idiopathic pulmonary fibrosis. An official ATS/ERS/JRS/ALAT clinical practice guideline. *Am J Respir Crit Care Med*. 2018;198(5):e44–e68.
11. Fischer A, Antoniou KM, Brown KK, et al. An official European Respiratory Society/American Thoracic Society research statement: interstitial pneumonia with autoimmune features. *Eur Respir J*. 2015;46(4):976–987.
12. Kaur A, Mathai SK, Schwartz DA. Genetics in idiopathic pulmonary fibrosis pathogenesis, prognosis, and treatment. *Front Med (Lausanne)*. 2017;4:154.
13. Seibold MA, Wise AL, Speer MC, et al. A common MUC5B promoter polymorphism and pulmonary fibrosis. *N Engl J Med*. 2011;364(16): 1503–1512.
14. Olson AL, Gifford AH, Inase N, et al. The epidemiology of idiopathic pulmonary fibrosis and interstitial lung diseases at risk of a progressive-fibrosing phenotype. *Eur Respir Rev*. 2018;27(150).
15. Lynch DA, Sverzellati N, Travis WD, et al. Diagnostic criteria for idiopathic pulmonary fibrosis: a Fleischner Society White Paper. *Lancet Respir Med*. 2018;6(2):138–153.
16. American Thoracic Society/European Respiratory Society International Multidisciplinary Consensus Classification of the Idiopathic Interstitial Pneumonias. This joint statement of the American Thoracic Society (ATS), and the European Respiratory Society (ERS) was adopted by the ATS board of directors, June 2001 and by the ERS Executive Committee, June 2001. *Am J Respir Crit Care Med*. 2002;165(2):277–304.
17. Wilson MS, Madala SK, Ramalingam TR, et al. Bleomycin and IL-1beta-mediated pulmonary fibrosis is IL-17A dependent. *J Exp Med*. 2010;207(3):535–552.
18. Sonnenberg GF, Nair MG, Kirn TJ, et al. Pathological versus protective functions of IL-22 in airway inflammation are regulated by IL-17A. *J Exp Med*. 2010;207(6):1293–1305.
19. Lee JS, Kim EJ, Lynch KL, et al. Prevalence and clinical significance of circulating autoantibodies in idiopathic pulmonary fibrosis. *Respir Med*. 2013;107(2):249–255.
20. Bauer Y, White ES, de Bernard S, et al. MMP-7 is a predictive biomarker of disease progression in patients with idiopathic pulmonary fibrosis. *ERJ Open Res*. 2017;3(1).
21. Ley B, Collard HR, King Jr TE. Clinical course and prediction of survival in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med*. 2011;183(4):431–440.
22. King Jr TE, Bradford WZ, Castro-Bernardini S, et al. A phase 3 trial of pirfenidone in patients with idiopathic pulmonary fibrosis. *N Engl J Med*. 2014;370(22):2083–2092.
23. Richeldi L, du Bois RM, Raghu G, et al. Efficacy and safety of nintedanib in idiopathic pulmonary fibrosis. *N Engl J Med*. 2014;370(22):2071–2082.
24. Hyzy R, Huang S, Myers J, et al. Acute exacerbation of idiopathic pulmonary fibrosis. *Chest*. 2007;132(5):1652–1658.
25. Katzenstein AL, Myers JL, Mazur MT. Acute interstitial pneumonia. A clinicopathologic, ultrastructural, and cell kinetic study. *Am J Surg Pathol*. 1986;10(4):256–267.
26. Ryu JH, Myers JL, Capizzi SA, et al. Desquamative interstitial pneumonia and respiratory bronchiolitis-associated interstitial lung disease. *Chest*. 2005;127(1):178–184.
27. Portnoy J, Veraldi KL, Schwarz MI, et al. Respiratory bronchiolitis-interstitial lung disease: long-term outcome. *Chest*. 2007;131(3):664–671.
28. Katzenstein AL, Fiorelli RF. Nonspecific interstitial pneumonia/fibrosis. Histologic features and clinical significance. *Am J Surg Pathol*. 1994;18(2):136–147.
29. Xu W, Xiao Y, Liu H, et al. Nonspecific interstitial pneumonia: clinical associations and outcomes. *BMC Pulm Med*. 2014;14:175.
30. Drakopanagiotakis F, Paschalaki K, Abu-Hijleh M, et al. Cryptogenic and secondary organizing pneumonia: clinical presentation, radiographic findings, treatment response, and prognosis. *Chest*. 2011;139(4):893–900.
31. Kamen DL, Strange C. Pulmonary manifestations of systemic lupus erythematosus. *Clin Chest Med*. 2010;31(3):479–488.
32. Ta R, Celli R, West AB. Diffuse alveolar hemorrhage in systemic lupus erythematosus: histopathologic features and clinical correlations. *Case Rep Pathol*. 2017;2017:1936282.
33. Tselios K, Gladman DD, Urowitz MB. Systemic lupus erythematosus and pulmonary arterial hypertension: links, risks, and management strategies. *Open Access Rheumatol*. 2017;9:1–9.
34. Mok MY, Tsang PL, Lam YM, et al. Bosentan use in systemic lupus erythematosus patients with pulmonary arterial hypertension. *Lupus*. 2007;16(4):279–285.
35. Lin E, Limper AH, Moua T. Obliterative bronchiolitis associated with rheumatoid arthritis: analysis of a single-center case series. *BMC Pulm Med*. 2018;18(1):105.
36. Raimundo K, Solomon JJ, Olson AL, et al. Rheumatoid arthritis-interstitial lung disease in the United States: prevalence, incidence, and healthcare costs and mortality. *J Rheumatol*. 2019;46(4):360–369.
37. Juge PA, Lee JS, Ebstein E, et al. MUC5B promoter variant and rheumatoid arthritis with interstitial lung disease. *N Engl J Med*. 2018;379(23):22092219.
38. Flaherty KR, Wells AU, Cottin V, et al. Nintedanib in progressive fibrosing interstitial lung diseases. *N Engl J Med*. 2019;381(18):1718–1727.
39. van den Hoogen F, Khanna D, Fransen J, et al. 2013 classification criteria for systemic sclerosis: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. *Ann Rheum Dis*. 2013;72(11):1747–1755.
40. Highland KB. Recent advances in scleroderma-associated pulmonary hypertension. *Curr Opin Rheumatol*. 2014;26(6):637–645.
41. Grunewald J, Grutters JC, Arkema EV, et al. Sarcoidosis. *Nat Rev Dis Primers*. 2019;5(1):45.
42. Arkema EV, Cozier YC. Epidemiology of sarcoidosis: current findings and future directions. *Ther Adv Chronic Dis*. 2018;9(11):227–240.
43. Greaves SA, Atif SM, Fontenot AP. Adaptive immunity in pulmonary sarcoidosis and chronic beryllium disease. *Front Immunol*. 2020;11:474.
44. Moller DR, Rybicki BA, Hamzeh NY, et al. Genetic, immunologic, and environmental basis of sarcoidosis. *Ann Am Thorac Soc*. 2017;14(suppl 5): S429–S436.
45. Levy A, Hamzeh N, Maier LA. Is it time to scrap Scadding and adopt computed tomography for initial evaluation of sarcoidosis? *F1000Res*. 2018:7.
46. Rosen Y. Pathology of sarcoidosis. *Semin Respir Crit Care Med*. 2007;28(1):36–52.
47. Selman M, Pardo A, King Jr. TE. Hypersensitivity pneumonitis: insights in diagnosis and pathobiology. *Am J Respir Crit Care Med*. 2012;186(4): 314–324.

73

Chronic Obstructive Pulmonary Disease and Emphysema

Tillie-Louise Hackett, Francesca Polverino, and Farrah Kheradmand

The large-scale industrial commercialization, advertisement, and addictive nature of nicotine led to the increase in smoking prevalence in the early to mid-twentieth century. In the following decades, after the surge in mass production and consumption of pre-packaged cigarettes, chronic obstructive pulmonary disease (COPD) became one of the significant causes of death in the world, a devastating legacy that until recently was not matched by progress in understanding disease pathophysiology. Observational studies of smokers in the 1960s that assessed lung function, smoking habits, and acute infection episodes ultimately established the causal link between smoking and COPD.

Historically physicians labeled smokers as "blue bloaters" when they presented with chronic cough and increased mucous production, which characterized a clinical manifestation consistent with chronic bronchitis. Alternatively, smokers were labeled as "pink puffers" when presenting with barreled chest, lung hyperinflation, and flat diaphragms consistent with the lung hyperinflation of emphysema. However, with the advent of comprehensive lung-based immune phenotyping, radiographic, and physiological studies, newer insights into smokers' different endotypes beyond the original description of "blue blotters" or "pink puffers" have emerged. Indeed, patients with chronic bronchitis can present with emphysematous lung changes, and harbor autoreactive immune cells, while the inflammatory signature in those with significant reversible airway obstruction can resemble asthmatic inflammation and is known as asthma/COPD overlap (ACO).[1]

Although most endotypes of COPD invariably result in poor oxygenation and dyspnea on exertion, clinically, they display highly variable disease progression from age-appropriate lung functional decline to rapid, unrelenting deterioration that leads to end-stage COPD and early death. Our current pharmacological interventions rely heavily on bronchodilation and exacerbation prevention, which can provide symptomatic relief but fails to reverse the underlying course of pathological changes in smokers with COPD. The small airways that offer little resistance in healthy lungs are the primary site of obstruction to airflow, as evidenced by post mortem examination of emphysematous lungs. Thereby the term "small airway disease" was introduced because smoking affected both the smaller bronchi, defined by the presence of cartilage in their walls, and the bronchioles. Some of the initial descriptions of COPD were limited to the location and patterns of cigarette-smoke-induced lung damage. However, newer advancements (e.g., imaging, molecular genetics, animal models, etc.) have provided a clearer picture of the pathophysiology of smoking-induced lung diseases, some of which will be highlighted in this chapter.

CHRONIC OBSTRUCTIVE PULMONARY DISEASE DEFINITION AND EPIDEMIOLOGY

The umbrella term COPD is defined by the Global Initiative for Obstructive Lung Disease (GOLD) as a "common, preventable and treatable disease, characterized by persistent respiratory symptoms and airflow limitation that is due to airway and/or alveolar abnormalities usually caused by significant exposure to noxious particles or gases".[2] Currently, nearly 10% of the population older than 40 years (11.8% for men and 8.5% for women) are estimated to have COPD.[3] COPD is one of the few chronic diseases where the incidence continues to rise, with the estimated worldwide prevalence increasing over the last 30 years.[4] The increases in global morbidity and mortality are largely attributable to some of the most populous countries in the world (e.g., China, India, Indonesia, etc.), where greater than 50% of men smoke or are exposed to high-level particulate air pollution, complicated by recurrent respiratory infections.[2] Epidemiological studies on the prevalence and morbidity of COPD suggest that the disease burden is underestimated because of underdiagnosis, and smokers often do not seek medical help until the disease is moderately advanced. COPD poses a substantial economic burden on healthcare services globally due to a large number of affected individuals and the lack of disease-modifying therapies.

Clinical Presentation of Chronic Obstructive Pulmonary Disease

Patients with COPD experience three cardinal symptoms: (1) dyspnea at rest or upon physical activity, (2) increased sputum production, and (3) chronic cough. Clinically, the gold standard for diagnosing COPD is the detection of airflow obstruction using spirometry. The GOLD recommendations define airflow obstruction as a post-bronchodilator forced expiratory volume in 1 second (FEV_1) ratio over forced vital capacity (FVC) of less than 0.7.[2] Following a diagnosis, the severity of COPD is further categorized by the extent of FEV_1 airflow limitation (Table 73.1)

Patients with a more severe type of disease often experience exacerbations, defined as a sustained worsening of respiratory condition, from their stable state and beyond normal day-to-day variations. Exacerbations are thought to be primarily triggered by respiratory viral infections, but bacterial or fungal infections and environmental factors may also initiate these events to a lesser degree.[2] These events are acute in onset and may warrant additional treatment or hospitalization in patients with underlying COPD. Longitudinal data have shown that exacerbation

TABLE 73.1 The GOLD Classifications for the Severity of Airflow Limitation in COPD

Classification of Severity of Airflow Limitation in COPD		
GOLD Stages	**COPD Severity**	**FEV_1 (% of predicted)**
GOLD-1	Mild limitation	≥80%
GOLD-2	Moderate limitation	50–79
GOLD-3	Severe limitation	30–49
GOLD-4	Very severe limitation	<30

rates in the first year of follow-up are 0.85 per person for patients with GOLD-2, 1.34 for patients with GOLD-3, and 2.00 for patients with GOLD-4.[5] Thus, the COPD recommendations have been revised to combine COPD assessment tools incorporating spirometry, with patient-reported outcomes/symptoms and exacerbation rates.

A revised grading system has been used to identify COPD patients in four distinct classes ranging from A to D (https://goldcopd.org). In addition to using post-bronchodilator FEV_1 measurements, this more comprehensive approach adds the impact of clinical symptoms using the modified Medical Research Council (mMRC) dyspnea scale and the COPD Assessment Test (CAT). The final additional stratification uses the number of COPD exacerbations in the previous 12 months. One significant advancement of the newer system is that the A-D grading factors in features related to clinical presentations in smokers, where measurement of FEV_1 alone may not account for disease severity or prediction of the therapeutic needs of COPD subtypes (Table 73.2).[6] This more comprehensive classification scheme allows for more reliable predictions as to COPD severity.

CHRONIC OBSTRUCTIVE PULMONARY DISEASE CLINICAL SUBPHENOTYPES

Chronic Bronchitis

Chronic bronchitis is defined as persistent cough with sputum production for at least three months over two consecutive years. Increased mucus production and secretion are attributed to mucous gland hypertrophy and bronchial epithelial goblet cell metaplasia. The Reid Index has historically been used to quantify the thickness of the mucous glands with airway wall thickness. Chronic bronchitis is associated with inflammation within the central airways (defined as larger than 4 mm in internal diameter), transformation into the mucus-producing glands, and thickening of the bronchial wall associated with extracellular matrix deposition. While most smokers will develop chronic bronchitis during their lifetime, the presence of chronic bronchitis alone does not predict the development or progression of airflow limitation. These observations suggest that the cough and sputum production that characterize chronic bronchitis can occur in parallel or independently of the disease process responsible for airflow obstruction in COPD patients.

Emphysema

Emphysema is characterized by abnormal enlargement of airspaces distal to the terminal bronchioles accompanied by the destruction of their walls and without obvious fibrosis. A more quantitative clinical noninvasive thoracic CT imaging has been used to validate radiographic correlations with regional lung function and to quantify emphysema (<−950 Hounsfield Units (HU). However, the spatial resolution of clinical CT scans, 800 to 1000 μm, does not permit analysis of the smallest conducting airways or parenchymal structures. MicroCT has been used to image tissue samples from explanted lungs from patients with very severe (GOLD-4) COPD and provided the first evidence that terminal bronchioles are destroyed in end-stage COPD, even in regions of the lung with no emphysema. Koo et al.[7] demonstrated that terminal and first-generation respiratory (e.g., transitional) airways and bronchioles are lost in lung tissue in which no emphysematous destruction is present. These findings indicate that small airways disease could represent an early pathological feature of mild and moderate COPD and that lung destruction may occur in stages. Three distinct histopathological forms of emphysema have been described based on the location of lung parenchymal loss compared to normal lung (Fig. 73.1A):

1. *Centrilobular emphysema (CLE)* is the most common form of emphysema primarily associated with cigarette smoking. CLE often develops first in the upper lobes and/or the superior segment of the lower lobes of the lung. As the disease progresses, it can be radiologically detected throughout the lung. CLE is characterized by enlargement of the centriacinar space, occurring in the proximal respiratory bronchioles, often leaving the alveolar ducts and sacs intact (Fig. 73.1B). MicroCT-based research studies that measure emphysema are only performed using human lung cadaveric tissue.
2. *Panlobular emphysema (PLE)* primarily occurs in smokers with α1-antitrypsin (A1AT) deficiency. Abnormal A1AT or complete lack of expression fails to counteract or inhibit neutrophil elastase, leading to widespread destruction of the connective tissue network, including elastin fibers. As illustrated in Fig. 73.1C, PLE is characterized by uniform dilation of the airspaces from the respiratory bronchioles to the alveoli, resulting in homogeneous destruction of the secondary lobules (acini).
3. *Paraseptal emphysema (PSE)* can present alongside lung fibrosis and can coexist with the other subtypes of emphysema. As shown in Fig. 73.1D, PSE is characterized by enlargement of the airspace at the periphery of the acini. Large paraseptal lesions or bullae favor a sub-pleural site of formation, whereas milder changes are often observed deeper within the lungs adjacent to the connective tissue septae. Patients with

TABLE 73.2 ABCD Criteria for the Severity of COPD

Classifications	Exacerbation History	mMRC/CAT score	Risk/Symptoms
A	<1 or no hospitalization	0–1/CAT<10	Low risk, fewer symptoms
B	<1 or no hospitalization	>2/CAT>10	Low risk, more symptoms
C	>2 or >1 hospitalization	0–1/CAT<10	High risk, fewer symptoms
D	>2 or >1 hospitalization	>2/CAT>10	High risk, more symptoms

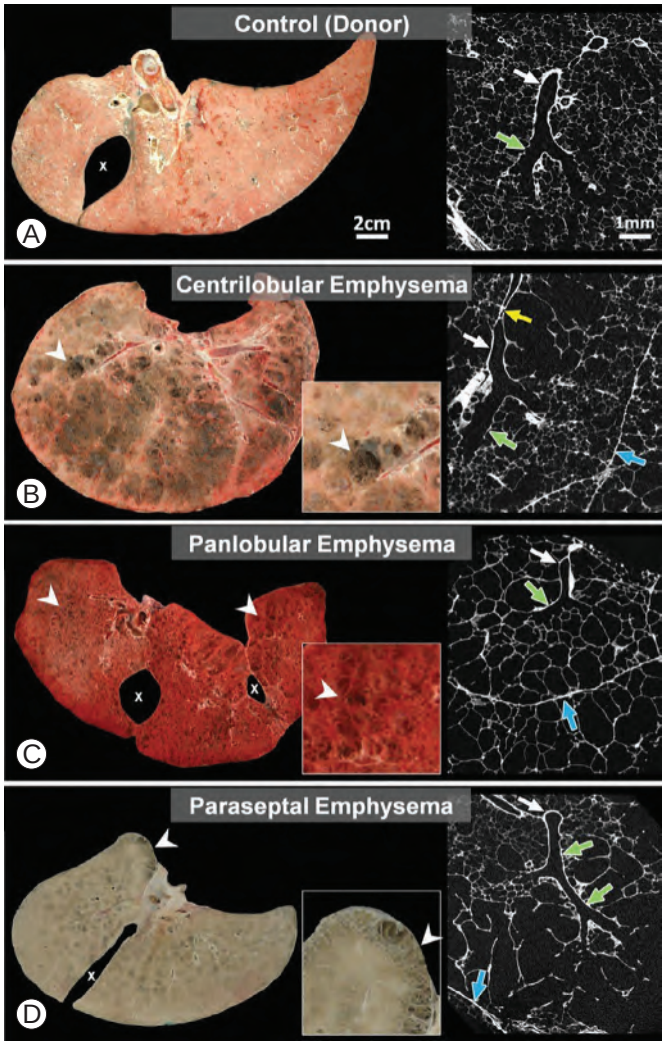


FIG. 73.1 Frozen Lung Slices and microCT Images Illustrating the Subtypes of Emphysema. (A) Normal appearance of parenchyma in the control lung slice and the microCT image shows a terminal bronchiole *(white arrow)* connected to respiratory bronchioles *(green arrow)* and intact, normal alveoli. (B) Advanced centrilobular destruction in the lung slice *(white arrows)* and microCT scan showing narrowing of the terminal bronchiole *(yellow arrow)*, dilation, and destruction of the proximal respiratory bronchioles *(green arrow)*, with sparing of the alveoli adjacent to the septa *(blue arrow)*. (C) Contrastingly, panlobular emphysema shows relatively mild destruction on the gross specimen, and the microCT scan shows uniform destruction of the alveoli proceeding up to the lobular septa *(blue arrow)*. The terminal *(white arrow)* and respiratory *(green arrow)* bronchioles appear normal. (D) paraseptal emphysema shows characteristic lesions *(arrowheads)* beneath the pleural surface in the lung slice. The microCT scan shows alveoli adjacent to the lobular septa are dilated and destroyed *(blue arrow)*, with preservation of the center lobule. The terminal *(white arrow)* and respiratory (green arrow) bronchiole appear normal. Fig. 73.1 is reproduced from a statement from the Fleischner society Lynch et al., 2015.[8]

paraseptal emphysema are typically asymptomatic, but the condition is associated with spontaneous pneumothoraces in young adults.

Chronic Bronchitis with Emphysema Features

Histological studies of explanted COPD lung tissue and airway biopsies have shown that airway remodeling and emphysema can occur in the large airways and lung parenchyma. Airway remodeling is the term used to describe structural changes within airway walls caused by inhalational injury followed by induction of chronic inflammation. Small airways show an accumulation of mucus in the lumen, which contributes to luminal obstruction. Mucus exudates may be caused by aspiration of mucus from the larger proximal bronchi or by secretion of mucins by goblet cells in the small airway epithelium itself. Inflammatory cells are recruited from the circulation to the small airways and accumulate within the airway walls and appear in inflammatory exudates within the lumen. Furthermore, with increasing disease severity, there are increased numbers of lymphoid aggregates that contain follicles[9] in the airway wall and scattered throughout the parenchyma, indicating activation of the adaptive immune system, which underlies lung destruction.[10] Each of the aforementioned pathological features shown in small airways has been consistently found in many studies and is associated with the severity of airflow obstruction in COPD.

Clinical Aspects and Diagnosis of Chronic Obstructive Pulmonary Disease

COPD is a highly heterogeneous disease; therefore, lung disease in patients may not be completely captured by the original GOLD classification or the revised ABCD description. These shortcomings in identifying COPD patients' specific characteristics have contributed to the failure to capture meaningful outcomes in multiple clinical trials. For example, it has become more evident, although not widely accepted, that a more precise measurement of lung parenchymal abnormality should include CT scanning, which can provide an exact location and quantification of lung parenchymal destruction in smokers.[11] A clear definition of the lung (airways and parenchyma) changes with or without symptoms is necessary to provide a better COPD characterization in smokers. Specifically, a chest CT can distinguish smokers who have airway disease (AD), parenchymal destruction (emphysema), or a combination of the two. Importantly, nearly 20% of patients with normal lung function tests (e.g., FEV_1 or FVC that are used for spirometric criteria in COPD diagnosis) have emphysema based on chest CT scan.[12] These findings indicate that physiological lung measurements in COPD diagnosis are insufficient for detecting smoke-induced lung pathologies. Indeed, careful examination of immune cells in the lungs has confirmed that the pathobiology of emphysema and airway disease are distinct and represent different pathogenic processes.

Obstructive and Emphysema Phenotypes

Proper identification of patients with the obstructive phenotype (e.g., a predominance of AD) is clinically relevant because patients with severe emphysema have more exacerbations and

are more likely to be hospitalized and die within three years as compared to patients with predominant AD.[13] Furthermore, from a clinical standpoint, patients with emphysema, compared with patients with AD, have an accelerated loss of lung function and enhanced loss of tissue mass in several other body compartments and show higher mortality than patients without emphysema.[14] In clinical practice, both visual and quantitative CT scans have become instrumental in the diagnosis and management of patients with COPD. Quantitative CT provides useful information regarding emphysema, airways, and air trapping and provides a means of objectively characterizing and following these pathologic processes. CT has been instrumental in showing that small airways disease/loss is evident even in low COPD stages (GOLD 1-2) and correlates with disease progression. Although microCT can measure the extent of emphysema in human lung tissue, it is exclusively used for research. However, other radiographic images, such as a high-resolution CT scan (HRCT), allow an in-depth assessment of emphysema and air trapping subphenotypes in COPD patients. In two large cohorts, (COPDgene (www.copdgene.org) and SPIROMICS (www.spiromics.org), HRCT scans were used to differentiate COPD subphenotypes. Four main subphenotypes that have emerged using radiographic studies of two large COPD cohorts are:

1. Chronic airflow obstruction defined by a low post-bronchodilator FEV_1/FVC
2. HRCT-defined emphysema (low attenuation at total lung capacity)
3. Hyperinflation or gas trapping on HRCT, defined as low attenuation of the lungs at low lung volume (functional residual capacity)
4. Airway inflammation/disease (bronchial wall thickening on HRCT; bronchiectasis).

Thus, the HRCT scan can provide more in-depth phenotyping of the lungs in patients with COPD. However, a comparison of chronic airflow obstruction versus emphysema showed striking disagreement between spirometric and HRCT-defined emphysema (regardless of disease severity),[15] further highlighting the insensitive nature of some of the commonly used diagnostic tools in COPD. HRCT scan is associated with high costs, radiation exposure, and a high incidence of false-positive findings (e.g., indeterminant pulmonary nodules). However, low-dose CT (LDC) scans could provide accurate information about lung parenchyma disease and will undoubtedly become the first-line choice in phenotyping emphysema in COPD patients.

Asthma-Chronic Obstructive Pulmonary Disease Overlap

Asthma-COPD overlap (ACO) in former or active smokers is defined by persistent airflow limitation with several clinical features usually associated with asthma (e.g., reversible obstructive airway disease) and those associated with COPD (chronic productive cough).[16] However, whether or not it represents a distinct clinical entity *per se* is still debated. A consensus statement on the clinical definition of ACO includes the following six criteria:

1. Three major (persistent airflow limitation, tobacco smoking, and previous asthma or reversibility >400 mL FEV_1), and,
2. Three minor (history of atopy or rhinitis, at least two positive bronchodilator tests, and ≥300 blood eosinophils per μL).

CT shows a higher emphysema index and a more significant upper-zone-predominant distribution of emphysema in patients with ACO than non-smokers with asthma. However, no specific pattern(s) of radiographic findings has been conclusive in the diagnosis of ACO.

Early Chronic Obstructive Pulmonary Disease

The diagnosis of "Early COPD" has been used to indicate the age of COPD onset in smokers who present with airway obstruction and are younger than 50 years of age.[17] This clinical distinction is different from detecting mild COPD in older individuals who may or may not have a progressive disease. Chronic lung conditions in young adults may begin due to prenatal or perinatal factors such as premature birth and bronchopulmonary dysplasia, or postnatal factors (e.g., environmental exposures, maternal smoke, childhood infections), which can cause persistent airflow limitation. However, little is known about those who develop COPD in their early adulthood because COPD is most often associated with smokers past the sixth or seventh decade of life. Thus, there is a critical need for improved distinction between the various phenotypes and pathophysiological mechanisms that begin in childhood and often lead to chronic airflow limitation and early diagnosis of COPD. In clinical practice, physicians rely heavily on pulmonary function tests, even though these are global measures that are not sensitive enough to detect an early obstruction and/or minor changes over time. The combination of proton MRI with hyperpolarized-gas magnetic resonance imaging (MRI) is a noninvasive imaging modality that provides multi-faceted structural and functional information, maps of gross parenchymal abnormalities, regional ventilation, alveolar-airspace size, and gas exchange abnormalities (^{129}Xe MRI). These techniques could provide a useful and safe tool to follow patients with pulmonary childhood diseases longitudinally who have a predisposition to developing fixed airflow limitation.[18]

Immunological Mechanisms of Chronic Obstructive Pulmonary Disease

Chronic inflammation is a central and well-recognized feature of the underlying lung disease pathogenesis in COPD.[2] Some of the initial studies that explored lung inflammation used immunohistochemistry (IHC) in biopsies or resected lung tissue to identify immune cell profiles in the lungs of COPD patients. The cross-sectional IHC-based examination of the lung inflammation in different stages of COPD captures one time-point in the disease pathogenesis; however, the progressive nature of COPD does not follow an exact temporal sequence marching from GOLD stage 1 to 4. Thus, the lack of longitudinal human lung tissue sampling makes it less clear whether the development of progressive disease is associated with innate immune responses that precede adaptive immune responses. Similarly, it is unclear whether small airway disease precedes emphysematous changes in COPD. Cigarette smoke can induce epigenetic reprogramming, remodeling, and hyperplasia of airway basal cells, the stem/progenitor cells that are central to pulmonary host defense, even in the absence of inflammatory cell infiltration.[19] Small airway epithelium basal/progenitor cells from smokers, with and without COPD, are limited in their ability to regenerate a fully differentiated epithelium and present abnormalities in airway epithelial junction formation. These findings support the hypothesis that decreased self-renewing cells and airway basal progenitor cells may relate to reduced lung function in smokers.

Little is known about the initial steps in the activation and recruitment of the innate and adaptive immune cells in COPD. It was initially postulated that innate immune inflammation drives the mild stages of COPD, whereas, in more advanced COPD, adaptive T and B cell responses become predominant. However, the recruitment of innate and adaptive immune responses might not be sequential. Specifically, cigarette smoke triggers pattern recognition receptors to release "danger signals" that act as ligands to Toll-like receptors, triggering the production of host-derived damage-associated molecular patterns (DAMPs) and cytokines and activating innate immune cells such as epithelial cells and macrophages. Levels of several DAMPs, including S100 proteins, defensins, and high-mobility group box-1 (HMGB1), are increased in extracellular lung fluids of COPD patients. Macrophages that are initially recruited into the lungs of smokers have dysregulated capacity to secrete proinflammatory mediators and proteases. However, they can induce oxidative stress, engulf microbes and apoptotic cells, and express surface and intracellular markers. Lung dendritic cell (DC) subsets have also emerged as critical mediators in both pathological processes from the earliest to the later stages of COPD pathogenesis. Specifically, there is strong emerging evidence that a subset of smokers develops autoimmune inflammation that is linked to the loss of tolerance to self and is driven in part by the activation of CD1a+ lung DCs. Emphysema, a major COPD subphenotype, is characterized by specific innate and adaptive immune profiles and is described below.

Evidence for Autoimmunity in Emphysema

Recent basic and translational studies have highlighted that the immunopathogenic hallmarks of emphysema are very distinct from AD. Similarly, network analysis of lung transcriptomics distinguished bronchiolitis and emphysema based on distinct molecular signatures, independent of airflow limitation severity. The pathobiology of emphysema in humans is associated with the activation of adaptive immunity, whereby antigen-specific CD4, CD8 T cells, and B cells have been identified. Studies using human lung tissue have identified extramedullary lymphoid follicle formation around the airways of active or former smokers characterized by the presence of autoreactive immune cells such as T-helper type 1 (Th1) expressing IFNγ, and Th17 cells expressing IL-17A, that correlate with emphysema severity. Specifically, emphysematous lung harbors antigen-specific Th1 and Th17 cells that secrete cytokines and chemokines, which further enhance the release of matrix metalloproteinases as compared to smokers without emphysema.

In preclinical models of cigarette smoke-induced emphysema, IL-17A, IFNγ expressing CD4 T cells (Th17 and Th1 respectively), and IFNγ expressing CD8 T cells are increased in the lungs. Preclinical studies of emphysema have shown that Th17 cells are partially responsible for the persistence of lung inflammation, as IL-17A signaling can induce and stabilize the transcripts of downstream chemokines, and mice lacking IL-17A are protected against emphysema development. Both IL-17 and IFNγ can induce proteolytic enzyme release from macrophages and neutrophils. $CD1a^+$ conventional dendritic cells (cDCs) are the primary lung antigen-presenting cells that induce Th17 cell differentiation through the action of a secreted cytokine, osteopontin (*Spp1*). To further support a role for adaptive immunity in the pathogenesis of emphysema, mice lacking IL17A or its receptor are highly protected from cigarette smoke-induced emphysema, and *Spp1* knockout mice manifest attenuated emphysema. These studies suggest redundant IL-17A induction pathways that include osteopontin but expand to other possible proinflammatory mediators. In contrast to the induction of proinflammatory mediators, a reduction in anti-inflammatory factors such as peroxisome proliferator-activated receptor gamma (PPARG/Pparg) and C1Q in lung innate immune cells (DCs, and macrophages) in humans and mice exposed to chronic smoke have been reported. In a preclinical therapeutic model of emphysema, exogenous restoration of Pparg or C1Q resulted in the attenuation of lung inflammation and emphysema. These studies provide hope that emphysema could become a treatable disease in the future by targeting specific proinflammatory pathways.[20]

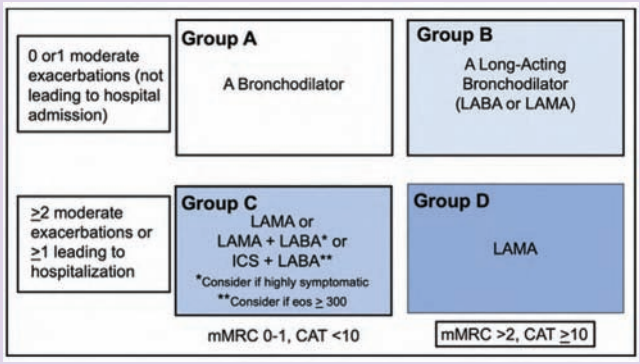
Evidence for Type 2 Inflammation in Asthma Chronic Obstructive Pulmonary Disease Overlap

Increased sputum eosinophils are present in both the stable and exacerbation phase of patients with COPD, implying the potential role of eosinophils in the pathogenesis of COPD. Eosinophilia is generally defined as greater or equal to 2% eosinophils in either blood or sputum or an absolute blood eosinophil count of 0.34×10^9 cells per liter. A subset of COPD patients present greater than 2% blood eosinophil count, which has been called "high eosinophil-COPD." Consistently in asthma and COPD, sputum eosinophilia is associated with an excellent response to corticosteroid therapy and strategies aimed to normalize sputum eosinophils to reduce exacerbation frequency and severity. Thus, the new GOLD guidelines in COPD suggest that high blood eosinophil counts (≥300 eosinophils/μL) could be used to identify patients with a greater likelihood of a beneficial response to inhaled corticosteroids (ICS). Eosinophil inflammation might be a common mechanism underlying the pathogenesis of ACO. However, eosinophils are recruited in the airway and play a key role in eradicating fungi,[21] further highlighting the need to identify mucosal fungal colonization in the airways of patients with ACO. Future studies are needed to clarify our understanding of fungal infections, eosinophil recruitment to the airway, and the consequence of eosinophil infiltrate to develop new therapies for ACO.[22]

Therapy

Evidence-based clinical guidelines for the diagnosis and management of COPD are now widely available. However, several studies on COPD have demonstrated that clinical practice may deviate significantly from guideline recommendations. The most recent clinical trials and longitudinal studies have repeatedly shown that current treatments for COPD (inhaled bronchodilators and steroids) can only relieve symptoms and reduce the frequency of exacerbations but do not improve disease outcomes over time.[23] Therefore, over 3 million COPD-related deaths occur per year worldwide.[24] To date, progress has been hampered by the heterogeneity of disease mechanisms and phenotypic expression. Perhaps more importantly, the presence of spirometrically detected airflow obstruction is not sensitive enough to detect early abnormalities in lung health or capture the endotypes and phenotypic heterogeneity of smoking-related lung disease.

TABLE 73.3 The GOLD Guidelines for Pharmacologic COPD Treatment



Abbreviations: Eos: blood eosinophil count in cells per microliter; LABA: long-acting beta-agonists; LAMA: long-acting muscarinic antagonists; ICS: inhaled corticosteroids; mMRC: modified Medical Research Council dyspnea questionnaire; CAT™: COPD Assessment Test™

Several important lifestyle changes are appropriate for all patients with COPD, including smoking cessation, avoidance of environmental pollutants, physical rehabilitation, vaccination against respiratory infections, and education about medication usage and inhaler technique. Smoking cessation can achieve a reduction in terms of all-cause mortality, even if those interventions are successful only in a minority of patients. Similarly, pulmonary rehabilitation that includes exercise, promotion of healthy behaviors, education, adherence to medication, and psychological support, has been shown to improve life quality, decrease dyspnea, decrease health care utilization and eventually reduce hospitalization and mortality of COPD.

KEY CONCEPTS

Immunopathogenesis of COPD

- Airway disease is characterized by (1) infiltration of innate and adaptive immune cells, and in particular neutrophils, monocytes/macrophages, and CD4, and CD8 T lymphocytes; (2) airway wall thickening, epithelial metaplasia, goblet cell hypertrophy, and smooth muscle hyperplasia of the walls of airways that are less than 2 mm in diameter (small airways)
- Emphysema is characterized by (1) an imbalance between protease and antiprotease activity due to chronic inflammation that can promote oxidative stress; and (2) activation of the innate (monocytes/macrophages, $CD1a^+$ DCs), and increase in adaptive immune (CD4, CD8 T, and B lymphocytes) with autoimmune features such as elastin-specific CD4 T cells, and size of B-cell-rich lymphoid follicles, and in circulating and pulmonary B-cell-secreting autoantibodies.
- Abnormal alveolarization during lung development in early life due to maternal smoking, prematurity, chronic infection, or genetic factors (e.g., A1AT) contributing to the lifetime burden of emphysema (Early COPD)
- Environmental factors. Inhalation of environmental pollution resulting from incomplete combustion of organic matter (indoor cooking with biomass fuel, industrial factory air pollution, carbon black, etc.) can promote chronic lung inflammation and emphysema.

The pharmacologic COPD treatment recommended by the current GOLD guidelines is mainly based on the symptoms and exacerbation risk summarized in Table 73.3.

Abbreviations: Eos: blood eosinophil count in cells per microliter; LABA: long-acting beta-agonists; LAMA: long-acting muscarinic antagonists; ICS: inhaled corticosteroids; mMRC: modified Medical Research Council dyspnea questionnaire; CAT™: COPD Assessment Test™.

In the 21st century, clinical management of COPD is moving towards precision-based medicine and individualized treatment approaches that incorporate multiple relevant imaging and physiological approaches to endotype patients. COPD is most often diagnosed when smokers present with shortness of breath, and changes in lung function are highly variable, thus limiting the success of most randomized phase two and three clinical trials. Therefore, future studies are needed to better classify young adults based on their underlying airway, emphysematous, and gas-exchange abnormalities identified with sensitive non-invasive technologies, such as lung MRI, and to establish earlier and patient-tailored therapies that could prevent progression and help alleviate the burden of COPD. There are exciting new goals to initiate global and multidisciplinary COPD commissions to drive new strategies for the prevention, detection, assessment, and treatment of COPD.

ON THE HORIZON

- Future studies are needed to define the clinical COPD endotypes and their circulating and pulmonary cellular signature that can benefit from future therapies targeting specific cellular pathways.
- Examples of selected clinical trials focused on interrupting key immunological pathways or repairing lungs that are registered in the clinicaltrials.gov (i.e., Dupilumab (a monoclonal antibody against IL-4 receptor), FP-025 (an MMP12 inhibitor), umbilical cord mesenchymal stem cells, Benralizumab (a monoclonal antibody against a-5 chain of the IL-5 receptor), RETHINC, etc.).
- Focused research on the importance of disease susceptibility and factors responsible for initiation and perpetuation of emphysema.
- Development of immunologically based diagnostic and prognostic strategies for obstructive and/or destructive features of smoking.
- Clinical studies that reassess the validity of the predominant use of targeted anti-inflammatory approaches.

REFERENCES

1. Vestbo J. COPD: definition and phenotypes. *Clinics in Chest Med*. 2014;35:1–6.
2. Vogelmeier CF, Criner GJ, Martinez FJ, et al. Global strategy for the diagnosis, management, and prevention of chronic obstructive lung disease 2017 report: GOLD executive summary. *Arch Bronconeumol*. 2017;53:128–149.
3. Buist AS, McBurnie MA, Vollmer WM, et al. International variation in the prevalence of COPD (the BOLD Study): a population-based prevalence study. *Lancet*. 2007;370:741–750.
4. Ruparel M, Quaife SL, Dickson JL, et al. Prevalence, symptom burden and under-diagnosis of chronic obstructive pulmonary disease in a lung cancer screening cohort. *Ann Am Thorac Soc*. 2020;17:869–878.
5. Hurst JR, Vestbo J, Anzueto A, et al. Susceptibility to exacerbation in chronic obstructive pulmonary disease. *The New England J Med*. 2010;363:1128–1138.
6. Cabrera Lopez C, Casanova Macario C, Marin Trigo JM, et al. Comparison of the 2017 and 2015 Global Initiative for Chronic Obstructive Lung Disease Reports. Impact on Grouping and Outcomes. *Am J Respir Crit Care Med*. 2018;197:463–469.

7. Koo HK, Vasilescu DM, Booth S, et al. Small airways disease in mild and moderate chronic obstructive pulmonary disease: a cross-sectional study. *Lancet Respir Med*. 2018;6:591–602.
8. Lynch DA, Austin JH, Hogg JC, et al. CT-definable subtypes of chronic obstructive pulmonary disease: a statement of the fleischner society. *Radiology*. 2015;277:192–205.
9. Hogg JC, Chu F, Utokaparch S, et al. The nature of small-airway obstruction in chronic obstructive pulmonary disease. *N Engl J Med*. 2004;350:2645–2653.
10. Xu C, Hesselbacher S, Tsai CL, et al. Autoreactive T cells in human smokers is predictive of clinical outcome. *Front Immunol*. 2012;3:267.
11. Tanabe N, Vasilescu DM, Hague CJ, et al. Pathological comparisons of paraseptal and centrilobular emphysema in COPD. *Am J Respir Crit Care Med*. 2020;202:803–811.
12. Hesselbacher S, Ross R, Schabath M, et al. Cross-sectional analysis of the utility of pulmonary function tests in predicting emphysema in ever-smokers. *Int J Environ Research Public Health*. 2011;8(5):1324–1340.
13. Mullerova H, Maselli DJ, Locantore N, et al. Hospitalized exacerbations of COPD: risk factors and outcomes in the ECLIPSE cohort. *Chest*. 2015;147:999–1007.
14. Oelsner EC, Hoffman EA, Folsom AR, et al. Association between emphysema-like lung on cardiac computed tomography and mortality in persons without airflow obstruction: a cohort study. *Ann Intern Med*. 2014;161:863–873.
15. Enright P. HRCT-defined emphysema is not COPD to be treated with inhalers. *Thorax*. 2014;69:401–402.
16. Global Initiative for Asthma. *Global Strategy for Asthma Management and Prevention* (updated 2019). 2019. https://ginasthma.org/wp-content/uploads/2019/06/GINA-2019-main-report-June-2019-wms.pdf
17. Martinez FJ, Han MK, Allinson JP, et al. At the Root: Defining and Halting Progression of Early Chronic Obstructive Pulmonary Disease. *Am J Respir Crit Care Med*. 2018;197:1540–1551.
18. Polverino F, Hysinger EB, Gupta N, et al. Lung MRI as a potential complementary diagnostic tool for early COPD. *Am J Med*. 2020;133:757–760.
19. Crystal RG. Airway basal cells. The "smoking gun" of chronic obstructive pulmonary disease. *Am J Respir Crit Care Med*. 2014;190:1355–1362.
20. Kheradmand F, Shan M, Xu C, Corry DB. Autoimmunity in chronic obstructive pulmonary disease: clinical and experimental evidence. *Exp Rev Clin Immunol*. 2012;8:285–292.
21. Porter P, Susarla SC, Polikepahad S, et al. Link between allergic asthma and airway mucosal infection suggested by proteinase-secreting household fungi. *Mucosal Immunol*. 2009;2:504–517.
22. Casanova C, Celli BR, de-Torres JP, et al. Prevalence of persistent blood eosinophilia: relation to outcomes in patients with COPD. *Eur Respir J*. 2017:50.
23. Singh D, Agusti A, Anzueto A, et al. Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Lung Disease: the GOLD science committee report 2019. *Eur Respir J*. 2019:53.
24. Vestbo J, Lange P. COPD drugs: the urgent need for innovation. *The Lancet Respiratory Med*. 2014;2:14–15.

Immunologic Ocular Disease

Michel M. Sun and Lynn K. Gordon

A broad overview of ocular inflammatory diseases, a highly diverse group of diseases of clinical and research importance, is presented in this chapter. Diagnoses that fall into this category are commonly grouped in terms of location of the primary site of inflammation in the eye: for example, the uvea or sclera. However, categorizing by location is heterogeneous in terms of causative etiology and immune pathophysiology. This chapter presents select common, clinically relevant, or scientifically interesting inflammatory diseases of the eye by anatomic location of inflammation and causative etiology where known. Some main topics of interest include the heterogeneous types of uveitis, neuro-ophthalmic immune-mediated diseases, paraneoplastic syndromes, ocular allergy, and issues in corneal transplant and transplant rejection. Absent from this chapter are topics such as orbital inflammatory diseases or immunologic factors in age-related macular degeneration (AMD). Several cutting-edge topics of active research interest, including the role of commensal microbiota in ocular immunology, cancer immunotherapy–associated ocular complications, and gene therapy, are briefly discussed.

THE OCULAR IMMUNE SYSTEM

The innate immune system of the eye includes the mechanical barriers of the eyelids, lashes, tears, and epithelium. The tear film has extensive antimicrobial properties, and contains secreted immunoglobulin A (IgA), lysozymes, lactoferrin, and peroxidases that protect the ocular surface in the absence of immune sensitization.[1] Additionally, dendritic cells (DCs) and resident macrophages in the conjunctiva and cornea produce antimicrobial substances, inflammatory cytokines, and chemokines, and act as antigen-presenting cells (APCs) of the eye. Mucosal-associated lymphoid tissue (MALT) associated with the conjunctiva and lacrimal glands are the primary locations of acquired sensitization. T cells outnumber B cells on the ocular surface, most of which are CD8 T cells in the conjunctiva, although immunoglobulin-producing B cells have been identified adjacent to lacrimal gland epithelia and contribute to the IgA component of tears.

Immune Privilege

The eye has a unique relationship with the immune system, and is one of few organs protected by immune privilege.[2] Immune privilege was originally defined as sites in the body where foreign tissue grafts can survive for extended periods of time without acute rejection, and is thought to be an evolutionary adaptation to protect indispensable tissues that have limited regeneration capacity.[2] The retina is of true central nervous system (CNS) origin and is protected by a tight blood-retina barrier, similar to that of the brain. Immune privilege of other ocular tissues is mediated by multiple anatomic, physiologic, and immunoregulatory mechanisms. Soluble immunosuppressive factors in the aqueous humor including transforming growth factor-β (TGF-β), neuropeptides such as α-melanocyte-stimulating hormone (α-MSH), vasoactive intestinal peptide (VIP), and others, and expression of immunomodulatory ligands including CD95L, TGF-β, and complement regulatory proteins contribute to ocular immune privilege.[3,4] Other unique mechanisms include low major histocompatibility complex (MHC) class I expression on the corneal endothelium, uvea, retinal pigment epithelium (RPE), and neural retina, peripheral tolerance of eye-derived antigens, and direct inhibition of T cells and preferential induction of T-regulatory (Treg) cells through both soluble and contact dependent mechanisms with ocular resident cells, including retinal glial Müller cells.[5]

The eye also employs mechanisms to actively suppress the systemic immune system response to an antigen introduced into the eye, known as anterior chamber–associated immune deviation (ACAID).[6] In animal models, antigens injected into the aqueous humor without any accompanying danger signal are taken up by APCs of the iris and ciliary body, which enter the venous circulation, bypassing the traditional lymphatic drainage. These APCs migrate through the bloodstream and can induce the formation of immunomodulatory suppressor cells in the thymus and spleen, including CD4 or CD8 Treg cells, marginal zone regulatory B cells, γδ Tregs, and invariant natural killer T cells (iNKT), which result in antigen-specific immune deviation.[6]

The eye uniquely has multiple complex layers of immune protection. However, one consequence of immune privilege is that the same mechanisms that protect the eye from immune activation may paradoxically leave the eye vulnerable to autoimmune diseases by preventing induction of peripheral tolerance.

Tolerance

Central tolerance occurs during T-cell development in the thymus where T cells are exposed to a wide cohort of self-antigens under control of the autoimmune regulator (AIRE) AIRE transcription factor.[7] T cells with high affinity for autoantigens undergo negative selection and apoptotic cell death. T cells with intermediate affinity for self-antigens may be converted to natural Treg cells to promote self-tolerance. Retinal antigens are expressed in the thymus under AIRE transcription factor control and there is detectable elimination of retina-specific T cells during negative selection in the thymus.[7] Thymic expression of ocular-specific genes is variable among individuals, and this variability may correlate with susceptibility to autoimmune uveitis and other ocular diseases. Additionally, the process of central tolerance alone is not sufficient to eliminate all autoreactive T cells.

Self-reactive T cells are further regulated in the process of peripheral tolerance, in which exposure to the cognate antigen in the absence of costimulatory signals, results in T-cell anergy or

conversion to induced Treg cells. Peripheral tolerance of ocular antigens is inefficient, as tissues of the eye are sequestered behind a blood ocular barrier with limited accessibility of unique ocular antigens. T cells that have escaped negative selection in the thymus persist in the periphery in a potentially autoreactive state due to inefficient peripheral tolerance mechanisms.[7]

INFLAMMATORY DISEASES OF THE EYE

The highly vascular nature of the uvea and prominent numbers of resident immune cells, including resident tissue macrophages, DCs, and mast cells, increases the propensity for immunologic disease when privilege is broken. As shown in experimental models of uveitis, privilege can be broken by even small numbers of activated effector T cells,[8] suggesting that the balance is very fine and can be easily tipped, leading to development of ocular inflammatory diseases.

Many systemic autoimmune diseases such as rheumatoid arthritis, sarcoidosis, inflammatory bowel disease, and systemic lupus erythematosus (SLE) develop ocular inflammatory manifestations. A wide variety of ocular involvement and clinical presentations may be seen, including uveitis, scleritis, conjunctivitis, or keratitis.[9] Ocular symptoms may also be the initial presentation of systemic autoimmune diseases, requiring physicians to have a high index of suspicion to make appropriate and timely diagnoses.

Inflammatory diseases can be loosely divided into primarily antibody-dependent or primarily cell-mediated etiologies, with most ocular diseases having contributions from both. Allergic conjunctivitis, atopic keratoconjunctivitis, vernal keratoconjunctivitis (VKC), and uveitis or scleritis associated with the joint diseases rheumatoid arthritis, juvenile idiopathic arthritis (JIA), and reactive arthritis, are believed to be predominantly driven by antibody-mediated mechanisms. Uveitis associated with sarcoidosis, Behçet disease, Vogt-Koyanagi-Harada (VKH) disease, and sympathetic ophthalmia are largely believed to be secondary to cell mediated immunity.[10] The balance between humoral and cellular immunity in these ocular inflammatory diseases frequently directs available treatment options.

Human leukocyte antigen (HLA) associations have been identified for a number of ocular inflammatory conditions. HLA-A29 carries a relative risk of 50, the highest HLA-related correlation, for developing birdshot chorioretinopathy, a form of uveitis.[11] Other associations include HLA-B27 with acute anterior uveitis and HLA-B51 with Behçet disease, among others.[12] It is commonly suggested that HLA expression in these diseases may confer a genetic susceptibility that requires additional insults, such as infection or environmental exposure, in order to result in clinical disease. HLA molecules may uniquely present disease-inducing peptides, facilitate molecular mimicry of endogenous peptides by infectious pathogens, or preferentially select disease-inducing subsets of T cells, among other proposed etiologic mechanisms (Table 74.1).[12]

UVEITIS AND INTRAOCULAR IMMUNE-MEDIATED INFLAMMATION

The uvea is a highly vascularized, middle layer of the eye composed of the iris, ciliary body, and choroid. Uveitis encompasses a spectrum of diseases resulting in inflammation of the uvea, including both infectious and immune-mediated etiologies. This chapter focuses on select non-infectious uveitis and their proposed immune mechanisms.

TABLE 74.1 Human Leukocyte Antigens Associated With Selected Ophthalmic Diseases

Disease	HLA association
Acute anterior uveitis	B27
Birdshot chorioretinopathy	A29
Behçet disease	B51
Sympathetic ophthalmia	A11, DRB1, DR4
Vogt-Koyanagi-Harada disease	DR4, DRB1
Giant cell arteritis	DR4, DRB1
Fuchs heterochromic iridocyclitis	DR53
Diabetic retinopathy	DQB1
Sjögren syndrome	DQB3, DBQ1
Mucus membrane pemphigoid	DR4, DQB1
Epidermal necrolysis	B58
Presumed ocular histoplasmosis syndrome	B7, DR2

HLA, Human leukocyte antigen.

TABLE 74.2 SUN Working Group Descriptors of Uveitis

Category	Descriptor	Comment
Onset	Sudden	
	Insidious	
Duration	Limited	≤3 months in duration
	Persistent	>3 months in duration
Course	Acute	Episode characterized by sudden onset and limited duration
	Recurrent	Repeated episodes separated by periods of inactivity without treatment ≥3 months in duration
	Chronic	Persistent uveitis with relapse in <3 months after discontinuous treatment

Jabs DA, Nussenblatt RB, Rosenbaum JT. Standardization of Uveitis Nomenclature Working G. Standardization of uveitis nomenclature for reporting clinical data. Results of the First International Workshop. *Am J Ophthalmol.* 2005;140(3):509–516.

Intraocular inflammation can be initiated by an antigen-specific or a nonspecific inflammatory response. Nonspecific ocular inflammatory responses can result from surgery, trauma, infections, and other causes of diffuse tissue damage. In contrast, antigen-specific immune responses are those directed against specific ocular antigens that are normally sequestered.[7] Frequently, development of an autoimmune response can be preceded by a viral or bacterial infection. An infectious agent may lead to initiation of an autoimmune response through molecular mimicry in which antigens on the pathogen cross-react with antigens in ocular tissue, or can induce an immune response by leading to breakdown of the blood-ocular barrier, ocular tissue destruction, and release of normally sequestered antigens.[13]

Uveitis is classified anatomically into anterior, intermediate, posterior, and panuveitis, and classified by etiology into infectious, non-infectious, and masquerade.[14] The duration and timing of inflammatory activity can be limited (3 months or less) or persistent, and disease can be acute, recurrent, or chronic (Table 74.2). Remission is defined as inactivity for 3 months or longer without treatment. Anterior chamber cell is a measure of the number of inflammatory cells in the aqueous and is a reliable indicator of inflammatory activity. A standardized grading scheme from 0 to 4+ for anterior chamber cell and flare has been defined to evaluate uveitis severity and monitor disease recovery (Tables 74.3 and 74.4). Aqueous flare is a measure of haziness of the aqueous fluid, reflecting increased protein due to breakdown of the blood-aqueous barrier. Large exudates

TABLE 74.3 SUN Working Group Grading Scheme for Anterior Chamber Cells

Grade	Cells in Field (1 × 1 mm slit beam)
0	<1
0.5+	1–5
1+	6–15
2+	16–25
3+	26–50
4+	>50

Jabs DA, Nussenblatt RB, Rosenbaum JT. Standardization of Uveitis Nomenclature Working G. Standardization of uveitis nomenclature for reporting clinical data. Results of the First International Workshop. *Am J Ophthalmol.* 2005;140(3):509–516.

TABLE 74.4 SUN Working Group Grading Scheme for Anterior Chamber Flare

Grade	Description
0	None
1+	Faint
2+	Moderate (iris and lens details clear)
3+	Marked (iris and lens details hazy)
4+	Intense (fibrin or plastic aqueous)

Jabs DA, Nussenblatt RB, Rosenbaum JT. Standardization of Uveitis Nomenclature Working G. Standardization of uveitis nomenclature for reporting clinical data. Results of the First International Workshop. *Am J Ophthalmol.* 2005;140(3):509–516.

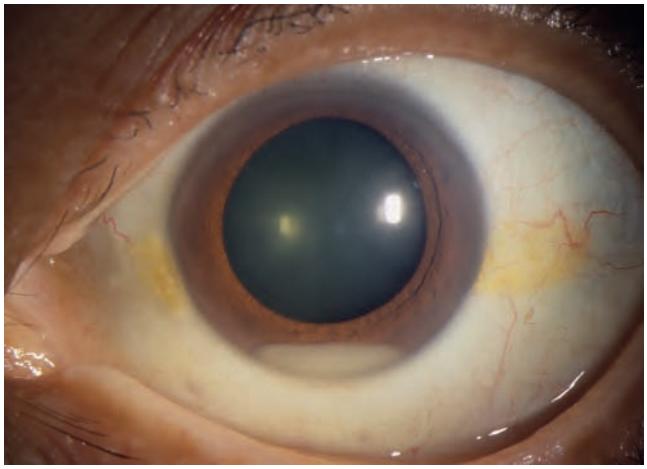


FIG 74.1 Hypopyon. A creamy white exudate rich in white blood cells produced as a result of intense intraocular inflammation. The exudate settles in the dependent aspect of the anterior chamber due to gravity.

of inflammatory white blood cells can result in layering of a purulent hypopyon in the anterior chamber and can also form deposits on the corneal endothelium known as keratic precipitates (Fig. 74.1). Other signs including iris nodules, fibrinous exudates, neovascularization, posterior synechiae, and reduced or elevated intraocular pressure, may also be observed (Figs. 74.2 and 74.3). Acute anterior uveitis is the most common presentation of uveitis. Idiopathic cases are common; however, a work-up to investigate systematic causes should be performed in cases of recurrent, severe, bilateral, or granulomatous uveitis.

Spondyloarthropathies

HLA-B27–associated seronegative spondyloarthropathies including ankylosing spondylitis, psoriatic arthritis, and reactive arthritis frequently have extra-articular manifestations. The most common ocular extra-articular finding is acute anterior uveitis, which is classically unilateral, but can recur in either eye at different times. Bilateral simultaneous involvement is rare. Up to 30% of patients with ankylosing spondylitis,[15] 20% of patients with reactive arthritis, and 10% of patients with psoriatic arthritis develop episodes of acute anterior uveitis, and less commonly may develop conjunctivitis, episcleritis, or scleritis.[16]

Common to all of these diseases is an association with HLA-B27 in rheumatoid-factor-negative joint disease. However, HLA-B27 is positive in approximately 5% of the general population, most of whom do not develop disease. The role of HLA-B27 in disease pathogenesis of the spondyloarthropathies and anterior uveitis is not fully understood. One proposed hypothesis is the arthrogenic/uveitogenic peptide hypothesis, in which HLA-B27 molecules present antigen that shares homology to those in the eye, joints, or other tissue.[17] Antigens from certain commensal gut bacteria, (*Chlamydia*) species, and numerous other organisms have been implicated, with some studies showing cross-reactivity with synovial fluid.[17] HLA-B27 heavy chains have a tendency to misfold, and an alternative hypothesis suggests that misfolding of HLA-B27

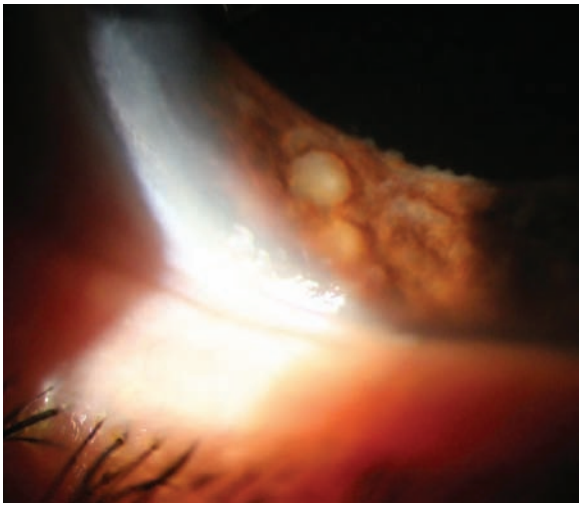


FIG 74.2 Iris Nodules. Bussaca nodules on the iris surface, which are characteristic of granulomatous uveitis. (Image courtesy of Dr. Edmund Tsui, Stein Eye Institute, David Geffen School of Medicine at UCLA.)

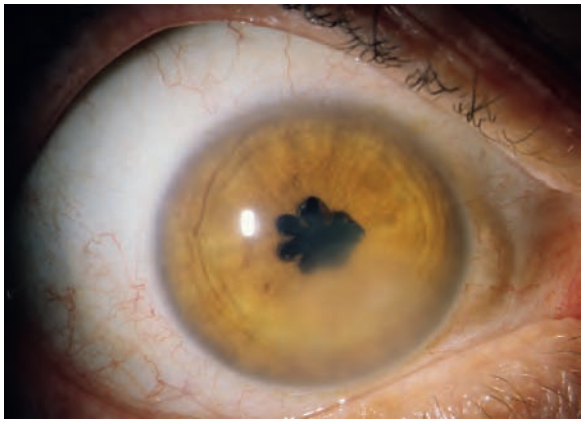


FIG 74.3 Posterior Synechiae. The irregular appearance of the pupil margin is a result of adhesions between the posterior iris and the anterior lens capsule, most commonly formed during states of inflammation. Opacification of the cornea in the inferonasal quadrant is also noted in this patient.

triggers the unfolded protein response, leading to production of proinflammatory cytokines that may play a role in mediating disease. Innate immune recognition of aberrantly folded HLA-B27 has also been proposed.

Juvenile Idiopathic Arthritis

JIA is the most common systemic cause of anterior, bilateral, non-granulomatous uveitis in children. The oligoarticular form is the most common and is also the most likely to develop uveitis. Antinuclear antibody (ANA) seropositivity and HLA-B27 positivity further increases risk of uveitis, which occurs in up to 30% of children with oligoarticular, ANA-positive JIA.[18] In most cases, arthritis is diagnosed before uveitis, and all children diagnosed with JIA should undergo routine eye screening, as the associated uveitis is typically asymptomatic and there may be no external signs of ocular disease in about 50% of affected patients. Patients rarely complain, even in acute exacerbations, making early and periodic screening critical, as undetected uveitis may lead to late complications including cataract, glaucoma, band keratopathy, and irreversible vision loss.

Pathogenesis of intraocular inflammation in JIA is unknown. Both B and T cells appear to be involved, with predominance of CD4 T cells and CD20 B cells identified in biopsy samples. The association with ANA suggests autoantibodies may be involved; however, ANAs have not been demonstrated to be pathogenic. There is a strong HLA association, with HLA-B27, HLA-DR5, and HLA-DRB1 shown to confer increased risk of uveitis, while HLA-CD1 and HLA-DQA appear protective.

Fuchs Heterochromic Iridocyclitis

Fuchs heterochromic iridocyclitis is a unilateral, chronic, low-intensity, non-granulomatous uveitis. Cellular activity in the anterior chamber is typically mild, and flare is faint. The affected eye is heterochromic due to iris stromal atrophy and, most commonly, the affected eye has a lighter iris color. Other ocular manifestations include cataracts in about 80% of patients, glaucoma in up to 59%, and, less commonly, iris nodules and iris neovascularization.

Evidence suggests that the rubella virus may be causative in this disease. Local intraocular synthesis of rubella antibodies in the aqueous humor has been identified in eyes affected by Fuchs heterochromic iridocyclitis, but not in other causes of anterior uveitis.[19,20] Rubella RNA has also been isolated from the intraocular fluid in some cases, and persistent intraocular rubella virus is thought to result in chronic low-grade inflammation in the affected eye. The incidence of Fuchs heterochromic iridocyclitis has declined significantly in patients born in the era of routine measles-mumps-rubella (MMR) vaccination, although incidence may be higher in countries with lower vaccination rates. Additional infectious associations include ocular toxoplasmosis, herpes simplex virus (HSV), and cytomegalovirus (CMV) infections.

Tubulointerstitial Nephritis and Uveitis

Tubulointerstitial nephritis and uveitis (TINU) is a rare immune-mediated disorder characterized by acute tubulointerstitial nephritis and a bilateral, non-granulomatous anterior uveitis. The disease presents most commonly in adolescent girls, with a median age of 15 years, but occurs in males with an age range from 9 to 74. Renal disease typically precedes uveitis, but uveitis can occur prior to renal disease in about 20% of cases; the severity of the renal and ocular disease appears independent.

Disease pathogenesis may involve both dysregulated cell-mediated immunity with loss of T-cell tolerance, as well as humoral immunity with autoreactive antibodies directed against renal and ocular antigens.[21] TINU patients were found to have elevated titers of anti-monomeric C-reactive protein (anti-mCRP) antibodies, targeting an acute phase reactant that deposits in the kidney and is believed to be predictive of developing subsequent uveitis.[21] One leading hypothesis is that an inciting event in the kidney stimulates an HLA class II response that targets an antigen common to the eye and kidney. This antigen could be a yet unknown native protein found in the uvea and renal interstitium, or it could directly be the mCRP acute phase reactant that deposits in both organs.

Sarcoidosis

Sarcoidosis is a chronic systemic disease characterized by non-caseating granulomatous inflammation that can affect any organ system. The disease has a high racial predilection, affecting Blacks 10 times more frequently than Caucasians. Ocular involvement is common in sarcoidosis, affecting 30% to 60% of patients, and can be the presenting symptom/sign of the disease.[22] Any ocular structure can be involved by sarcoidosis, including acute granulomatous anterior uveitis, intermediate, posterior, or panuveitis, optic nerve, or orbit. Sarcoid-associated uveitis is typically characterized by bilateral large granulomatous mutton-fat keratic precipitates, iris nodules, trabecular meshwork nodules, vitreous opacities, chorioretinal peripheral lesions, and retinal phlebitis (Fig. 74.4).

The etiology of sarcoidosis is unknown. Phagocytosis and presentation of an unidentified antigen by macrophages or DCs is thought to initiate formation of sarcoid granulomas through an exaggerated cell-mediated immune response. No single etiologic agent has been clearly implicated; however, mycobacteria, propionibacteria, vimentin, serum amyloid A, environmental exposures, and other agents have all been investigated as having potential roles in the disease pathophysiology.

Behçet Disease

Behçet disease is an autoimmune vasculitis that typically presents with recurrent aphthous oral ulcers, genital ulcers, and uveitis. The disease is more common in patients from eastern Asia, the Middle East, and the Mediterranean, and is strongly associated with HLA-B51. Ocular inflammation occurs in approximately 70% of patients and can be the presenting symptom. A bilateral, acute panuveitis with retinal vasculitis is typically observed, and usually presents with a relapsing and remitting course. The retinal vasculitis can present as a mix of arteritis and phlebitis.

Vasculitis in Behçet disease may cause severe and permanent visual disability and may involve both arterial and venous structures of any size (Fig. 74.5). The etiology is unknown, but it has been proposed to be caused by environmental or infectious triggers that induce inflammatory episodes in genetically susceptible individuals. HLA-B51 is the strongest susceptibility factor; however, other associations have been identified including Interleukin (IL)-23 and IL-10 receptor variants and endoplasmic reticulum aminopeptidase 1 (ERAP1), an enzyme involved in processing of peptides to be loaded onto the antigen-binding groove of MHC molecules.[23] One hypothesis is that ERAP1 polymorphisms may affect the specificity of peptides loaded onto HLA-B51, resulting in activation of a cytotoxic T-cell response or, alternatively, an antigen-independent innate inflammatory response caused by endoplasmic reticulum (ER) stress

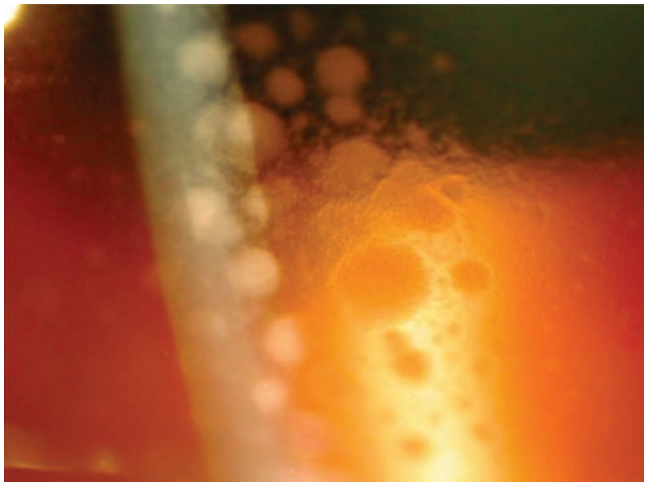


FIG 74.4 Granulomatous Keratic Precipitates. Large, greasy-appearing, yellowish-white mutton-fat keratic precipitates, which are manifestations of leukocyte aggregates adherent to the corneal endothelium. These deposits are characteristic of granulomatous uveitis. (Reproduced with permission from Keenan JD, Tessler HH, Goldstein DA. Granulomatous inflammation in juvenile idiopathic arthritis–associated uveitis. *Journal of AAPOS*. 2008(12):540–550.)

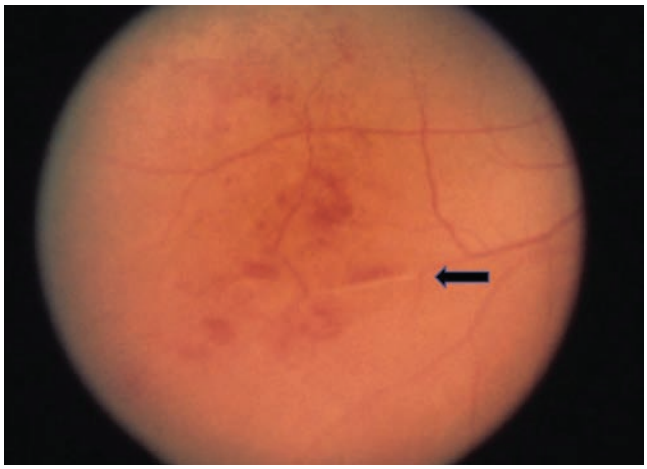


FIG 74.5 Fundus photograph showing retinal vasculitis in a patient with behçet disease. *Arrow* pointing to area of vascular sheathing and occlusion. Vascular occlusion and scattered hemorrhages are prominent.

from inappropriate matches between HLA-B51 and the available peptides.[23] Patients with Behçet disease also have exaggerated innate inflammatory responses, and other hypotheses propose that defects in sensing or processing of pathogen-associated or danger-associated molecular patterns leads to the strong nonspecific inflammatory response characteristic of Behçet disease.

Birdshot Chorioretinopathy

Birdshot chorioretinopathy is a chronic bilateral posterior uveitis predominantly seen in middle-aged Caucasian females. Classically, birdshot chorioretinopathy presents as a white, painless eye with minimal anterior segment inflammation, retinal vascular leakage, and cystoid macular edema in late stages (Fig. 74.6). Over 96% of patients are HLA-A29 positive.[24] An association with ERAP2 has also been identified, suggesting selective antigen processing and presentation on HLA-A29 may create a pathogenic immunogenic signal in this primarily T-cell-mediated disease. The disease-causing antigen is unknown; however, numerous viral antigens have been proposed with potential cross-reactivity against ocular antigens, including the retinal specific S-antigen, intraretinal-binding protein (IRBP), and a number of melanocyte-derived peptides. Increased disease risk is further conferred by specific combinations of killer-cell immunoglobulin-like receptor (KIR) genes, while other KIR genotypes are protective, emphasizing the multifactorial and polygenic nature of this disease.[25]

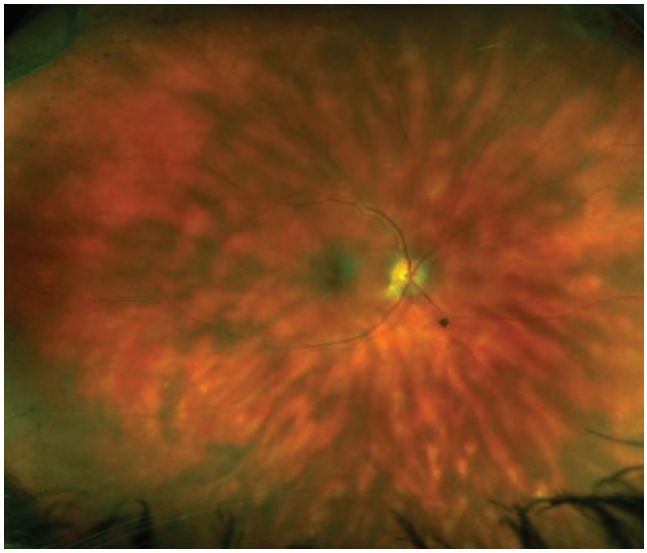


FIG 74.6 Fundus photograph showing multiple small, cream-colored fundus lesions scattered around the optic disc and radiating to the equator in a "shotgun" pattern, typical of birdshot chorioretinopathy. (Image courtesy of Dr. Edmund Tsui, Stein Eye Institute, David Geffen School of Medicine at UCLA.)

Vogt-Koyanagi-Harada

VKH is an immune, T-cell-mediated disease characterized by bilateral panuveitis in association with neurologic, auditory, and skin manifestations such as poliosis and vitiligo. It may be associated with an autoimmune response against melanin antigens. The associated uveitis may be granulomatous acutely with a thickened choroid, optic disc edema and hyperemia, and serous retinal detachments. Neurologic signs can include headache, cranial nerve abnormalities, cerebrospinal fluid (CSF) pleocytosis, and neck stiffness, and auditory manifestations can include tinnitus, hearing loss, and vertigo.[26,27] Patients are at high risk for complications including cataract, glaucoma, subretinal fibrosis, and choroidal neovascularization.

The incidence of VKH varies worldwide, with the disease common in Asia, the Middle East, and South America.[26] In contrast, the disease is rare in European populations.[26] HLA-DR4, HLA-DR1, HLA-DRB1, and HLA-DRw53 have all been associated with a significantly increased risk for VKH. KIR combinations on natural killer cells and cytotoxic T cells have also been implicated in conferring disease susceptibility.[28] It is hypothesized that certain combinations of HLAs and KIRs preferentially present immunogenic antigen epitopes from melanocyte-associated antigens; however, the exact antigens involved are unknown.[29] Environmental factors such as preceding viral infection, and molecular mimicry between

viral antigens and melanocyte proteins, have been hypothesized to play an inciting role.[27] VKH uveitis has also been reported to be associated with cutaneous pigmented malignant melanoma, with VKH thought to be a consequence of autoimmune reaction against melanoma in patients with good cancer immunosurveillance.[30]

Sympathetic Ophthalmia

Sympathetic ophthalmia is a rare type of bilateral granulomatous uveitis that occurs following penetrating eye trauma or eye surgery in one eye, leading to disease in the fellow eye. The pathophysiology is thought to involve a T-cell-mediated response against ocular self-antigens that are unmasked following trauma that abrogates immune privilege and tolerance mechanisms.[31] Multiple antigens, including rod photoreceptor outer segments, and RPE and choroidal melanocyte antigens, have been suggested to be causative and specific HLA associations have been made; however, the exact mechanisms remain elusive (Fig. 74.7).

The time between the inciting trauma or surgery and sympathetic ophthalmia development is typically rapid, with 80% of patients developing symptoms within 3 months, and 90% within 1 year.[31] Clinical presentation can be varied, with the exciting eye typically having photophobia, decreased vision, and mutton-fat keratic precipitates. Inflammation in the fellow eye may develop insidiously or more rapidly, and classically presents with photophobia, mild pain, epiphora, paresis of accommodation, cells and flare in the anterior chamber, mutton-fat keratic precipitates, and posterior synechiae.[32] Intraocular pressure may either be high from inflammatory cell blockage of trabecular meshwork flow or low due to ciliary body shutdown. Clinical course may have intermittent episodes of acute inflammation with highly variable outcomes. The disease is characterized by a diffuse, granulomatous, non-necrotizing inflammatory response of the entire uveal tract, with cellular infiltrates dominated by T cells. Sympathetic ophthalmia can also present with alopecia, vitiligo, poliosis, and deafness, overlapping significantly with VKH-like symptoms.

Prevention of sympathetic ophthalmia may be possible with enucleation of an injured eye within 2 weeks of the initial insult, especially in cases with little or no prognosis for recovery of useful vision. There is controversy over the benefit of enucleation once disease has begun in the sympathizing eye.

THERAPEUTIC PRINCIPLES

Current Treatment Options for Uveitis

Corticosteroids
- Topical steroids: prednisolone, dexamethasone, difluprednate, loteprednol etabonate, betamethasone, prednisolone, fluorometholone
- Long-term depot steroids: subconjunctival or sub-Tenons steroid deposit
- Intravitreal steroids: intravitreal triamcinolone acetonide, slow-release intravitreal steroid implants
- Oral: prednisone

Steroid-Sparing Agents
- Antimetabolites: azathioprine, methotrexate, mycophenolate mofetil
- T-cell inhibitors: cyclosporine, tacrolimus
- Alkylating agents: cyclophosphamide, chlorambucil
- Biologic therapy: adalimumab (FDA approved for non-infectious intermediate, posterior, and panuveitis)

Ancillary Therapy
- Cycloplegic agents: cyclopentolate, homatropine, atropine
- NSAIDs: naproxen, tolmetin

NSAIDs, Non-steroidal anti-inflammatory drugs.

Uveitis Treatment

Damage from uveitis occurs secondary to inflammatory cell infiltrates and release of inflammatory mediators that facilitate destruction of ocular tissues. The main approach to treating non-infectious ocular inflammation has been to inhibit the activity of the immune response. Topical corticosteroids are common first-choice options for anterior uveitis. Gradual tapering of the corticosteroid is initiated once the inflammation is controlled. The treatment goal is near-complete suppression of inflammation with no anterior chamber cell or flare. Common complications of topical steroids include elevation of intraocular pressure and cataract formation. Ancillary therapy including cycloplegic agents may also be used to prevent formation of posterior synechiae, break formed synechiae, or relieve ciliary muscle spasm.

For more severe cases, or cases where steroid tapering is unsuccessful, long-term depot steroid preparations can be administered via subconjunctival or sub-Tenon routes. Intravitreal triamcinolone acetonide or slow-release intravitreal steroid implants may also be considered in refractory cases. Systemic oral steroids can be used in treatment of refractory disease; however, they come with potential adverse systemic effects. Antimetabolites including azathioprine, methotrexate, and mycophenolate mofetil, T-cell inhibitors such as cyclosporine and tacrolimus, and alkylating agents such as cyclophosphamide and chlorambucil may be useful as steroid-sparing agents for patients with severe or corticosteroid-refractory disease. Newer biologic therapy has proven to be successful in cases where more conventional immunosuppressive therapy has failed, and may be particularly useful in patients with concurrent systemic symptoms. The tumor necrosis factor (TNF) inhibitor adalimumab is the first US Food and Drug Administration (FDA)-approved biologic for the treatment of non-infectious uveitis. Other agents which target inflammatory pathways, including IL-1 and IL-6 inhibitors, may have future roles in the treatment of refractory uveitis.[33]

NEURO-OPHTHALMIC IMMUNE-MEDIATED DISEASES

The eye is an extension of the CNS, and ocular symptoms can frequently manifest concurrently with or as the presenting symptom of a systemic CNS disease. Immune-mediated diseases also frequently target the optic nerve, or neuromuscular junctions (NMJs) of extraocular muscles and can have varying levels of systemic involvement. Topics of neuro-ophthalmic interest that will be presented here include optic neuritis, including the AQP4- and MOG-associated forms, giant cell arteritis (GCA), the Miller Fisher variant of Guillain-Barré syndrome (GBS), myasthenia gravis, and Lambert-Eaton myasthenic syndrome (LEMS).

Optic Neuritis

Optic neuritis, or inflammation of the optic nerve, can arise from demyelinating, autoimmune, and infectious etiologies. It classically presents with acute, unilateral, painful vision loss, with retroorbital pain typically worsening with eye movement. Patients may have visual field loss, color vision deficits, and an afferent pupillary defect. Optic neuritis is a common clinical manifestation of the demyelinating disease multiple sclerosis (MS; Chapter 66). Optic neuritis affects approximately 70% of patients with MS and

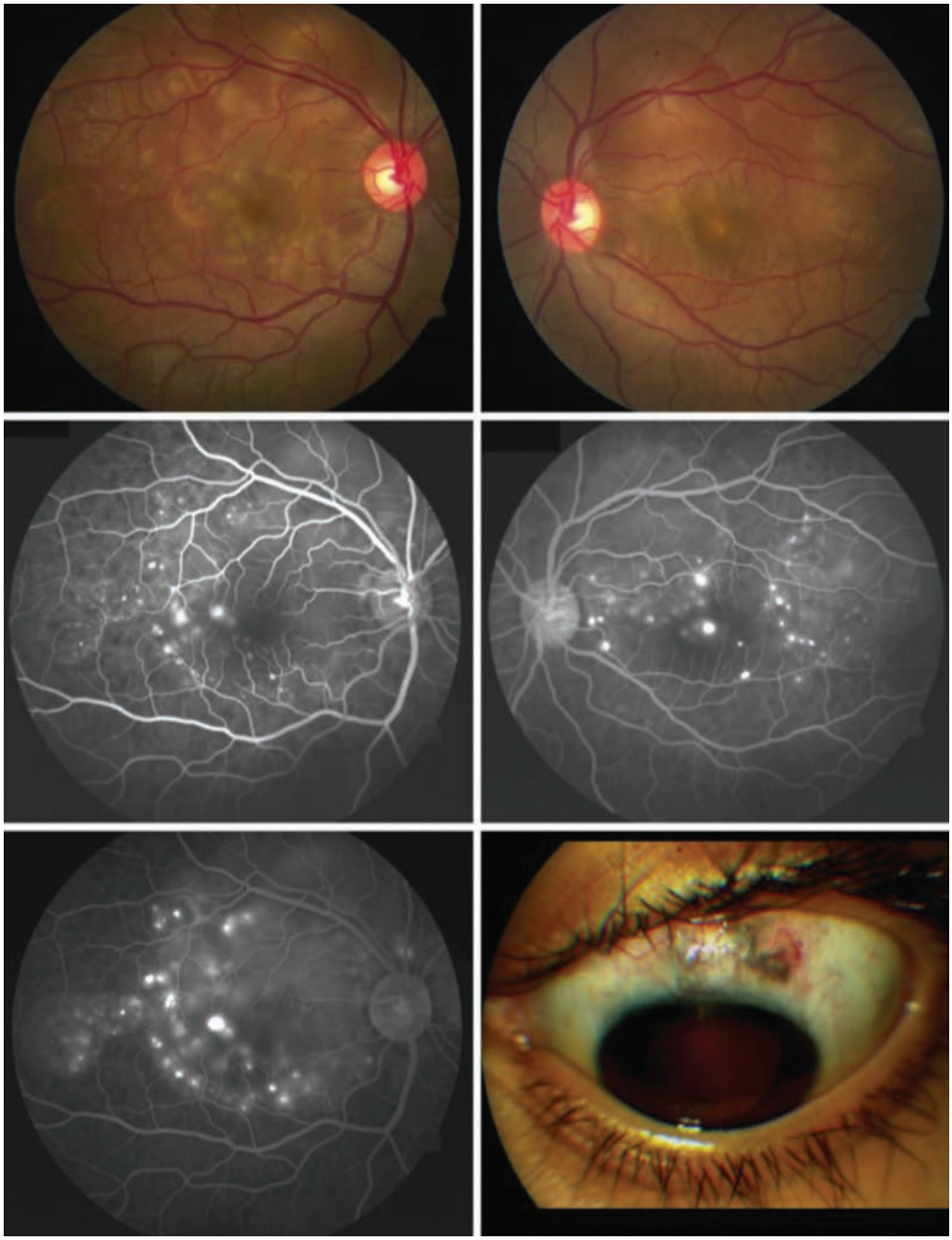


FIG 74.7 Fundus photographs and fluorescein angiography of a patient with sympathetic ophthalmia following a penetrating left eye globe injury. Foci of exudative retinal detachment are present in the posterior poles of both eyes. Fluorescein angiography demonstrating multiple areas of pin-point leakage at the level of the retinal pigment epithelium. (Reproduced with permission from Vasconcelos-Santos DV, Rao, NR. *Sympathetic Ophthalmia*. Ryan's Retina. Chapter 77 Fig. 74.5:14961–504.)

is frequently the presenting symptom of the systemic disease.[34] Magnetic resonance imaging (MRI) is required to evaluate the extent of optic nerve as well as brain involvement. The presence of a single, typical demyelinating lesion on MRI can stratify the risk of developing systemic MS in cases where optic neuritis is the initial presentation. Optic neuritis associated with MS typically results in good recovery of visual acuity, however, there is often permanent damage that is visualized using optical coherence tomography (OCT) to evaluate the retinal nerve fiber layer, as well as the thickness of the retinal ganglion cell and inner plexiform layers of the macular retina.

Two unique forms of optic neuritis have recently been identified that are defined by the presence of specific antibodies: namely, anti-aquaporin 4 (AQP4)-associated and anti-myelin oligodendrocyte glycoprotein (MOG)-associated disease. These forms present with more severe vision loss, are more commonly bilateral, and recur more frequently than typical MS-associated demyelinating optic neuritis.[35]

Neuromyelitis optica spectrum disorder (NMOSD) is an inflammatory CNS condition characterized by a variety of "core" syndromes, one of which is optic neuritis. When optic neuritis is associated with other "core" syndromes that are disseminated

in space and are associated with certain findings on MRI or the presence of antibodies against aquaporin-4 (AQP4), an abundant water channel in the CNS, then the findings meet diagnostic criteria for NMOSD.[36] AQP4 antibodies are shown to be pathologic in animal models, mediating complement activation and cytotoxic damage.[35] The AQP4 water channel is predominantly expressed on astrocytic foot processes, and NMOSD is characterized primarily by astrocytic damage, with oligodendrocyte loss and demyelination as a secondary feature.

Optic neuritis associated with antibodies against MOG is bilateral in about 40% of affected individuals. Edema of the optic disc and recurrences are both common. Optic neuritis may be isolated or may be associated with transverse myelitis or encephalitis. In contrast to AQP4-associated disease, this disease is characterized histologically by oligodendrocyte loss and demyelination in the absence of astrocytopathy.

Management of acute attacks includes high-dose intravenous corticosteroid and may require plasma exchange or intravenous immunoglobulin (IVIG). Management for relapsing disease may require long-term immunosuppressant agents.[37] Patients with NMOSD may benefit from early plasma exchange. Additionally, the complement C5 neutralizing antibody eculizumab has recently been approved by the FDA for the treatment of NMOSD. Non-FDA-approved therapies include B-cell depletion or immunosuppression.

CLINICAL PEARLS

Optic Neuritis

Multiple sclerosis–associated optic neuritis:
- Unilateral, (bilateral presentation is rare)
- Young adults, female predominance
- Acute demyelination
- Good recovery of visual acuity

Neuromyelitis optica spectrum disorder (NMOSD):
- Associated with anti-aquaporin-4 (AQP4) antibodies
- Predominantly astrocyte damage
- Older adults, female predominance
- May have severe visual acuity loss
- 20% bilateral
- Recurrences common
- Poor visual outcome

Myelin oligodendrocyte glycoprotein (MOG)-associated optic neuritis:
- Associated with anti-myelin oligodendrocyte glycoprotein (MOG) antibodies
- Predominantly oligodendrocyte loss and demyelination
- Pediatric and adult, no sex predilection
- 40% bilateral
- May have severe visual acuity loss
- Recurrences common
- Most with optic disc edema
- Good visual outcome

Giant Cell Arteritis

GCA is an autoinflammatory granulomatous disease primarily targeting medium-and large-sized arteries, with DCs, T cells, and macrophages identified as the main drivers of the inflammatory response. Abnormal activation of vascular DCs recruits other immune cells to the vessel wall through the vasa vasorum, a normally immune-protected tissue microenvironment.[38] T cells and large multinucleated giant cells derived from activated macrophages are found in granulomatous infiltrates in affected arterial walls. A reactive vascular injury response can cause endothelial cell damage, intimal hyperplasia, vessel lumen occlusion, and subsequent tissue ischemia that leads to the major clinical complications of GCA. Age is the greatest risk factor for GCA, and likely contributes to the immune and vascular system dysfunction seen in this disease. Most patients with GCA appear to have systemic inflammation, with elevated acute phase markers including erythrocyte sedimentation rate and C-reactive protein.[38] Evidence suggests that circulating blood monocytes and macrophages in GCA patients may be in an abnormally activated state, producing proinflammatory cytokines, including IL-1β and IL-6.[39] The disease also has an association with HLA-DR4, which is commonly associated with other autoimmune diseases such as rheumatoid arthritis.

Clinical presentation of the disease includes headache, jaw or tongue claudication, new neck pain, scalp pain, fatigue, and weight loss. Involvement of the aorta can lead to aneurysm, aortic dissection and rupture. Vasculitis and occlusion of the vessels supplying the eye and optic nerve can lead to arteritic anterior ischemic optic neuropathy (AAION) due to ischemic damage to the optic nerve. AAION classically presents with acute loss of vision in one eye with severe optic disc edema and a relative afferent pupillary defect. Permanent occlusion may be preceded by episodes of transient vision loss or double vision, and if untreated, the risk to the fellow eye is high. GCA is an ophthalmologic emergency and can progress quickly to permanent vision loss.

Diagnosis of GCA is histologic, with temporal artery biopsy showing intimal thickening, internal elastic lamina fragmentation, and chronic inflammatory infiltrate with giant cells. Due to the high risk of vision loss in the fellow eye, early aggressive treatment is recommended prior to biopsy. The gold standard of therapy is high-dose glucocorticoids, and the IL-6 inhibitor tocilizumab is used in select patients as a steroid-sparing agent.

Miller Fisher Variant of Guillain-Barré Syndrome

Miller Fisher syndrome (MFS) is a rare, acute demyelinating peripheral polyneuropathy variant of GBS, occurring in about 5% of cases. Clinically the presenting symptoms include an infectious prodrome followed by diplopia, ptosis, ophthalmoplegia, supranuclear palsies and facial palsies, ataxia, and areflexia. Presence of anti-GQ1b antibodies with a CSF albuminocytologic dissociation is suggestive of MFS. Electromyography (EMG) and nerve conduction studies can also be supportive. The pathologic antibody is targeted against sialylated glycosphingolipids located in neuronal and glial plasma membranes and neuromuscular junctions (NMJs).[40] High concentrations of the GQ1b ganglioside are found in the oculomotor nerve, trochlear nerve, and abducens nerve, explaining the relationship with ophthalmoplegia in this subtype of the disease (Fig. 74.8). Viral or bacterial organisms, including *Campylobacter jejuni*, have been found to contain lipopolysaccharides that are antigenically identical to those on human gangliosides, supporting molecular mimicry as the method of immune sensitization and disease initiation. Treatment includes IVIG, plasmapheresis, and supportive care, with a good overall prognosis.

Myasthenia Gravis

Myasthenia gravis is characterized by autoantibodies against postsynaptic acetylcholine receptors (AChRs) on skeletal muscle leading to dysfunction at the NMJ. Functionally, this leads to impaired signal transduction at the NMJ, resulting in muscle weakness and

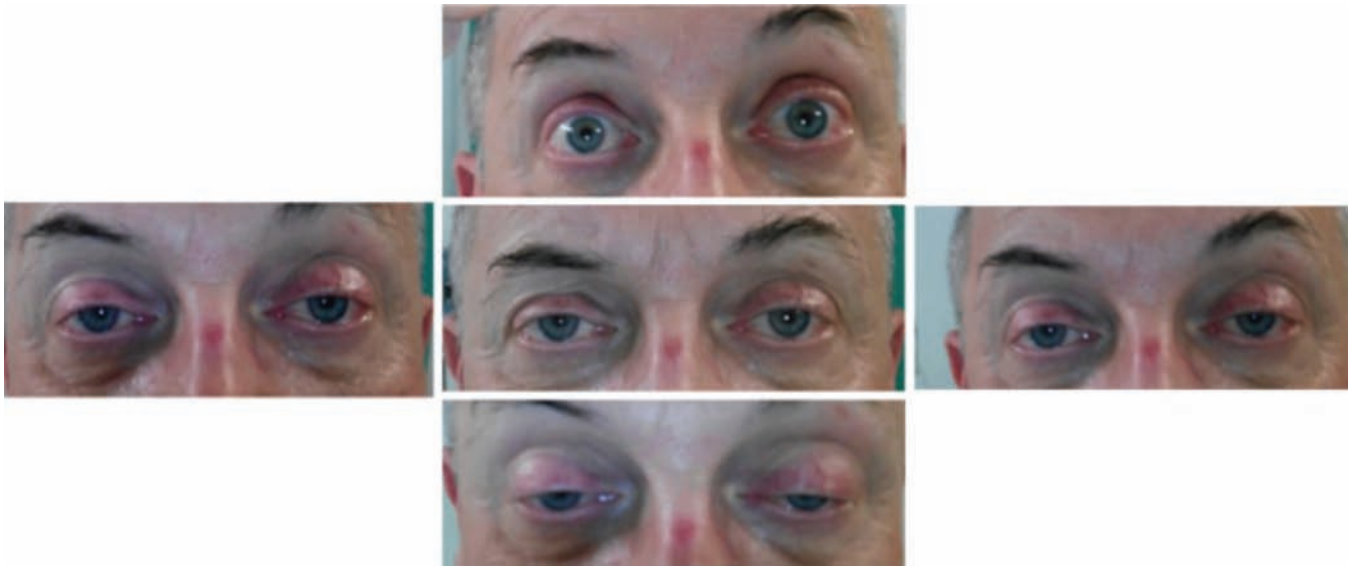


FIG 74.8 Extraocular Motility in a Patient with the Miller Fisher Variant of Guillain-Barré syndrome. There is mild right ptosis and ophthalmoplegia, with decreased movement in all fields of gaze. (Reproduced with permission from Pellegrini F, Prosdocimo G, Barton JJS. *Survey of Ophthalmology*. 2016;61:248–254. See Fig. 74.2.)

fatigability.[41] Typical ocular symptoms include ptosis and diplopia with fatigability of extraocular muscles that worsens with sustained muscle use and improves with cold, rest, and sleep. It can present localized to the palpebral and extraocular muscles, or occur in conjunction with systemic symptoms including difficulty swallowing, change in voice, shortness of breath, and weakness of the extremities. Ocular manifestations are commonly the initial presentation of generalized myasthenia gravis, with most patients developing systemic symptoms within 2 years.[42]

Serologic tests for disease-causing circulating antibodies against nicotinic AChRs are diagnostic; however, many patients are seronegative with no detectable antibodies.[43] Antibodies against other NMJ proteins including anti-muscle-specific tyrosine kinase (MuSK) and low-density lipoprotein (LDL)-related receptor-related protein 4 (LRP4) have also been reported and may be causative in some AChR seronegative cases. Electromyography is diagnostic and may be important in establishing the diagnosis in seronegative patients. Single-fiber EMG, the diagnostic "gold standard," shows characteristic muscle fiber jitter. Repetitive nerve stimulation shows a decrement in the muscle action potential, in contrast to the paraneoplastic mimicker Lambert-Eaton syndrome, which is characterized by increase in amplitude with repetitive nerve stimulation.

Treatment includes acetylcholinesterase inhibitors, most commonly pyridostigmine, for symptomatic management of generalized disease, but may be ineffective for ocular manifestations. Immunosuppressive drugs including glucocorticoids and steroid-sparing agents such as azathioprine and mycophenolate mofetil may be required for treatment of generalized disease. Chest computed tomography (CT) imaging should be obtained in every patient to assess for presence of a thymoma, which, if present, would require thymectomy.[41]

PARANEOPLASTIC SYNDROMES

Paraneoplastic syndromes are those that arise secondary to immune responses against a remote tumor, resulting in autoimmune responses against normal tissue. In some diseases, specific causative antibodies or target antigens have been identified. Seropositivity and a history of known malignancy may suggest the presence of a paraneoplastic syndrome; however, the ophthalmic consequences of these paraneoplastic syndromes are varied in presentation and may precede a cancer diagnosis by months to years, adding to the diagnostic difficulty of this group of diseases.[44] This section describes cancer-associated retinopathy (CAR) and melanoma-associated retinopathy (MAR), the most common paraneoplastic retinopathies, as well as LEMS, a common mimicker of myasthenia gravis. However, other paraneoplastic syndromes have been identified, including bilateral diffuse uveal melanocytic proliferation (BDUMP), polyneuropathy, organomegaly, endocrinopathy, monoclonal plasma cell disorder and skin changes (POEMS), paraneoplastic optic neuropathy, and opsoclonus-myoclonus, but are not be discussed here. Myasthenia gravis, discussed in the prior section, can result from a primary autoimmune condition, or rarely a paraneoplastic syndrome associated with thymoma or thymic hyperplasia.

Lambert-Eaton Myasthenic Syndrome

LEMS is mediated by antibodies targeting presynaptic P-Q voltage-gated calcium channels at the NMJ, causing a decrease in voltage-gated calcium expression at the synapse, subsequent decrease in calcium internalization, and ultimately decreased release of acetylcholine at the NMJ. Clinical symptoms can be similar to those seen in myasthenia gravis and include proximal muscle weakness, autonomic dysfunction, and hyporeflexia or areflexia. Ocular symptoms include diplopia and a milder ptosis that may, in contrast to myasthenia gravis, decrease with prolonged upgaze. Other features that distinguish LEMS from myasthenia gravis include an improvement in strength or tendon reflexes after muscle use or exercise, and EMG that shows a low-amplitude compound muscle action potential that increases in amplitude with high-frequency repetitive nerve simulation or after maximal muscle contraction.[44]

LEMS is strongly associated with small cell lung cancer, and patients with suspected LEMS are recommended to undergo chest CT imaging first to identify a causative tumor. Treatment of the underlying malignancy can often greatly improve neurologic

symptoms. Other treatment options include amifampridine, a voltage-dependent potassium channel blocker, pyridostigmine, corticosteroids, IVIG, or plasma exchange.

Autoimmune Retinopathies

Autoimmune retinopathies are rare ophthalmic disorders characterized by autoantibodies against retinal proteins leading to progressive visual loss. Anti-retinal antibodies can be paraneoplastic and occur in the setting of an underlying malignancy, with CAR and melanoma-associated retinopathy (MAR) being the most common classic retinal-based paraneoplastic syndromes.

CAR is the most prevalent paraneoplastic retinopathy and is most often associated with underlying small cell lung carcinoma, followed by gynecologic and breast malignancies. Vision loss is typically painless, progressive, and bilateral, and can occur before a systemic malignancy is diagnosed in up to 50% of cases. Molecular mimicry is proposed as the causative mechanism, with tumor cells expressing antigenic epitopes that cross-react with retinal photoreceptors or other retinal antigens. Numerous retinal antigens have been identified, including recoverin, α-enolase, transducin, arrestin, photoreceptor cell-specific nuclear receptor, and multiple others.[45] Recoverin is expressed and functional in many tumors, and anti-recoverin antibodies may have tumor-suppressing effects,[46] suggesting anti-tumor immunosurveillance may be at the cost of retinal autoimmunity in these patients.

MAR typically presents following diagnosis of the primary cutaneous melanoma with a latency of 2 to 19 years and can frequently be a harbinger of non-ocular metastasis.[45] MAR is hypothesized to result from autoantibodies against postsynaptic receptors of the depolarizing ON bipolar cells, disrupting rod photoreceptor transmission, and mediating rod-predominant dysfunction. A wide variety of retinal antigen targets have been implicated, emphasizing the immunologic heterogeneity associated with these paraneoplastic disorders. Rare cases of melanoma patients developing an unusual paraneoplastic vitelliform maculopathy have also been reported, including one patient with circulating autoantibodies against bestrophin,[47] and another with antibodies against an RPE peroxidase.[48]

Treatment for CAR and MAR involves long-term immunosuppression with corticosteroids, with possible benefit from plasmapheresis and IVIG to decrease circulating anti-retinal antibodies. Overall prognosis is poor, with rapid vision loss, and treatment of the underlying malignancy does not appear to improve vision.

EXTRAOCULAR IMMUNE-MEDIATED INFLAMMATION

Ocular Allergy

Allergic conjunctivitis is extremely common and often underdiagnosed and undertreated. It is a type I hypersensitivity, IgE-mediated reaction to a specific allergen, and can be seasonal due to airborne pollens or perennial due to year-round allergens. Activation of mast cells increases histamine, tryptase, prostaglandins, and leukotrienes in tears, as well as activation of vascular endothelial cells to secrete chemokines and adhesion molecules for inflammatory cell recruitment, leading to inflammation in the conjunctival mucosa.[49] Ocular symptoms include conjunctival hyperemia, acute chemosis, and itching. Management includes avoidance of the allergen when possible, topical antihistamines or mast cell stabilizers, topical non-steroidal anti-inflammatory drugs (NSAIDs), or mild topical steroids.

VKC is a nonspecific hyperreactivity reaction mediated by Th2 lymphocytes. It is more common during the spring and summer months in warm climates and is thought to be due to nonspecific stimuli such as wind, dust, and sunlight, rather than airborne allergens. Skin tests and serum IgE tests against common seasonal allergens are often nonreactive. VKC often involves a chronic ocular surface inflammation with increased eosinophils, neutrophils, mononuclear cells, mast cells, and lymphocytes. The formation of giant papillae filled with inflammatory cells and edema on the upper tarsal conjunctiva is a characteristic finding. The cornea can become involved with development of a punctate keratitis that may lead to central scarring and declining vision (Fig. 74.9). Horner-Trantas dots, consisting of clumps of necrotic eosinophils, are a characteristic finding in the active phase of the disease.[49] VKC is more common in children and typically resolves by age 20. Management includes topical antihistamines or mast cell stabilizers, topical steroids, supratarsal injection of steroids, or topical cyclosporine.

Atopic keratoconjunctivitis is a chronic ocular inflammatory disease, considered to be the ocular counterpart to atopic dermatitis. It involves both type I and type IV hypersensitivity with chronic IgE-mediated mast cell degranulation as well as altered T-cell-mediated immunity. Over 30% of patients have a history of atopic dermatitis or other systemic atopy. Patients typically present with chronic papillary conjunctivitis without seasonal variation and eczematous lesions on the eyelids (Fig. 74.10). The presence of giant papillae or Horner-Trantas dots may be variable. Anterior shield-like atopic cataracts may also develop. More severe manifestations can include neovascularization, limbal stem cell deficiency, and corneal opacification. Treatment includes the topical options for VKC but may also require adding systemic therapy for chronic management.

KEY CONCEPTS

Ocular Allergy

Allergic Conjunctivitis
- Type I hypersensitivity—IgE-mediated activation of mast cells
- Reaction to specific seasonal or perennial allergen

Vernal Keratoconjunctivitis
- Nonspecific hyperreactivity with chronic ocular surface inflammation
- Reaction against nonspecific stimuli: for example, wind, dust, sunlight
- Giant papillae on upper tarsal conjunctiva
- Horner-Trantas dots of necrotic eosinophils

Atopic Keratoconjunctivitis
- Type I and type IV hypersensitivity—IgE-mediated activation of mast cells with altered T-cell immunity
- Associated with history of atopic dermatitis and systemic atopy
- Chronic papillary conjunctivitis

Sjögren Syndrome

Sjögren syndrome is a chronic autoimmune disease characterized by a T-cell-mediated destruction of exocrine glands. Classic symptoms include dry eye and dry mouth from progressive damage to the lacrimal and salivary glands. Sjögren syndrome has a 9:1 female to male predominance. It can present as a primary disease or in association with a systemic connective tissue disease, commonly rheumatoid arthritis. Diagnosis of Sjögren syndrome requires the presence of ocular symptoms, including keratoconjunctivitis sicca, and ocular surface damage from dryness of the

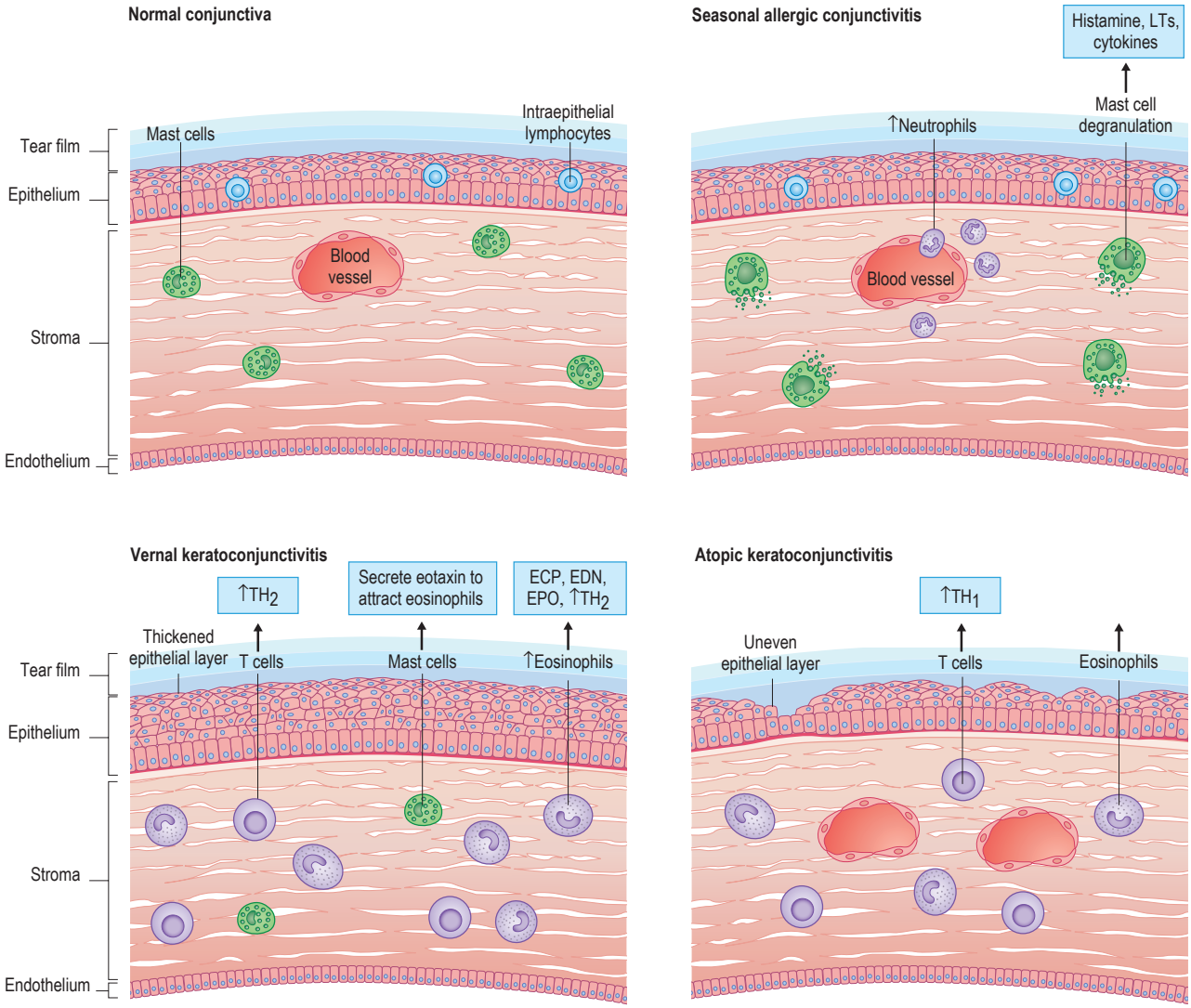


FIG 74.9 A schematic diagram of a cross-section of conjunctival tissue, showing the cell processes involved in normal vs sac- vs vkc- vs akc- affected tissues. *AKC*, Atopic keratoconjunctivitis; *ECP*, eosinophil cationic protein; *EDN*, eosinophil-derived neurotoxin; *EPO*, eosinophil peroxidase; *SAC*, seasonal allergic conjunctivitis; *VKC*, vernal keratoconjunctivitis.

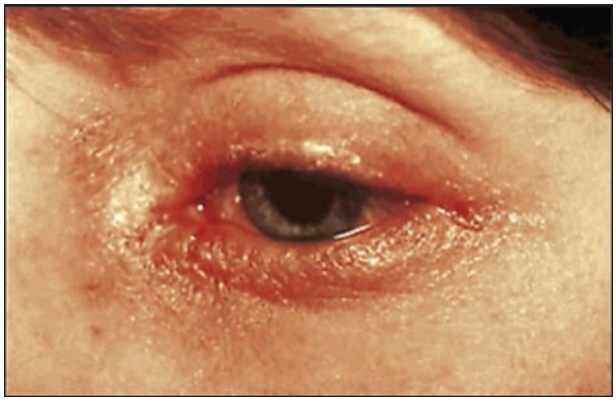


FIG 74.10 Allergic Keratoconjunctivitis (AKC). Patient with AKC demonstrating conjunctival injection, increased lacrimation, and erythematous, thickened eyelids. Patients typically have a personal history of atopy, including allergies, atopic dermatitis, eczema, or other atopic conditions. (Reproduced with permission from Conjunctivitis. Elsevier Point of Care 2019. Copyright Elsevier BV. All rights reserved.)

conjunctiva and cornea. The disease can also have optic nerve involvement.

Patients are typically seropositive for anti-SSA (Ro) or anti-SSB (La) and may also be rheumatoid factor or ANA positive.[50] Histopathology classically shows focal lymphocytic sialoadenitis. Extra-glandular manifestations are common and include polyarticular arthralgias or synovitis, gastrointestinal motility problems related to decreased salivary flow, autoimmune pancreatitis, celiac disease, chronic cough due to desiccation of the upper airways, interstitial lung disease, recurrent respiratory infections due to impaired mucociliary clearance, anemia of chronic disease, leukopenia, and lymphopenia. Approximately 5% of patients will develop non-Hodgkin B-cell lymphoma, most often presenting in the parotid gland, or other hematologic malignancy.[50]

Management strategies for the dry eye associated with Sjögren syndrome include tear substitutes, nighttime gels and ointments, topical cyclosporine, topical steroid drops, punctal occlusion, and autologous serum tears. More severe cases may require systemic anti-inflammatory medication or tarsorrhaphy (surgical closure of part of the eyelids).

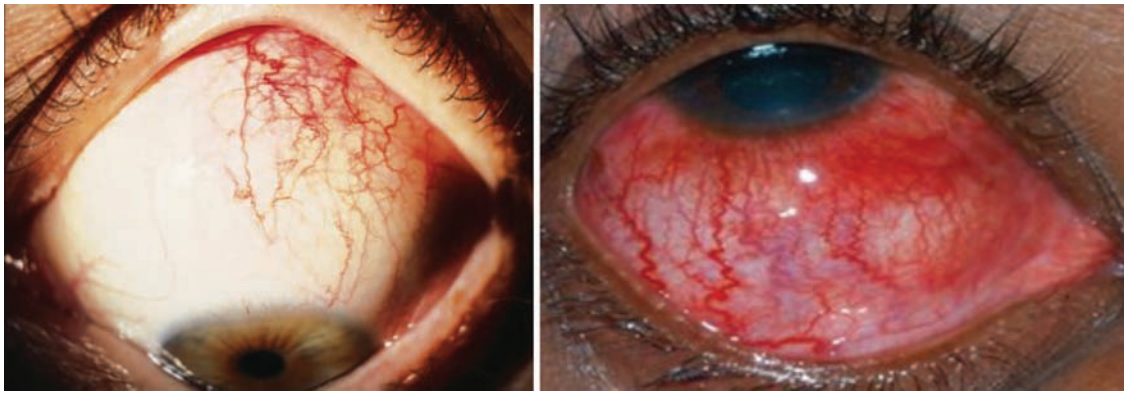


FIG 74.11 Episcleritis and Scleritis. Episcleritis with sectoral vascular dilatation of conjunctival and superficial episcleral vessels. Markle JG, Deenick EK, (Left) Cornea, 4e. Scleritis with inflammation of the sclera and deeper scleral vessels and notable violaceous hue. A discrete, raised scleral nodule is also present near the limbus at the 5 o'clock position. (Right) Scleritis is frequently associated with rheumatoid arthritis and other collagen-vascular diseases. (Reproduced with permission from Galoir A, Thorne, JE. Scleritis and peripheral ulcerative keratitis. *Rheumatic Disease Clinics of North America*. 2007; 33: Fig. 74.4.)

Episcleritis and Scleritis

Episcleritis is a mild, self-limiting inflammation of the episcleral tissue, and a common cause of red eye. Vascular congestion of the superficial episcleral plexus is typically sectoral, but may also present diffusely. The nodular form will have a discrete, elevated area of inflamed tissue (Fig. 74.11). The episclera and Tenon capsule are infiltrated with inflammatory cells on histopathologic analysis. Episcleral vessels will blanch with phenylephrine, differentiating this from a deeper scleritis. Management includes NSAIDs and a brief course of mild topical steroids. Most cases are idiopathic, but autoimmune work-up may be indicated for recurrent cases.

Scleritis is a deeper immune-mediated vasculitis of the sclera that can result in scleral thinning and perforation in severe cases. Patients present with severe pain, deep hyperemia with violaceous hue that does not blanch with phenylephrine, and scleral edema. Autoimmune work-up for systemic conditions should be pursued, with one-third of patients with diffuse or nodular anterior scleritis and two-thirds of patients with necrotizing anterior scleritis having an underlying systemic autoimmune disease, most commonly rheumatoid arthritis, lupus, spondyloarthropathies, and vasculitides such as granulomatosis with polyangiitis and polyarteritis nodosa.[51] Histopathology can show either a non-granulomatous or granulomatous process with or without associated scleral necrosis. Management includes high-dose oral NSAIDs, oral steroids, and systemic immunosuppression. Surgical intervention may be needed for scleral perforation or excessive scleral thinning.

CORNEAL TRANSPLANT AND TRANSPLANT REJECTION

The cornea was the first successfully transplanted solid tissue and remains the most commonly transplanted tissue worldwide. Corneal transplants are the only solid organ that is not typically HLA matched. Additionally, patients are not routinely cross-matched for the presence of HLA-specific antibodies. Despite this, graft survival is overall excellent due to the healthy cornea's lack of blood and lymphatic flow and unique immune privilege.

Corneal graft survival is inversely related to the degree of host corneal bed vascularity, presence of inflammation, and number of antigen-presenting Langerhans cells in the cornea.[52] Data from corneal transplant registries have shown that transplant survival for penetrating keratoplasty is primarily dependent on the indication for surgery, with 10-year graft survival rates ranging from 89% for keratoconus, 73% for Fuchs corneal dystrophy, 70% for non-herpetic corneal scars, 60% for herpetic corneal scars, and 40% for pseudophakic and aphakic corneal edema.[53] Eyes that failed one corneal transplant and required a regraft showed the lowest graft survival rate of 37% after 10 years.[53] Notably, the diagnoses with the best transplant outcomes such as keratoconus and Fuchs corneal dystrophy present with the least vascularization, while regrafts typically occur in vascularized corneal beds, with consequently lower graft survival.

Immune-mediated rejection in solid organ transplants is typically characterized as hyperacute, acute, or chronic. Hyperacute rejection mediated by alloantibodies against ABO blood group or HLA antigens does not occur in corneal transplantation. The presence of alloantibodies against a graft, for example in a patient undergoing regraft after a graft failure, can exacerbate and accelerate rejection, but alloantibodies alone are not believed to be capable of hyperacutely rejecting corneal grafts.[52] Acute rejection is initiated by host recognition of alloantigens, is both antibody and cell mediated, and manifests in days to months. Acute rejection events can occur and are typically managed with medical immunosuppression. Chronic rejection develops slowly over months or years, is largely cell mediated with T-cell and macrophage involvement and is the major source of corneal allograft failure.

Rejection of the graft epithelial layer is typically asymptomatic, as donor epithelium is replaced by host epithelium derived from the corneal limbus. Destruction of the epithelium can immunize the host against the donor but does not greatly affect graft survival.[54] Stromal rejection is relatively common, but easily treated by intensive use of topical corticosteroid drops, and rarely leads to endothelial rejection. Immune-mediated endothelial rejection is the most serious and frequently leads to corneal graft failure. Human endothelial cells are nonreplicative, and their loss is irreversible. Endothelial cell density below a critical level results in chronic corneal edema due to inability to maintain a clear dehydrated state. Endothelial cell loss due to

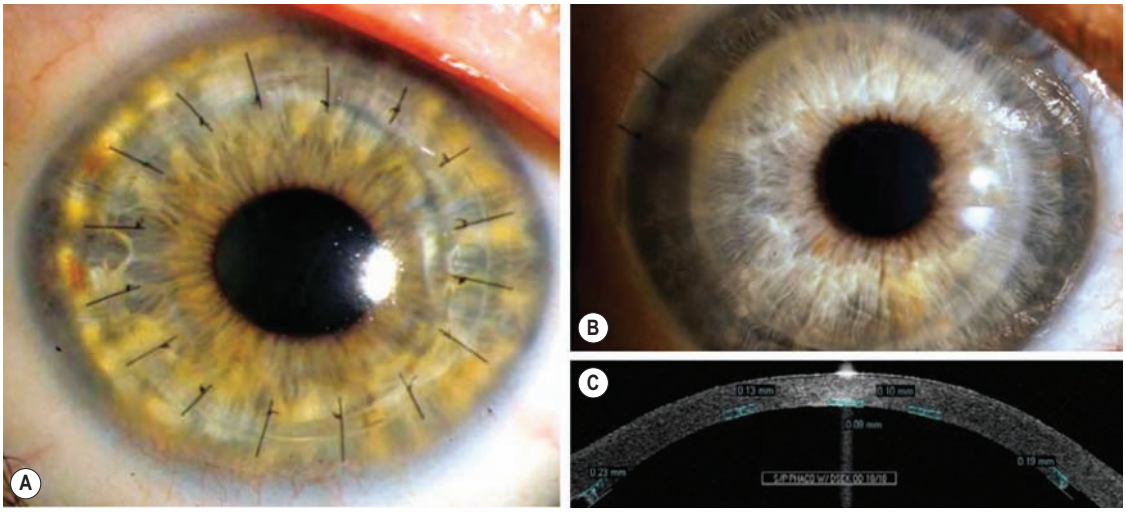


FIG 74.12 External Photograph of Cornea. (A) After a penetrating keratoplasty with interrupted sutures. (B) After Descemet's stripping automated endothelial keratoplasty. (C) With corresponding anterior segment optical coherence tomography image. (Reproduced with permission from Tan DTH, Dart JKG, Holland EJ, Kinoshita S. Corneal transplantation. *Lancet*. 2012;379: 1749–1761. Copyright © 2012 Elsevier Ltd.)

recurrent acute rejection events or chronic cell-mediated rejection can eventually lead to end-stage graft failure. Donor endothelial cells can also fail by non-immune-mediated pathways. Endothelial cell density rapidly falls during ex vivo storage, and graft endothelial cells have accelerated rate of loss even in the absence of overt rejection.[54] Donor endothelial cells after penetrating keratoplasty have been measured to have a half-life of approximately 21 years, over 10-fold less than anticipated for normal aging cell loss.[55] Progressive endothelial cell death over time can thus result in late graft failure decades after transplant. Other cases of transplant failure can occur due to recurrence of disease, most commonly for infectious disorders such as herpetic and fungal keratitis, or failure to heal in anesthetic corneas.

Although HLA matching of corneal transplants has been generally assumed to be unnecessary, extensive studies do show a reduction in rejection in both high-risk and normal-risk recipients who are HLA matched.[52] Additionally, pre-transplant cross-matching to screen for HLA-specific antibodies may be of benefit in high-risk or retransplant patients. However, the value of routine HLA matching for corneal transplants is controversial, given the cost and complexity of matching, delay in surgical treatment, high success rates in the absence of matching, and availability and success of immunomodulatory agents and biologics for the prophylaxis and treatment of corneal graft rejection.

Newer forms of lamellar transplantation surgery in which only specific layers of the cornea are removed and transplanted to address individual diseases have gained favor over traditional full-thickness penetrating keratoplasty.[56] Rejection rates for these selective lamellar keratoplasty techniques have the potential to be lower than for traditional penetrating keratoplasty; however, long-term follow up will define the success of the newer surgical techniques. Corneal transplantation remains the most successful type of organ transplantation, with long-term graft survival rates in non-matched corneal transplant patients even higher than rates seen in HLA-matched transplants of other solid organs (Fig. 74.12).

KEY CONCEPTS

Corneal Transplant

Factors mediating corneal transplant graft survival:
- Indication for surgery
- Host corneal bed vascularity
- Presence of inflammation
- Number of antigen-presenting Langerhans cells in cornea
- Prior graft failure

Indications with high corneal graft survival:
- Keratoconus
- Fuchs corneal dystrophy
- Other endothelial and stromal dystrophies
- Non-herpetic corneal scars

Indications with low corneal graft survival:
- Herpetic corneal scars
- Chemical burns
- Corneal regraft for previous transplant failure

EMERGING TOPICS IN IMMUNOLOGIC OCULAR DISEASES

Commensal Microbiota in Ocular Immunology

The role of the microbiome has been traditionally focused on its role in gastrointestinal diseases; however, increasing evidence suggests that gut commensals affect all aspects of immune homeostasis and have many effects in distant tissues, including those of the eye.[57] There is evidence that gut microbiota can trigger development of uveitis. Studies in the R161H mouse model of spontaneous T-cell-driven uveitis show that depletion of the gut microbiota by broad-spectrum antibiotics or rearing in germ-free conditions attenuates disease, and disease development was associated with increased populations of Th17 cells in the intestinal lamina propria.[58] Prevailing hypotheses include the idea that retina-specific T cells may be primed in the gut, possibly though an unidentified commensal antigen mimic of

retinal antigens. Bacteria-rich intestinal contents can activate R161H T cells in vitro and transfer of these activated cells can confer disease to naïve wild-type mice, supporting an antigenic mimicry hypothesis.[58] Alternatively, microbial products and metabolites may provide adjuvant innate signals that modify the host response toward developing uveitis. An activation step in the periphery is likely important for autoimmune ocular diseases, as antigens in a healthy eye are sequestered, requiring T cells to be activated to breach the blood-retinal barrier. Microbial exposure may be an important causative or modulatory factor in the initiation and severity of uveitis in some models. However, commensals are not the only disease trigger, as R161H mice will develop uveitis with reduced disease scores over time, even in germ-free conditions. Additionally, the AIRE$^{-/-}$ mouse model develops spontaneous autoimmune uveitis that is not dependent on the presence of commensal microbiota.[59] The relationship of the commensal microbiome to ocular disease is likely complex, and further research is needed. However, the possibility of modulating the gut microbiota through antibiotics, probiotics, prebiotics, or fecal transplants as a therapeutic approach in the treatment of uveitis and other autoimmune diseases is intriguing and warrants further exploration.

Cancer Immunotherapy-Associated Ocular Complications

A revolution is taking place not only in the treatment of autoimmune diseases with increasing development of biologics but also in the field of cancer therapy with the advent of cancer immunotherapy drugs. We are now in an era where we are able to harness the natural power of the immune system to detect and eliminate transformed cells. Cancer immunotherapy acts by enhancing immune function to amplify antitumor response and aid in eradication of existing cancers and metastases. However, the side effects seen with immunotherapy drugs reflect this immune enhancement and disruption of self-tolerance, resulting in high rates of autoimmune toxicity. Autoimmune-related side effects are common and can affect almost any organ, including causing both uveitis and neuro-ophthalmic complications of the eye. The study of immune-related adverse events (irAEs) and their management is a critical area of immunotherapeutic research, as the development of autoimmune-related toxicity frequently limits the use of these otherwise effective cancer therapeutics.

There are currently several distinct classes of immunotherapeutic drugs, including antitumor vaccination, cellular immunotherapy, cytokine therapy, and monoclonal antibody-based therapies. Uveitis associated with anti-CTLA-4 and anti-PD-1/PD-L1 checkpoint inhibitor use has been well documented.[60] Other ophthalmic complications have recently been recognized, ranging from optic neuritis, myasthenia gravis–like presentations, and muscle/orbit involvement. This illustrates the vast range of ocular presentations that may result from immunotherapy treatment, and highlights the fact that management of patients on complex immunotherapeutic regimens will require a multidisciplinary approach.

Gene Therapy

The eye is an ideal location to pioneer advancements in gene therapy. The eye is uniquely immune privileged, yet has easy access for surgical delivery with subretinal, intravitreal, and other minimally invasive available intraocular routes of administration. The highly compartmentalized nature of the eye and separation from the rest of the body by the blood-brain barrier also minimizes concerns for off-target effects. The use of gene therapy to permanently correct monoallelic genetic disorders has been an elusive goal, until recently. Voretigene neparvovec (Luxturna, Spark Therapeutics) was approved by the US FDA in 2017 as the first in vivo gene replacement therapy. It is designed as an AAV2 vector containing a functional copy of human RPE65 cDNA with a modified Kozak sequence, administered subretinally, and is the first and only treatment of Leber congenital amaurosis, an autosomal recessive RPE65-mediated inherited retinal disease that causes progressive blindness.[61] This targeted gene therapy resulted in significant visual improvement which is maximal by 30 days after administration and durable for at least 4 years with continued ongoing observation.[62]

The early success of voretigene neparvovec provides tremendous encouragement to the field of gene therapy and future treatments of diseases such as Stargardt macular degeneration, X-linked retinoschisis, choroideremia, Leber hereditary optic neuropathy, and others. The introduction of CRISPR gene editing has further opened new possibilities in this field. We expect additional rapid advancements, and we anticipate that ophthalmology will continue to play a leading role due to the unique location and immune environment of the eye.

ON THE HORIZON

Topics of research interest:
- Immune mechanisms in degenerative diseases, including age-related macular degeneration (AMD)
- Role of commensal microbiota in ocular immunology

Development of new drugs and new therapeutics:
- Biologics: TNF inhibitors, IL-1 inhibitors, IL-6 inhibitors, etc.
- Gene therapy for ocular disease

Investigation of new diseases:
- Checkpoint-inhibitor-associated uveitis and neuro-ophthalmologic complications

SUMMARY

Multiple mechanisms have evolved to protect vision, including the blood-retinal barrier, immune privilege, and tolerance mechanisms to safeguard the eye from immune cell infiltration. However, this special immune status of the eye can paradoxically leave it vulnerable in states of autoimmunity and disease, as eye antigens are normally sequestered and not available during the development of induced Tregs that mediate peripheral tolerance. Immune diseases of the eye can affect any ocular tissue, and as such are highly varied in etiology and presentation. Frequently, these ocular diseases present in the setting of a broader systemic autoimmune disease or in the setting of general autoimmune susceptibility. Many MHC associations have been identified, and some putative targets or initiating factors have been proposed for some diseases. However, the targets for most autoimmune ocular diseases, preceding infectious triggers, and other risk factors are still not well defined and require further study. We expect that there will be many cutting-edge developments in the field of ocular immunology in the coming years, with resolution of many of these unanswered questions.

REFERENCES

1. Knop E, Knop N. Anatomy and immunology of the ocular surface. *Chem Immunol Allergy.* 2007;92:36–49.
2. Niederkorn JY. See no evil, hear no evil, do no evil: the lessons of immune privilege. *Nat Immunol*. 2006;7(4):354–359.
3. Zhou R, Caspi RR. Ocular immune privilege. *F1000 Biol Rep.* 2010;2:3.
4. Streilein JW. Ocular immune privilege: therapeutic opportunities from an experiment of nature. *Nat Rev Immunol*. 2003;3(11):879–889.
5. Forrester JV, Xu H. Good news-bad news: the Yin and Yang of immune privilege in the eye. *Front Immunol.* 2012;3:338.
6. Vendomele J, Khebizi Q, Fisson S. Cellular and molecular mechanisms of anterior chamber-associated immune deviation (ACAID): what we have learned from knockout mice. *Front Immunol.* 2017;8:1686.
7. Gery I, Caspi RR. Tolerance induction in relation to the eye. *Front Immunol.* 2018;9:2304.
8. Caspi RR. Ocular autoimmunity: the price of privilege? *Immunol Rev*. 2006;213:23–35.
9. Patel SJ, Lundy DC. Ocular manifestations of autoimmune disease. *Am Fam Physician*. 2002;66(6):991–998.
10. Iqbal MM, Hodge WG. Immunologic diseases of the eye. In: Riordan-Eva P, Augsburger JJ, eds. *Vaughan & Asbury's General Ophthalmology*. 19th ed. New York, NY: McGraw-Hill Education; 2017.
11. Nussenblatt RB, Mittal KK, Ryan S, et al. Birdshot retinochoroidopathy associated with HLA-A29 antigen and immune responsiveness to retinal S-antigen. *Am J Ophthalmol.* 1982;94(2):147–158.
12. Goverdhan SV, Lotery AJ, Howell WM. HLA and eye disease: a synopsis. *Int J Immunogenet*. 2005;32(6):333–342.
13. Whitcup SM. The double-edged ocular immune response: the Cogan lecture. *Invest Ophthalmol Vis Sci*. 2000;41(11):3243–3248.
14. Jabs DA, Nussenblatt RB, Rosenbaum JT. Standardization of Uveitis Nomenclature Working Group. Standardization of uveitis nomenclature for reporting clinical data. Results of the First International Workshop. *Am J Ophthalmol*. 2005;140(3):509–516.
15. Rosenbaum JT. New developments in uveitis associated with HLA. *Curr Opin Rheumatol*. 2017;29(4):298–B303.
16. Rosenbaum JT. Uveitis in spondyloarthritis including psoriatic arthritis, ankylosing spondylitis, and inflammatory bowel disease. *Clin Rheumatol*. 2015;34(6):999–1002.
17. Wakefield D, Yates W, Amjadi S, McCluskey P. HLA-B27 Anterior uveitis: immunology and immunopathology. *Ocul Immunol Inflamm*. 2016;24(4):450–459.
18. Sen ES, Ramanan AV. Juvenile idiopathic arthritis-associated uveitis. *Best Pract Res Clin Rheumatol*. 2017;31(4):517–534.
19. Quentin CD, Reiber H. Fuchs heterochromic cyclitis: rubella virus antibodies and genome in aqueous humor. *Am J Ophthalmol*. 2004;138(1):46–54.
20. Liu Y, Takusagawa HL, Chen TC, Pasquale LR. Fuchs heterochromic iridocyclitis and the rubella virus. *Int Ophthalmol Clin*. 2011;51(4):1–12.
21. Pakzad-Vaezi K, Pepple KL. Tubulointerstitial nephritis and uveitis. *Curr Opin Ophthalmol*. 2017;28(6):629–635.
22. Jamilloux Y, Kodjikian L, Broussolle C, Seve P. Sarcoidosis and uveitis. *Autoimmun Rev*. 2014;13(8):840–849.
23. Gül A. Pathogenesis of Behçet's disease: autoinflammatory features and beyond. *Semin Immunopathol*. 2015;37(4):413–418.
24. Vitale AT. Birdshot retinochoroidopathy. *J Ophthalmic Vis Res*. 2014;9(3):350–361.
25. Levinson RD. Killer immunoglobulin-like receptor genes in uveitis. *Ocul Immunol Inflamm*. 2011;19(3):192–201.
26. Du L, Kijlstra A, Yang P. Vogt-Koyanagi-Harada disease: novel insights into pathophysiology, diagnosis and treatment. *Prog Retin Eye Res*. 2016;52:84–111.
27. Sakata VM, da Silva FT, Hirata CE, et al. Diagnosis and classification of Vogt-Koyanagi-Harada disease. *Autoimmun Rev*. 2014;13(4–5):550–555.
28. Levinson RD, Du Z, Luo L, et al. KIR and HLA gene combinations in Vogt-Koyanagi-Harada disease. *Hum Immunol*. 2008;69(6):349–353.
29. Burkholder BM. Vogt-Koyanagi-Harada disease. *Curr Opin Ophthalmol*. 2015;26(6):506–511.
30. Aisenbrey S, Luke C, Ayertey HD, et al. Vogt-Koyanagi-Harada syndrome associated with cutaneous malignant melanoma: an 11-year follow-up. *Graefes Arch Clin Exp Ophthalmol*. 2003;241(12):996–999.
31. Chang GC, Young LH. Sympathetic ophthalmia. *Semin Ophthalmol*. 2011;26(4–5):316–320.
32. Cunningham ET, Kilmartin D, Agarwal M, Zierhut M. Sympathetic ophthalmia. *Ocul Immunol Inflamm*. 2017;25(2):149–151.
33. Pasadhika S, Rosenbaum JT. Update on the use of systemic biologic agents in the treatment of noninfectious uveitis. *Biologics.* 2014;8:67–81.
34. Bennett JL. Optic neuritis. *Continuum (Minneap Minn)*. 2019;25(5):1236–1264.
35. Chen JJ, Pittock SJ, Flanagan EP, et al. Optic neuritis in the era of biomarkers. *Surv Ophthalmol*. 2020;65(1):12–17.
36. Wingerchuk DM, Banwell B, Bennett JL, et al. International consensus diagnostic criteria for neuromyelitis optica spectrum disorders. *Neurology.* 2015;85(2):177–189.
37. Horton L, Bennett JL. Acute management of optic neuritis: an evolving paradigm. *J Neuroophthalmol*. 2018;38(3):358–367.
38. Weyand CM, Liao YJ, Goronzy JJ. The immunopathology of giant cell arteritis: diagnostic and therapeutic implications. *J Neuroophthalmol*. 2012;32(3):259–265.
39. Watanabe R, Goronzy JJ, Berry G, et al. Giant cell arteritis: from pathogenesis to therapeutic management. *Curr Treatm Opt Rheumatol*. 2016;2(2):126–137.
40. Saul RF. Neuro-ophthalmology and the anti-GQ1b antibody syndromes. *Curr Neurol Neurosci Rep*. 2009;9(5):379–383.
41. Fortin E, Cestari DM, Weinberg DH. Ocular myasthenia gravis: an update on diagnosis and treatment. *Curr Opin Ophthalmol*. 2018;29(6):477–484.
42. Nair AG, Patil-Chhablani P, Venkatramani DV, Gandhi RA. Ocular myasthenia gravis: a review. *Indian J Ophthalmol*. 2014;62(10):985–991.
43. Smith SV, Lee AG. Update on ocular myasthenia gravis. *Neurol Clin*. 2017;35(1):115–123.
44. Gordon L, Dinkin M. Paraneoplastic syndromes in neuro-ophthalmology. *Continuum (Minneap Minn)*. 2019;25(5):1401–1421.
45. Rahimy E, Sarraf D. Paraneoplastic and non-paraneoplastic retinopathy and optic neuropathy: evaluation and management. *Surv Ophthalmol*. 2013;58(5):430–458.
46. Ling CP, Pavesio C. Paraneoplastic syndromes associated with visual loss. *Curr Opin Ophthalmol*. 2003;14(6):426–432.
47. Eksandh L, Adamus G, Mosgrove L, Andreasson S. Autoantibodies against bestrophin in a patient with vitelliform paraneoplastic retinopathy and a metastatic choroidal malignant melanoma. *Arch Ophthalmol*. 2008;126(3):432–435.
48. Koreen L, He SX, Johnson MW, et al. Anti-retinal pigment epithelium antibodies in acute exudative polymorphous vitelliform maculopathy: a new hypothesis about disease pathogenesis. *Arch Ophthalmol*. 2011;129(1):23–29.
49. La Rosa M, Lionetti E, Reibaldi M, et al. Allergic conjunctivitis: a comprehensive review of the literature. *Ital J Pediatr*. 2013;39(1):18.
50. Vivino FB. Sjogren's syndrome: clinical aspects. *Clin Immunol*. 2017;182:48–54.
51. Okhravi N, Odufuwa B, McCluskey P, Lightman S. Scleritis. *Surv Ophthalmol*. 2005;50(4):351–363.
52. van Essen TH, Roelen DL, Williams KA, Jager MJ. Matching for human leukocyte antigens (HLA) in corneal transplantation—to do or not to do. *Prog Retin Eye Res*. 2015;46:84–110.
53. Williams KA, Lowe M, Bartlett C, et al. Risk factors for human corneal graft failure within the Australian corneal graft registry. *Transplantation.* 2008;86(12):1720–1724.
54. George AJ, Larkin DF. Corneal transplantation: the forgotten graft. *Am J Transplant*. 2004;4(5):678–685.
55. Armitage WJ, Dick AD, Bourne WM. Predicting endothelial cell loss and long-term corneal graft survival. *Invest Ophthalmol Vis Sci*. 2003;44(8):3326–3331.
56. Tan DT, Dart JK, Holland EJ, Kinoshita S. Corneal transplantation. *Lancet.* 2012;379(9827):1749–1761.
57. Horai R, Caspi RR. Microbiome and autoimmune uveitis. *Front Immunol.* 2019;10:232.

58. Horai R, Zarate-Blades CR, Dillenburg-Pilla P, et al. Microbiota-dependent activation of an autoreactive T cell receptor provokes autoimmunity in an immunologically privileged site. *Immunity.* 2015;43(2):343–353.
59. Gray DH, Gavanescu I, Benoist C, Mathis D. Danger-free autoimmune disease in Aire-deficient mice. *Proc Natl Acad Sci U S A.* 2007;104(46):18193–18198.
60. Sun MM, Levinson RD, Filipowicz A, et al. Uveitis in patients treated with CTLA-4 and PD-1 checkpoint blockade inhibition. *Ocul Immunol Inflamm.* 2019:1–11.
61. Russell S, Bennett J, Wellman JA, et al. Efficacy and safety of voretigene neparvovec (AAV2-hRPE65v2) in patients with RPE65-mediated inherited retinal dystrophy: a randomised, controlled, open-label, phase 3 trial. *Lancet.* 2017;390(10097):849–860.
62. Maguire AM, Russell S, Wellman JA, et al. Efficacy, safety, and durability of voretigene neparvovec-rzyl in RPE65 mutation-associated inherited retinal dystrophy: results of phase 1 and 3 trials. *Ophthalmology.* 2019;126(9):1273–1285.

Immunologic Diseases of the Gastrointestinal Tract

Peter J. Mannon

GASTRITIS

Gastritis is a histological term that differs from normal gastric mucosa (Fig. 75.1, A). It describes stomach inflammation resulting from toxic exposures, infection, idiopathic inflammation, and autoimmunity. Although symptoms tend to be very nonspecific, the etiology and treatment of gastritis can be extremely specific, particularly for *Helicobacter pylori* infection, which has important implications for outcomes and natural history.

Atrophic Gastritis/Pernicious Anemia

The most classic autoimmune disease of the gastrointestinal (GI) tract is atrophic gastritis (AG) characterized by loss of acid-producing parietal cells, anti-intrinsic factor, and anti-parietal cell autoantibodies, and an association with autoimmune thyroiditis, vitiligo, and type 1 diabetes.[1] The resulting loss of acid production interferes with inorganic iron absorption, causes hypergastrinemia and enterochromaffin cell hyperplasia, and B_{12} deficiency ("pernicious anemia [PA]"). Despite the age-old stereotype of the AG/PA patient as an older Caucasian woman (there is a female preponderance 2:1), 25% of cases are diagnosed below age 50 years, and disease does occur in African and Asian ethnicities. The detection of both anti-intrinsic factor and anti-parietal cell serum antibodies has 73% sensitivity and 100% specificity for PA.[2] Murine models show that immunity to H^+-K^+-ATPase (the parietal cell membrane protein that secretes H ions into the gastric lumen) results in AG.[3] Although potentially pathogenic H^+-K^+-ATPase–responsive CD4 T cells can be isolated from the gastric mucosa of patients with AG, it remains unclear how these cells arise, but observations implicate a role for *H. pylori* infection (>80% of patients with AG have antibodies to *H. pylori*).[4] Antral inflammation (up to 92%) and mucosal atrophy (up to 30%) together with histological evidence of *H. pylori* infection (up to 30%) occur in AG, with antigastric antibodies present in up to 65% of patients with *H. pylori* infection (and none in noninfected patients with gastritis).[5]

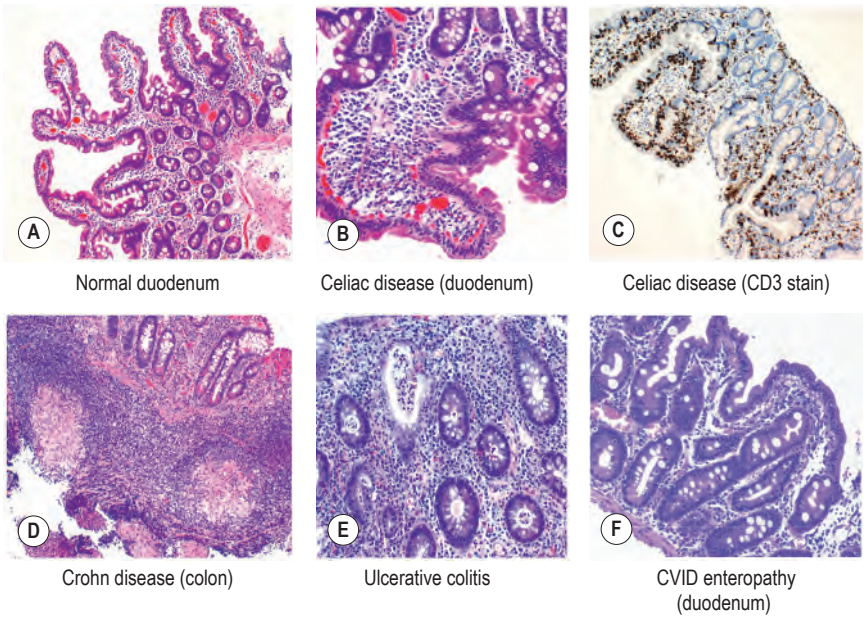


FIG. 75.1 Gastrointestinal Histology in Health and Immune-Mediated Disease. (A) Normal duodenal histology. (B) Celiac disease with blunted villus, increased plasma cell infiltrate, increased intraepithelial lymphocytes. (C) CD3 staining in celiac disease showing increased intraepithelial lymphocytes. (D) Crohn colitis showing mucosa expanded with lymphoplasmacytic infiltrate and two granulomata. (E) Ulcerative colitis showing crypt dropout, cryptitis, crypt abscess, and lymphoplasmacytic infiltrate. (F) Common variable immunodeficiency *(CVID)* enteropathy showing villus blunting, increased intraepithelial lymphocytes, and epithelial apoptosis. (Photomicrographs courtesy of Dr. Leona Council, Department of Pathology, University of Alabama at Birmingham, Birmingham, Alabama, USA.)

The diagnosis of AG/PA relies on the combination of laboratory, serological, and histological data.[6] AG itself does present a risk of cancer (neuroendocrine cell and adenocarcinoma) and guidelines are still in development as to surveillance.[7]

Helicobacter pylori Gastritis

H. pylori infection of the gastric mucosa is the leading cause of peptic ulcer disease and is a World Health Organization (WHO)-designated class I carcinogen for gastric carcinoma. Primary infection is largely acquired in childhood, and poor sanitation enhances fecal/gastro-oral transmission. Although acute infection can cause abdominal pain and dyspepsia, there is little clinical recognition of acute infection. The burden of *H. pylori* ensues from chronic infection of the stomach; *H. pylori* is uniquely adapted to the acidic environment of the stomach through its ability to metabolize urea to acid-buffering ammonia that allows prolonged asymptomatic colonization.

The ability of *H. pylori* to confer risk of peptic ulcer disease in 15% of chronically infected persons has been linked to *H. pylori* infection inhibition of local bicarbonate secretion by the gastric epithelium which enhances permeability to damaging hydrogen ions. In addition, expression of cytotoxins (vacuolating cytotoxin, VacA) and pro-inflammatory virulence factors, such as CagA, may contribute to pathogenesis.

The development of peptic ulcer disease and adenocarcinoma caused by chronic *H. pylori* infection correlates with the anatomical distribution of inflammation. When *H. pylori* chronic gastritis affects the antrum predominantly, there is an association with duodenal ulcers, increased serum gastrin levels and excess acid production, and no gastric mucosal atrophy.[8] However, when *H. pylori* affects the body and the antrum in a confluent or patchy manner, intestinal metaplasia develops, oxyntic mucosa atrophies, and acid production decreases; this type of *H. pylori* chronic gastritis is associated with gastric ulcerations and increased risk for adenocarcinoma and mucosa-associated lymphoreticular tissue (MALT) B-cell lymphoma. Although eradication of *H. pylori* may reverse the mucosal atrophy and restore acid production at this stage, mucosal restoration occurs only in a minority of patients and does not necessarily reverse the intestinal metaplasia.[4] Histologically, this form of AG is differentiated from primary autoimmune AG (above) by the presence of *H. pylori* organisms in the specimen (and concurrent involvement of the antrum). Specific immunohistochemical methods for detecting *H. pylori* are required when organisms are not seen on hematoxylin and eosin staining, when intestinal metaplasia occurs widely (*H. pylori* does not colonize intestinal metaplasia heavily), or when confirming *H. pylori* eradication after antibiotic treatment.[9] Immune mechanisms contribute to the persistence of *H. pylori* infection in the stomach, including pro-regulatory effects of local interleukin-10 (IL-10) production, increased regulatory T cells (Tregs) in the gastric mucosa, and increased antigen-presenting cell (APC) phagocytosis of apoptotic cells.[10–12] Establishing the diagnosis of active *H. pylori* infection is essential in the setting of active gastroduodenal ulcer disease, gastric MALT lymphoma, and, in highly endemic areas such as Central America, dyspepsia symptoms alone (upper abdominal pain, bloating, early satiety, and nausea). Active disease is diagnosed with endoscopic biopsy, which has high sensitivity and specificity, while simultaneously assessing peptic and malignant complications. Noninvasive testing for *H. pylori* infection includes serum antibody detection (best used in highly endemic areas to predict active infection), urea breath testing (limited by expense and possible false-positive results), and fecal antigen testing (which has potential advantages in the setting of intestinal metaplasia and after antibiotic treatment).

Once *H. pylori* infection is diagnosed, there are many effective eradication therapies that need to be tailored to patients' drug tolerance and allergy history as well as local antibiotic resistance patterns. In general, a 14-day course with a proton pump inhibitor (PPI; histamine 2 [H_2] blockers may be substituted) and two antibiotics (clarithromycin with amoxicillin or metronidazole) is recommended as first-line treatment.[13] Alternative regimens, including bismuth or sequential therapy, may be needed in cases of antibiotic resistance. Eradication of infection can be confirmed by either invasive or noninvasive (but not serum antibody) methods. In addition to the cure of recurrent gastroduodenal ulcer disease, eradication of *H. pylori* can cause regression of gastric MALT lymphoma in a majority of coinfected patients.[14]

KEY CONCEPTS

Gastritis

- Atrophic gastritis (AG) results from immune-mediated loss of the acid-producing cells of the gastric body. AG is associated with anti-parietal cell antibodies, an intact antral gastric mucosa, and no evidence of *Helicobacter pylori* infection.
- However, *H. pylori* infection has been linked to atrophic gastritis via increased rates of anti-*H. pylori* antibodies in AG patients and gastric mucosal CD4 T cells with cross-reactivity to H^+-K^+-ATPase and *H. pylori* antigens.
- *H. pylori* infection of the stomach is chronic because of metabolic and immune adaptations by the pathogen that allow persistence in the acidic gastric environment.
- *H. pylori* infection is a leading cause of peptic ulcer disease and a World Health Organization (WHO) class I carcinogen because of its link to gastric adenocarcinoma.

ON THE HORIZON

Gastritis

- Development of a *Helicobacter pylori* eradication treatment with affordable global implementation without significant adverse effects or antibiotic resistance that work to enhance current regimens or replace them (e.g., vaccines).

CELIAC DISEASE

Celiac disease is the most prevalent (0.5% to 1% of the general population) immune-mediated disease of the human GI tract.[15] Gluten-derived peptides not only disrupt the tight junction complex (via CXCR3-stimulated release of zonulin) but also gluten peptide–specific T cells drive specific autoantibody formation and an inflammatory response that leads to small intestinal villus atrophy and malabsorption caused by subsequent intestinal surface injury. The identification of the specific gluten peptide-presenting human leukocyte antigen (HLA) molecules (DQ2 and DQ8) and the gluten peptide cognate ligands that activate the T cells has advanced the understanding of the pathophysiology and genetics of celiac disease, driving the development of innovative therapies.[16]

Presentation

The classic clinical presentation of celiac disease—chronic diarrhea, weight loss, and abdominal bloating—results from defective nutrient absorption by the small intestine as a result

of inflammatory injury. However, adult patients with celiac disease are more likely to present with complications of nutrient deficiency (anemia, osteoporosis, sarcopenia) without overt GI complaints. Celiac disease can also present with an associated skin eruption (dermatitis herpetiformis), cerebellar ataxia, infertility and miscarriage, and chronic fatigue, and can be associated with autoimmune disorders, such as type 1 diabetes and thyroiditis, Addison disease, Sjögren syndrome, autoimmune hepatitis, and primary biliary cirrhosis.[15] The high frequency of nonspecific GI symptoms, such as abdominal pain and constipation, which are reported in over one-fifth of newly diagnosed subjects, complicates targeted screening. The variations of celiac disease presentation reflect several features: *(i)* the dose effect of celiac disease risk alleles on severity; *(ii)* the development of celiac disease involves additional genetic and environmental factors as well as the required HLA background; *(iii)* celiac disease activity depends on the quantity and quality of gluten exposure; and *(iv)* celiac disease effects are proportionate to the inflammatory activity and the extent of involved bowel. The concept of a vast subclinical celiac disease prevalence has led to an "iceberg model," where the visible tip is the group with symptomatic GI disease and the characteristic gut mucosal lesion, and below the surface are those with subclinical disease (no overt symptoms but gut mucosal lesions) and latent disease (no symptoms or mucosal lesions but possibly positive serology against a background of HLA genetic susceptibility).

The key to diagnosing celiac disease is simply to consider it in the differential diagnosis of classic gut malabsorption as well as subtle manifestations of malabsorption (e.g., unexplained iron deficiency anemia). Conversely, the differential diagnosis of villus blunting or intraepithelial lymphocytosis (without positive serologies) includes small-intestinal bacterial overgrowth, tropical sprue, autoimmune enteropathy, common variable immunodeficiency (CVID) enteropathy, and *H. pylori* gastritis, emphasizing that celiac disease is also not solely a histological diagnosis.

Immune Pathophysiology

Celiac disease results from the activation of T cells by gluten peptide–major histocompatibility complex (MHC) complexes on APCs in the lamina propria of the gut (primarily the small intestine). Dietary gluten, largely from wheat, barley, and rye, exists in polymeric (glutenin) and monomeric (gliadin) form and is incompletely digested to small peptides by gut luminal enzymes because of their high glutamine and proline content.[17] These large gluten peptides cross the zonulin-disrupted epithelial barrier and bind to specific HLA-DQ2 or -DQ8 molecules (tissue transglutaminase [TTG] can deamidate gluten peptides, resulting in negatively charged gluten peptides with increased affinity for the HLA binding site). Gut microbes also can affect the immunogenicity of gluten peptides via their own proteolytic enzymes but a specific celiac disease dysbiotic signature has not been defined.[18,19] The gluten peptide-activated T cells produce proinflammatory cytokines interferon-γ (IFN-γ), interleukin-18 (IL-18), and IL-21. The activated T cells also induce B-cell maturation to plasma cells, producing antibodies to gluten peptides as well as to TTG. The reason for TTG being targeted for autoantibody production is unknown.

Observed in animal models, the activated T cells are not required, but are sufficient to induce the epithelial damage and villus blunting. The characteristic villus atrophy is induced by gut APC- and epithelial cell–produced IL-15, enhancing CD8 T-cell infiltration into the epithelium where intraepithelial T lymphocytes (IELs) with upregulated NKG2D receptors, interacting with Major histocompatibility complex (MHC) class I chain-related protein A (MICA) and B upregulated on epithelial cells, induce cysteinyl leukotrienes that drive IEL cytotoxicity.[20]

Beyond the HLA associations with disease, genome-wide association studies (GWAS) have linked over 30 disease susceptibility loci with non-HLA regions such as the polymorphism in *lnc13*, a long noncoding RNA, which affects its binding with a heterogeneous nuclear ribonucleoprotein to undo its repression of other genes, including inflammatory mediators.[21]

Diagnosis

Celiac disease patients can present with both suggestive symptoms and signs (weight loss, chronic diarrhea) as well as in atypical ways, such as with specific micronutrient deficiencies or unexplained hyperamylasemia or hypertransaminasemia. Initial tests include measuring serum immunoglobulin A (IgA) antibodies against TTG and endomysial proteins, which have an estimated specificity/sensitivity of 95%/95% and 100%/>90%, respectively. Total serum IgA level must be made at the same time to prevent false–negative results; however, in the setting of IgA deficiency (there is an association with celiac disease), elevated IgG anti–TTG or anti-deamidated gliadin levels and identification of celiac disease susceptibility HLA genes can help determine the risk and presence of disease.[15]

An important part of celiac diagnosis is biopsy of the upper small intestine mucosa (see Fig. 75.1, B and C); typically three to four endoscopic biopsy specimens are obtained from both the duodenal bulb and the distal duodenum. Although the absolute requirement for histological diagnosis of celiac disease may be debated (and cannot be used alone because of lack of specificity), it remains important for several reasons: *(i)* the serological markers should only be used as a screening test, identifying which patients are at highest risk for the disease and appropriate for biopsy confirmation; *(ii)* even in people with HLA DQ-2 or DQ-8 backgrounds, only a minority of persons will develop symptomatic celiac disease, so the evaluation of atypical presentations especially will require histological examination; and *(iii)* because the treatment can be life altering, it is essential to make a definitive diagnosis.

Treatment

The treatment of celiac disease is avoidance of gluten, specifically foods containing wheat, barley, and rye. Successful gluten avoidance should first resolve symptoms and clinical lab abnormalities (iron deficiency, for instance). Follow-up endoscopy to assess response to therapy should be done only after 6 to 12 months of a strict gluten-free diet (GFD), although full recovery of the villus mucosa may take several years. There are no accurate biomarkers to monitor adherence to a GFD, although one indication may be a fall in the pretreatment level IgA anti-TTG serum antibodies; therefore, follow-up endoscopy with biopsy is needed to document restoration of the villus architecture.

About 5% of patients with celiac disease do not respond to a GFD. Ensuring a strict adherence to a GFD is important to identify the reasons for nonresponse, whether through inadvertent gluten exposure or whether the inflammation is truly refractory to a strict GFD. One group of patients with the so-called refractory celiac disease with polyclonal IEL populations may respond to corticosteroids and immunosuppressant treatment; another group with monoclonal IELs do not respond to such treatment and are at increased risk of lymphoma.[22]

The majority of patients with celiac disease respond positively to a GFD with return of normal gut absorption. However, ongoing inflammation is associated with risk of small-bowel lymphoma, so ensuring adherence to a GFD and documenting mucosal healing can affect the natural history of this disease. Finally, since first-degree relatives are at increased risk of celiac disease, patients should be advised of recommended serum antibody screening of these family members.

KEY CONCEPTS

Celiac Disease

- Patients with celiac disease present more often with complications of malabsorption than chronic diarrhea and weight loss.
- Human leukocyte antigen (HLA)-DQ2 or -DQ8 alleles are necessary but not sufficient for celiac disease to develop, as they are present in more persons unaffected by celiac disease.
- Immunoglobulin A (IgA) antibodies to tissue transglutaminase and endomysial proteins should be used for screening only (not diagnosis) and should be measured along with total IgA for validity.
- The goals of gluten-free diet treatment are relief of symptoms, reversal of malabsorption, and restoration of villi.

ON THE HORIZON

Celiac Disease

Discovery of additional genetic and environmental factors (beyond gluten) that strongly confer risk for celiac disease in the setting of human leukocyte antigen (HLA)-DQ2 and -DQ8 alleles leading to strategies to eliminate that risk and prevent disease.

- Novel treatment approaches to modify gluten to nonimmunogenic forms, anti-IL-15 strategies that interrupt the inflammatory cascade of events, and even vaccines to downregulate responses to gluten peptides or that induce tolerance to gliadin.
- Identification of specific celiac disease–relevant components of the microbiome that influence the expression of disease.

CROHN DISEASE

Crohn disease is a chronic idiopathic inflammation of the gut characterized by transmural involvement of the bowel wall (mucosa, muscle layer, and serosa) (see Fig. 75.1, D). Although often referred to as "terminal ileitis," the majority of patients with Crohn disease have colonic inflammation in addition or solely. Crohn disease typically runs a chronic, relapsing course often complicated by bowel obstruction as a result of fibrous strictures and by abscesses and fistulae caused by extension of inflammation beyond the bowel wall. Most Crohn patients will require surgical treatment at rates up to 70% after 20 years' disease duration. Crohn disease is treated with broad antiinflammatory drugs (corticosteroids and immunosuppressants) as well as antibodies targeting tumor necrosis factor (TNF-α), L-12/IL-23 p40, and the integrin molecules α_4 and $\alpha_4\beta_7$.

Crohn disease is thought to result from a dysregulated immune response to gut microbes. Despite evidence for genetic disease heritability, it is clear that the complex interactions of environmental exposures (including the gut microbiome and its metabolome), innate and adaptive immune dysfunction, and complex genetic and epigenetic features all contribute to disease causation and expression.[23]

Presentation

Patients with Crohn disease most often come to medical attention because of abdominal pain, altered bowel habits, and rectal bleeding. Abdominal pain may indicate bowel obstruction (especially if the pain is postprandial), an inflamed viscus, or a penetrating complication, such as an abscess or fistula. Diarrhea is related to malabsorption and dysmotility secondary to the effects of inflammatory cytokines on gut function. Diarrhea can also result from noninflammatory mechanisms like bile salt wasting or small intestinal bacterial overgrowth (SIBO). Conversely, constipation in Crohn disease can be a sign of stricturing of the bowel. Rectal bleeding results from mucosal friability and ulceration. In addition, fever, unexplained weight loss, fatigue, anemia, and failure to thrive (in children) can accompany the GI complaints or even be the primary presentation. Extraintestinal manifestations of Crohn disease include arthritis, uveitis, inflammatory skin lesions (pyoderma gangrenosum and erythema nodosum), and stomatitis. The arthritis can affect the axial (spine and pelvis) and articular skeleton, with the latter more often mirroring the activity of the gut disease. The joint complaints range from usual arthralgias to frank synovitis with swelling and tenderness (without erosive joint destruction). The uveitis most commonly occurs as episcleritis and iritis. Many of these lesions will subside with effective therapy aimed at the gut, but they can also have independent courses that require site-targeted treatment.

The incidence of Crohn disease and ulcerative colitis (UC) in Western populations has been estimated to be 5 to 15 cases/100,000 person-years.[24] Ashkenazi Jewish heritage confers increased Crohn risk in Caucasians, whereas African Americans seem to have rates similar to those of non-Jewish Caucasians, and Hispanics and Asians have much lower rates. There is a genetic risk with up to 10-fold increased rates of inflammatory bowel disease (IBD) in relatives of patients with Crohn disease and a 30% concordance rate in monozygotic twin pairs. The typical patient is diagnosed in his or her second or third decade, and there is no significant gender preference. The only environmental exposure that has been repeatedly linked to risk of Crohn disease has been tobacco use.

The majority (up to 70%) of Crohn patients experience a remitting and relapsing course, but some have chronically active symptoms refractory to medication. There are several phenotypes of disease. These include inflammatory disease (manifesting primarily as intestinal edema and ulceration), fibrostenotic disease (luminal narrowing by fibrous strictures dominate with symptoms of painful obstruction), and fistulizing disease (inflammatory tracts between the bowel and other intestines, the bladder, vagina, skin, and other structures). Although the majority of patients have inflammatory disease at the time of diagnosis, over time this phenotype changes so that after 20 years of disease duration, up to 70% and 18% of patients with Crohn disease report penetrating and fibrostenotic complications, respectively, often leading to surgery.[25]

Immune Pathophysiology

The current paradigm of Crohn disease pathogenesis is a dysregulated immune response to gut commensal microbial components (antigens and pathogen-associated molecular patterns). Initial rodent colitis models showed a predominant T-helper-cell-1 (Th1) inflammation, where the colitis could be blocked or reversed by treatment with anti–IL-12 antibodies. Roles for IL-23 and IL-17 in Crohn disease were later supported

by the IL-10–deficient mouse model of spontaneous colitis and the cell transfer model of induced colitis, where colitogenic naïve $CD4^+CD45RB^{high}$ T cells from wild-type mice are infused into T-cell–deficient mice.[26,27] IL-12 (p35/p40 dimer), IL-23 (p19/p40 dimer), IFN-γ, and IL-17 are significantly elevated in Crohn disease, but it is not known which cytokines play a more important role in specific patients. The first and most widely used biological in IBD targets TNF-α, which itself is a more downstream cytokine in the inflammatory cascade. The anti–IFN-γ monoclonal antibody (mAb) fontolizumab showed no clinical efficacy in Crohn disease (but did decrease C-reactive protein [CRP] level).[28] However, anti–IL-12/IL-23 p40 subunit mAbs targeting the Th1 and Th17 pathways resulted in clinical improvement that led to the use of ustekinumab for Crohn disease.[29,30] Currently, agents targeting IL-23 alone (anti-p19 antibodies) are being evaluated for efficacy in IBD. Unexpectedly, antibodies targeting IL-17A alone (secukinumab) or the IL-17AR receptor subunit (transducing IL-17A, IL-17E, and IL-17F intracellular signals) either did not show clinical improvement (and was associated with increased fungal infections) or even induced clinical deterioration (the anti–IL-17AR antibody).[31]

To date, it is estimated from multiple genome polymorphism studies that more than 200 IBD susceptibility loci exist, most associated with risk of both Crohn disease and UC.[32] However, it is estimated that all the loci together could account for no more than 15% of overall IBD risk. While most of these genetic risk loci are in noncoding regions of genes thought to modulate gene expression, the actual genes implicated have roles in the immune response, cell trafficking, and epithelial integrity, providing biological plausibility for their involvement in IBD. Several examples stand out. *(i)* The *NOD2* gene encodes an intracellular protein that binds muramyl dipeptide (MDP), a component of the bacterial cell wall TLR2 ligand peptidoglycan. Disease-associated mutations in the MDP-sensing leucine-rich repeat domain of *NOD2* are associated with defective activation of nuclear factor (NF)-κB. *(ii)* The *ATG16L1* autophagy gene is important for the metabolism of autologous cell proteins as well as intracellular microbes. Expression of the Crohn disease–associated Thr300Ala polymorphic *ATG16L1* sequence in a colon cancer cell line showed in vitro inhibition of packaging of intracellular *Salmonella* into autophagosomes, supporting the hypothesis that this mutation could lead to impaired clearance of microbes and chronic inflammation.[33] *(iii)* A polymorphism in the coding region of the IL-23 receptor gene (Arg381Gln) found in 14% of healthy controls is associated with protection from Crohn disease (less so UC) and is associated with decreased Th17 cascade effector cells.[34]

The key role of the commensal microbiota in gut immune homeostasis was established by the observations that germ-free mouse models are largely protected from experimental IBD. A number of studies show a dysbiosis compared with the healthy gut microbiota.[35] This change in the microbiota in IBD has been characterized as a loss of microbial community diversity, reversal of the Bacteroidetes-to-Firmicutes phyla ratio with reduction in Firmicutes, and metabolic and community changes associated with activity of disease.[36] The roles of antibiotics, probiotics, prebiotics/diet, or actively changing the microbiota through fecal transplant have not been clarified sufficiently to employ these strategies as conventional treatments in Crohn disease.

Diagnosis

The diagnosis of Crohn disease relies on radiographic, endoscopic, and histological examinations. In general, a combination of colonoscopy (with ileal examination) and small-bowel imaging (barium small-bowel follow-through, computed tomography [CT] or magnetic resonance enterography, or capsule video endoscopy) is usually sufficient to demonstrate active inflammatory disease of the colon and small bowel. Observed mucosal ulceration and friability in a patchy distribution separated by unaffected mucosa ("skip areas") are the endoscopic hallmarks of Crohn disease. Diagnostic imaging showing patchy bowel wall thickening, mucosal hyperemia, luminal narrowing (stricturing), and penetrating complications of the bowel wall, such as fistulae and abscesses, all suggest Crohn disease. Histologically, although the appearance of noncaseating granulomata is highly supportive of a diagnosis of Crohn disease, in practice, they are not often detected by endoscopic biopsy, particularly in adults. More often, evidence of chronic inflammation, such as architectural crypt distortion and basal lymphoplasmacytosis, help to differentiate the inflammation from an acute, self-limiting colitis or enteritis. Other findings, such as fecal leukocytes or elevated fecal calprotectin, may indicate an inflammatory colitis but are not specific for diagnosis of a chronic idiopathic IBD, such as Crohn disease or UC. In the setting of suggestive imaging or endoscopy appearances of colitis, the measurement of certain serum antibodies can further support the diagnosis of Crohn disease and even help differentiate it from UC, but they should not be used by themselves as diagnostic tests.[37] It has been shown that up to 68% of patients with Crohn disease are seropositive for antibodies targeting microbial antigens, such as anti–*Saccharomyces cerevisiae* antibody (up to 16% of patients with UC are seropositive). Additional antimicrobial antibodies, such as anti-OmpC, anti-I2, anti-flagellin 3, X, and CBir, also develop in Crohn disease. Crohn disease may need to be differentiated from other similarly presenting conditions, including UC, chronic ischemic colitis, infectious enteritis/colitis (amebiasis, *Yersinia enterocolitica* infection, *Mycobacterium tuberculosis* infection), intestinal lymphoma, celiac disease, diverticula-associated colitis, and radiation- and nonsteroidal antiinflammatory drug (NSAID)–induced enteropathy.

Treatment

The treatment of Crohn disease includes medical and surgical approaches. Since there is no cure for Crohn disease, the principles of therapy are to first make sure that symptoms are secondary to the underlying idiopathic inflammation and not caused by infectious or noninflammatory factors, such as coexisting irritable bowel syndrome (IBS). The goal is to induce quick remission of symptoms and establish therapy to maintain the remission with aggressive initial therapy commensurate with the extent and activity of the disease. Current management continues to evolve as new drugs, patient disease risk assessment, and treatment goals are introduced into therapeutic decision making.[38] For instance, limited anatomic extent of gut disease with little clinical impact could be treated with episodic corticosteroids and mesalamine (for colonic disease); however, for more advanced disease with more symptoms, inflammatory burden and even complications, initial therapy with biological medicines such as anti-TNF-α drugs (with or without immunosuppressants like thiopurines or methotrexate depending of disease severity) could be used; for patients with contraindications, intolerances,

prior anti-TNF-α exposure, or risk mitigation considerations, then anti-p40 or anti-integrin ($\alpha_4\beta_7$) agents could be used. Typically, maintenance therapy with the same agent used to achieve remission is needed. Continuous assessment of response (symptoms, biomarkers, mucosal healing) and use of therapeutic drug monitoring can optimize successful long-term treatment outcomes.

Surgery is required in cases of complications, such as bleeding, pain/obstruction, and fistulae that are refractory to medical therapy. Surgery typically involves resection limited to inflamed segments of small intestine and colon but small strictures can be treated in situ by stricturoplasty. In addition, surgery is required for treatment of intestinal adenocarcinoma, which also complicates the chronic inflammation of the bowel. There is a high incidence of endoscopic and symptomatic recurrence of inflammation by 2 years after surgery, and anti-TNF-α agents can successfully prevent this. Strategies to identify patients who would benefit from early postoperative drug prophylaxis are being developed.[39]

KEY CONCEPTS

Crohn Disease

- Crohn disease affects the full thickness of the bowel wall, resulting in fistulae and abscesses in 70% of patients by 20 years' disease duration.
- Crohn disease is a chronic, relapsing inflammation of the bowel, and up to 80% of patients will require surgical treatment at some point.
- Over 200 genetic loci are associated with disease risk, but several coding region polymorphisms, including *NOD2* and *ATG16L1*, highlight a role for dysfunctional innate immune responses.

ULCERATIVE COLITIS

UC also involves chronic idiopathic inflammation of the gut, but this is limited to the mucosal layer of the colon (see Fig. 75.1, E). UC can involve the rectum alone (ulcerative proctitis), the distal transverse colon to rectum (left-sided colitis), or the entire colon (pancolitis). Without the transmural inflammation of Crohn disease, penetrating complications, such as fistulae and abscesses, generally don't complicate UC. Oral and topical (per rectum) mesalamine preparations are most often used to treat UC, although patients also may require corticosteroids, immunosuppressants, and biological agents to induce and maintain remission. Unlike Crohn disease, total colectomy eliminates the disease. However, there can be ongoing complications of the surgery such as ostomy dysfunction or pouchitis if an ileal pouch–anal anastomosis is performed. Over 40% of patients with UC require surgery to treat medically refractory symptoms or development of dysplasia over their lifetime.[40]

Presentation

UC has rectal inflammation (proctitis) as its primary location but can extend to involve the whole colon. Bloody loose stool (including nocturnal) and abdominal pain are common symptoms along with proctitis-specific complaints of rectal urgency and incomplete evacuation. As in Crohn disease, fever, unexplained weight loss, fatigue, and anemia can accompany the GI complaints or be the primary presentation. Extraintestinal manifestations may include arthritis, uveitis, inflammatory skin lesions, and stomatitis. An interesting genetic connection exists between UC and HLA-B27–positive spondyloarthropathy (see Chapter 58), with 60% of patients with ankylosing spondylitis showing inflammation on colonoscopy. UC is also closely associated with primary sclerosing cholangitis (PSC; see Chapter 76); up to 3% of UC patients develop PSC, and up to 75% of all patients with PSC have UC.[41]

UC incidence is up to 20 to 25 cases per 100,000 person-years; the diagnosis of UC is typically made in the second or the third decade without significant gender preference.[42] Approximately 6% to 15% of first-degree relatives of patients with UC have a history of UC, but in general, the genetic contribution to risk is lower than in Crohn disease. There is a higher incidence of UC in European and North American populations compared with that in Asian and Latin American countries, but newly industrialized countries are increasing in incidence.[24] The only environmental exposures linked to risk of UC are the protective effect of tobacco exposure and appendectomy in the first decade of life.

The natural history of UC shows that most patients experience a remitting and relapsing course (60%), although some have prolonged remission after one episode of disease (20%), and others have chronically active symptoms refractory to medical remission (20%). Chronic inflammatory UC (and Crohn colitis) is accompanied by an increased incidence of colorectal cancer, so much so that recurring colonoscopic surveillance for dysplasia with biopsy is recommended starting 8 to 10 years after diagnosis.[43] Total colectomy is performed to treat refractory symptoms or in certain instances of detected dysplasia.

Immune Pathophysiology

UC was originally characterized as a Th2-like disease because of the increased IL-5 and IL-13 production in inflamed gut tissue seen in an animal model of UC, oxazolone-induced colitis, as well as from patient specimens.[44] In this model, mucosal natural killer T (NKT) cells produced the excess IL-13 and the colitis was reversed by immunoneutralizing anti–IL-13. When translated to humans, patients with UC were found to have high capacity for IL-13 production, also by type II NKT cells. IL-13 is a biologically plausible effector cytokine in UC injury because it disrupts the epithelial tight junction by upregulating claudin-2 and has a direct toxic effect on human gut epithelial cells in vitro.[45] However, results from a clinical trial using an anti–IL-13 antibody in UC did not show significant treatment efficacy.[46]

Data from genetic susceptibility studies in UC have provided less-compelling examples of disease-specific mechanisms compared with those in Crohn disease, but there are associations with HLA class II genes distinct from Crohn disease and loci associated with genes involved in epithelial barrier function, such as *GNA12* for tight junction formation, *CDH* for epithelial cadherin-1, *LAMB1* for the laminin constituent of basement membranes, and the ubiquitin ligase gene *RNF186*.[47,48] When specifically assaying epithelial cells (vs. whole tissue) for epigenetic markers associated with genomic risk loci, there is an enhanced association supporting the epithelium as a primary target for dysfunction in UC.[49,50]

Diagnosis

The diagnosis of UC is made by colonoscopy to show the confluent extent of the inflammation from the rectum to points proximal (ileal intubation can confirm that the inflammation is limited to the colon) and biopsy, which should contain histological features of chronic inflammation, including crypt distortion, crypt dropout, and lymphoplasmacytosis. The presence

of acute inflammatory features alone (polymorphonuclear cells, crypt abscess, and cryptitis) may also be seen but, when in isolation, these features suggest other etiologies, such as acute infectious, drug-induced, ischemia, and toxic exposures. Although no blood test can be used to diagnose UC, the presence of high-titer perinuclear antineutrophil cytoplasmic antibodies (pANCAs) can be seen in up to 70% of patients with UC; this information can help differentiate chronic indeterminate colitis when coupled with anti-*Saccharomyces cerevisiae* mannan antibodies (ASCAs; see above). At all times, acute infections from enteric pathogens, including *Clostridium difficile*, should also be excluded, as these may also occur during active IBD treatment and mimic exacerbation of disease (evaluation for cytomegalovirus [CMV] infection should be performed in the setting of UC seemingly refractory to immunosuppressants). Once a diagnosis is established, elevations in transaminases or alkaline phosphatase should prompt an evaluation for PSC (imaging by magnetic resonance or endoscopic cholangiopancreatography).

Treatment

Treatment is tailored to the extent and activity of disease.[51] Mild to moderate proctitis can respond to topical corticosteroids or mesalamine alone (enemas and/or suppositories). Most often with more extensive colonic involvement, oral mesalamine is required, which can be useful for induction and maintenance of remission. In moderate to severe disease, corticosteroids may be required to induce rapid responses and early introduction of anti–TNF-α, anti-p40 or anti-integrin agents are used to induce fairly rapid clinical responses and remission and may also be used as maintenance therapy. Janus kinase (JAK) inhibitors may be used if there is no response to the biologicals. Severe disease may require hospitalization for high-dose intravenous steroids followed by anti-TNF or cyclosporine if no response.

Mesalamine-based drugs are a cornerstone of therapy in UC and are generally included in most ongoing UC medical regimens. Whether use of mesalamine even in quiescent disease confers chemoprotection from developing dysplasia is still being debated, but because of their generally low adverse event rate and high tolerance, long-term use is a reasonable choice.

As discussed, surgery has a definite role in treating medically refractory disease or addressing dysplasia discovered by surveillance colonoscopy (dysplasia surveillance is done every 1 to 2 years after 8 to 10 years of disease by taking four-quadrant biopsy specimens every 10 cm). After total colectomy, options include an ileal pouch–anal anastomosis or a permanent end ileostomy. However, even the pouch can become chronically inflamed; generally, antibiotic responsive, this inflammation can also become refractory and require steroid, immunosuppressant, or biological therapy and even pouch removal.

KEY CONCEPTS

Ulcerative Colitis

- Ulcerative colitis (UC) is a chronic, relapsing inflammation of the colon that is limited to the mucosa and is not transmural.
- Many medical treatments for UC (including anti-TNF-α, anti-p40 or anti-integrin biologicals) overlap with Crohn disease, but total colectomy is typically reserved for UC when surgery is required for complicated or refractory disease.
- Although less strongly associated with genetic inheritance compared with Crohn disease, several UC susceptibility loci have been identified and are associated with epithelial barrier function.

ON THE HORIZON

Inflammatory Bowel Disease

- Improvement in detection of dysplasia using molecular techniques to better understand the incidence, natural history, prevention of, and role of inflammation-induced somatic mutations in inflammatory bowel disease (IBD)-related colonic epithelial neoplasia.
- Defining biomarkers that identify genotype–phenotype linkages that predict responses to highly targeted therapies like monoclonal antibodies.
- Determining whether the dysbiosis in IBD is a cause or effect and how manipulation of the microbiome can influence inflammation and response to treatment.

OTHER IDIOPATHIC INFLAMMATORY BOWEL DISEASES

Microscopic Colitis

Microscopic colitis is an increasingly recognized clinical condition that links chronic watery diarrhea with intraepithelial lymphocytosis (lymphocytic colitis) or with increased subepithelial collagen deposition (collagenous colitis). It differs from Crohn disease and UC by lacking endoscopic mucosal damage or evidence of histological features of chronic inflammation (no architectural crypt distortion, basal lymphoplasmacytosis, or loss of goblet cells). However, it causes significant morbidity and may require chronic immunosuppression for treatment. Although its etiology is unknown, there are associations with autoimmune conditions as well as with certain drug exposures.[52]

The hallmark symptom of microscopic colitis is poorly tolerated chronic, copious, watery, nonbloody diarrhea that can cause fatigue, arthralgias, and weight loss. Microscopic colitis typically begins in the sixth and seventh decades with a female predominance and is associated with a history of autoimmunity, especially thyroid disease, rheumatoid arthritis, and CREST (calcinosis, Raynaud phenomenon, esophageal dysmotility, sclerodactyly, and telangiectasia) syndrome and even celiac disease.[53] Although celiac-like villus blunting may be seen in less than 10% of patients with microscopic colitis, celiac serologies will not be positive, indicating that classic gluten enteropathy is not playing a role.

There is no current information on the immune mechanism of microscopic colitis. Activated NF-κB, increased nitric oxide and prostaglandin production have been measured and have been linked to the diarrhea. Excessive transforming growth factor (TGF)-β has also been measured in collagenous colitis, consistent with its role in collagen production and fibrosis. Concomitant medications associated with microscopic colitis include H_2 blockers, PPIs, selective serotonin reuptake inhibitors, ticlopidine, and NSAIDs.

The diagnosis of microscopic colitis relies on histology. The sine qua non for microscopic colitis is increased numbers of intraepithelial lymphocytes (>20 lymphocytes/100 epithelial cells) on colonic mucosal biopsy. This can be accompanied by a chronic inflammatory infiltrate in the lamina propria and, less often, a limited appearance of neutrophils (especially cryptitis), the latter finding suggesting that the etiology of the microscopic colitis may actually be related to an injurious drug effect, such as NSAID exposure. In collagenous colitis, a prominent subepithelial collagen band ≥10 μm is seen in addition to the intraepithelial lymphocytosis.

The general treatment approach to the patient with microscopic colitis begins with eliminating suspect concomitant medications, such as NSAIDs. Coexisting celiac disease should be considered, where appropriate, and addressed with gluten withdrawal. Medical therapies can begin with antidiarrheals (loperamide, diphenoxylate/atropine), a trial of cholestyramine (bile salt malabsorption has been hypothesized to play a role in microscopic colitis), and even a course of mesalamine. Bismuth subsalicylate has been reported to benefit a minority of patients with microscopic colitis; in patients with initially severe symptoms or who are refractory to these first treatments, short-term corticosteroids have been very effective. In particular, limited courses of oral budesonide 9 mg taken once per day has reliably improved diarrhea and induced remission, a result supported by placebo-controlled studies.[54] However, balancing long-term relief of symptoms with side effects of corticosteroid use in relapsing or steroid-dependent disease has led to possible use of steroid-sparing immunosuppressive medications, such as azathioprine and methotrexate, for long-term maintenance. Reports of anti–TNF-α drugs for refractory microscopic colitis have been published.[55] Occasionally, patients may need colectomy and ileostomy to manage refractory symptoms or drug intolerance.

Eosinophilic Esophagitis

Eosinophilic esophagitis (EoE) is a more recently recognized disease defined by symptomatic idiopathic eosinophilic inflammation of the esophagus in the absence of other known causes, especially chronic gastroesophageal reflux disease (GERD). Although its etiopathology is unclear, there are gene associations (*FLG* for epithelial barrier effects and *eotaxin-3* and *TSLP* for immune response effects), and data from animal models and human disease have implicated central roles for loss of tolerance to food antigens involving Th2 cytokines IL-5 and IL-13. EoE is being recognized with increasing frequency against a background of increased incidence of inflammatory allergic diseases.

EoE has an incidence rate of 0.1 to 0.2 cases/10,000 persons/year, has a male predominance (up to 70%), and peak incidence in neonates to 3 year olds. Symptoms include failure to thrive and feeding difficulty in infants (e.g., food refusal, limited variety diet, prolonged feeding times) and abdominal pain and vomiting in older children and adolescents. In adults, the primary symptom is typically intermittent dysphagia, with the first presentation possibly being an acute food impaction in the esophagus. Adult patients may report GERD symptoms that do not respond to adequate acid-suppression therapy. Patients report a high rate (>50%) of atopy (rhinoconjunctivitis, wheezing, or family history of atopy) as well as food allergies (including positive skin prick test, allergen-specific IgE test, or anaphylactic response to a dietary allergen).[56] There is also an association of esophageal disease (strictures or EoE) in the parents of up to 10% of patients.

The diagnosis of EoE requires endoscopic biopsy of the esophagus since increased eosinophils in the esophageal epithelium must be measured. The endoscopic appearance of the esophagus can show multiple thin rings ("feline esophagus"), with linear longitudinal furrows and whitish papules that represent eosinophilic microabscesses at the surface of the squamous epithelium. Biopsies show an infiltrate of eosinophils in the epithelium of at least 15 to 20 eosinophils/high-power field (hpf).[57] These often concentrate just under the epithelial surface and also form microabscesses in groups of ≥4 eosinophils (systemic eosinophilia can be seen in over 70% of patients with EoE). It is important to take at least three biopsy specimens, since involvement may be variable and patchy; in addition, it is advisable to take biopsy specimens from the distal and mid-to-upper esophagus (to help differentiate changes seen in GERD that can be limited to the distal esophagus) and specimens from the gastric and duodenal mucosae (to show that the eosinophilic infiltration is limited to the esophagus and does not represent a diffuse process, such as that found in eosinophilic gastroenteritis or hypereosinophilic syndrome). However, the introduction of the EoE diagnostic panel, a set of 94 differentially expressed genes in the esophageal mucosa of EoE patients, may possibly supplant histological examination as a way to follow response to therapy and stratify risk for recurrent disease as a way to monitor treatment.[58]

In terms of GERD, it is important to make sure that any excessive acid reflux is treated and controlled; persistent symptoms (and persistent biopsy abnormalities) may prompt a 24-hour ambulatory pH study of the distal esophagus to ensure that the acid-suppression treatment results in a normal acid-contact time. In fact, there seems to be a form of EoE that, while seeming to be without excess exposure to gastric acid, does, in fact, respond well to proton pump inhibitor (PPI) treatment; this may be attributed to the non–acid-suppressive effects of a drug, such as omeprazole, that can suppress eotaxin-3 secretion from squamous mucosae.[59] Multiple types of esophageal dysmotility, often reversible with treatment, have been described in EoE, contributing more to dysphagia than to strictures.

The pathogenesis of EoE seems to be linked to environmental and food allergen hypersensitivity. Given a familial association of EoE, atopy, and food allergy, a genetic component may be contributing to disease susceptibility. Several candidate disease susceptibility gene/gene loci in EoE include the 3′ untranslated region of eotaxin (*CCL26*), the TGF-β_1 promoter, a filaggrin (*FLG*) exon, and a thymic stromal lymphopoietin (*TSLP*) intron and TSLP receptor (*CRLF2*) exon. These associations are biologically plausible, since eotaxin is excessively expressed in EoE mucosa, filaggrin is a structural skin protein that helps maintain barrier function (and is downregulated by IL-13), and TSLP has been shown to stimulate IL-13 production by innate helper cells in the lamina propria. Moreover, the inflammation in EoE is characterized by increased IL-13 and IL-5 production; animal models of aeroallergen induction of an EoE-like lesion is blocked in both IL-13–deficient and signal transducer and activator of transcription 6 (STAT6)–deficient (an intracellular molecule important for IL-13 receptor α_1 signaling) animals. These data suggest that IL-13 secretion induces production of eotaxin from epithelial cells, which, together with IL-5, drives the local eosinophilic infiltration. Finally, the association with food allergy has led to successful therapy of EoE by using strict elimination diets (sometimes informed by skin testing) or even the use of elemental diet tube feedings.

Given the strong association with food allergies, elimination of foods commonly associated with IgE-mediated responses may be restricted first. Lack of improvement in symptoms would lead to a trial of amino acid–based elemental liquid diet necessitating nasogastric (or later percutaneous gastrostomy) placement, but despite its high efficacy, this approach can be uncomfortable, impractical, and expensive. If this dietary approach is successful, then after several weeks, individual foods may be added back every 5 to 7 days. For patients not responding to dietary therapy or with no identifiable dietary

allergens, corticosteroid treatment has been used successfully. Both systemic oral and swallowed topical corticosteroids (e.g., fluticasone propionate metered dose inhalers) for 4 to 6 weeks have been shown to relieve symptoms and resolve histological inflammation; one or the other may be more or less effective based on body weight, dose, steroid resistance, or severity of the EoE. However, relapse rates are high over the year following a course of steroids, which might suggest a trial of azathioprine or 6-mercaptopurine. Lastly, endoscopic therapy to treat strictures using dilation may incur higher risk of mucosal tears so that conservative treatment (smaller dilators, assessment for tears during the procedure before proceeding further) is encouraged.

GASTROINTESTINAL COMPLICATIONS OF PRIMARY IMMUNODEFICIENCIES

Certain primary immunodeficiency diseases increase the risk for GI complications. These complications fall into three main categories: infectious, idiopathic inflammatory/autoimmune, and neoplastic. CVID and chronic granulomatous disease (CGD) have some of the most frequent and significant gut manifestations, and these are discussed in detail below. However, a variety of inherited conditions of broad lymphocyte or innate immune cell dysfunction can also lead to gut disease. The study of monogenic diseases that are associated with inflammatory diseases of the gut have come to the attention of investigators, as they may be models for the mechanisms that could contribute to the pathophysiology of more genetically complex conditions, such as Crohn disease and UC.[60]

Common Variable Immunodeficiency

CVID is a syndrome of hypogammaglobulinemia (low levels of IgG and IgA and/or IgM) accompanied by recurrent sinopulmonary infections (see Chapter 33). Patients have no isohemagglutinins and cannot mount an adequate antibody response to T cell-dependent and -independent vaccine antigens. In general, Ig replacement therapy improves the sinopulmonary infection rate but does not affect other complications, such as autoimmune and gastrointestinal. Case series of patients with CVID have shown that up to 60% of patients experience GI symptoms and that endoscopic and histological abnormalities can occur in the majority of patients with CVID.[61] When considering the differential diagnostic possibilities for the GI complications of CVID, it is helpful to separate them into infectious, immune-mediated, and neoplastic processes. Among the infectious agents, *Giardia lamblia*, nontyphoidal *Salmonella*, and *Campylobacter jejuni* are frequently seen, but *Cryptosporidium* may also be found, and *C. difficile* and viral agents (CMV) may be encountered. In addition to the higher rate of gastric *H. pylori* infection, small-intestinal bacterial overgrowth is present in up to 30% of patients with CVID. The immune-mediated GI complications of CVID include idiopathic enteropathy (villous atrophy, increased intraepithelial lymphocytes/microscopic colitis, nodular lymphoid hyperplasia) seen in less than 15% of patients (see Fig. 75.1, F), manifesting as severe malabsorption (see below) and, even more rare, macroscopic ulcerating disease resembling UC or Crohn disease. CVID-associated autoimmune disease involving the GI tract also includes type II gastritis that can lead to achlorhydria and vitamin B_{12} deficiency ("pernicious anemia") and even autoimmune hepatitis and primary biliary cirrhosis. Increased rates of intestinal lymphoma and gastric adenocarcinoma (related to achlorhydric autoimmune gastritis) have been found.[62]

Presentation

The most common symptom of GI complications in CVID is episodic diarrhea that can progress to chronic diarrhea, regardless of the etiology. Weight loss and evidence of vitamin or mineral deficiencies may occur, with underlying maldigestion or malabsorption. In addition, abdominal pain related to splenomegaly (with or without portal hypertension) or ascites (portal hypertension secondary to hepatic nodular regenerative hyperplasia) can occur. Fever, together with intestinal obstruction and GI bleeding, may indicate the development of small-bowel lymphoma. Patients with CVID may report increased frequency of GERD symptoms as well as dyspepsia, but these complaints are nonspecific and have been difficult to link with immune defects or autoimmune complications.

Immune Pathophysiology

It is not clear if lack of Igs alone causes susceptibility to gut infections and inflammation because patients with X-linked agammaglobulinemia infrequently develop overt GI disease (they lack B cells but have functionally intact T-cell populations). Even persons with selective IgA deficiency have little in the way of GI disease, including infections, though CVID enteropathy risk may be higher in patients with very low tissue IgA levels.[63] The T-cell dysfunction that often accompanies CVID likely contributes substantially to susceptibility to GI disease.

In terms of an increased GI infection risk, autoimmune gastritis-induced achlorhydria could increase small-bowel exposure to swallowed commensals and pathogens escaping this innate gastric barrier to infection. Mucosal nodular lymphoid hyperplasia, characterized by disorganized secondary lymphoid nodules with poorly formed germinal centers, is likely related to the inability of B cells to undergo class switching when presented with antigen in situ. Last, CVID enteropathy is characterized clinically by severe malabsorption and histologically by blunted villi and increased intraepithelial lymphocytes, and epithelial apoptosis is associated with excess Th1 cytokine secretion (IL-12 and IFN-γ).[64] It remains unknown why this inflammatory lesion occurs in the first place, but a recent report implicates chronic norovirus infection as a possible etiology.[65]

Diagnosis

The work-up of GI complications of CVID often takes place as symptoms are usually first episodic but may eventually turn into chronic, progressive complaints. The key to successful diagnosis is the initial search for treatable infectious causes. This requires stool assay and culture for bacterial (especially *Campylobacter*) and protozoal pathogens as well as stool polymerase chain reaction (PCR) for norovirus RNA. In addition, hydrogen breath testing can help detect SIBO to provide another antibiotic-responsive etiology of the symptoms.

Negative results for infectious causes require endoscopy for biopsy of the proximal small intestine and colon. Routine histology will detect the features of enteropathy, including blunted villi and increased IELs, which are often interpreted as celiac disease. Unlike celiac disease, there usually is a lack of plasma cell infiltrate and crypt hyperplasia, as well as a preserved brush border and goblet cells, and there is often an increase in epithelial cell apoptosis, particularly in the colon.[66] If the diagnosis of

celiac disease must be considered, then genetic testing for celiac susceptibility HLA alleles should be performed; the absence of combinations of A and B alleles for HLA-DQ2 or -DQ8 that confer risk will rule this out. However, even if celiac disease gene alleles are detected, it still does not indicate that this is a gluten-driven celiac disease lesion but will require a trial of GFD nonetheless. The functional significance of any histological lesion in the small intestine can be evaluated by a measure of steatorrhea (fecal fat excretion) and small-bowel absorption (D-xylose absorption test), values that can be used to track improvement following treatments.

Treatment

None of the GI complications of CVID is treated by IgG replacement (initial treatment might alleviate GI symptoms, it does not treat overt gut inflammation and ulcers). Patients with any bacterial or protozoal pathogens should be treated with a recommended course of conventional antibiotics, including those positive for the *C. difficile* toxin. SIBO should be treated, and recurrent SIBO may need cycling antibiotic regimens. Norovirus infection responds unreliably to ribavirin.

The treatment of the idiopathic enteropathy is very challenging. Although this seems to be a late complication in a subset of patients, it has a high mortality.[67] In the early stages, it may be responsive to a short course of oral corticosteroids, either prednisone or budesonide. Case reports attest to the efficacy of infliximab and anti-p40 antibodies are being tested in clinical trials. It is possible that immunosuppressants may be used to control the inflammatory response underlying the small-bowel mucosal damage, but this should only be done in a closely observed clinical setting, with monitoring for infections. At all times, the patient's nutritional status should be maintained, initially using the oral route but administering parenteral nutrition to complement oral nutrition and when oral feedings are not adequately absorbed and exacerbating the diarrhea.

Portal hypertension may complicate CVID, typically from the nodular regenerative hyperplasia of the liver though little fibrosis occurs. It needs to be clarified in individual patients with splenomegaly (≈20%) that portal hypertension is not caused by excessive splenic vein flow associated with splenomegaly (induced by antibody-mediated autoimmune cytopenias) that might be ameliorated by splenectomy. In any scenario, the management of such late complications requires an especially experienced team of internists, surgeons, and nutritionists.

KEY CONCEPTS

Common Variable Immunodeficiency

- The majority of gastrointestinal (GI) complications of common variable immunodeficiency (CVID) are infectious and generally do not respond to intravenous or subcutaneous immunoglobulin (IV/SCIG) therapy (compared with sinopulmonary suppurative infections).
- CVID enteropathy is a rare immune-mediated complication of CVID that also does not respond to IV/SCIG.
- CVID enteropathy is often confused with celiac disease because of similar villus damage on biopsy, but additional features (lack of plasma cells, increased epithelial apoptosis, absence of celiac gene risk alleles) can help differentiate them.
- CVID enteropathy has no established therapy though judicious use of short courses of oral steroids or conventional immunosuppression may relieve the malabsorption and diarrhea temporarily.

Chronic Granulomatous Disease

CGD results from defects in the nicotinamide adenine dinucleotide phosphate (NADPH) oxidase complex that impair the ability of phagocytic cells to produce the reactive oxygen species required to kill bacteria and fungi within intracellular phagolysosomes (see Chapter 39). Patients with CGD have recurrent infections of the skin, lungs, liver, and bone, and nearly half develop gut inflammatory complications affecting areas anywhere from the mouth to the anus. Interestingly, the frequency of GI disease in CGD is higher in the X-linked $gp91^{phox}$ defect, but a recently described $p40^{phox}$ defect occurred in a young male patient who presented with granulomatous colitis alone.[68]

Presentation

GI symptoms (abdominal pain and diarrhea, with or without rectal bleeding) usually begin in the first decade of life, sometimes preceding the diagnosis of CGD. Although infectious diarrhea (especially *Salmonella* spp. and *C. difficile*) occurs in CGD, an idiopathic CGD-associated IBD also develops. In the mouth, granulomatous stomatitis and dental abscesses cause pain and difficulty eating; in the esophagus, dysphagia, chest pain, and vomiting may result from luminal narrowing or dysmotility related to granulomatous inflammation and fibrosis; in the stomach, loss of motility caused by thickened walls and smaller volume leads to vomiting, epigastric pain, and weight loss as a result of feeding difficulty; and in the small and large intestines, diarrhea (including protein-losing enteropathy), bowel obstruction (large granulomata compromising the size of the lumen), rectal bleeding, and tenesmus may result from active colitis/enteritis with mucosal ulceration, anal fissures, and perianal abscesses. In addition, because of the transmural nature of the granulomatous inflammation, penetrating complications, such as fistulae and abscesses, can occur. Feeding difficulties and the chronic inflammatory state itself predispose to growth delay that often affects pediatric patients with CGD.

Hepatic abscesses represent another frequent complication in CGD, occurring in up to 45% of patients.[69] These patients present with fever as well as abdominal pain, fatigue, and, less often, abdominal tenderness and hepatomegaly. The erythrocyte sedimentation rate (ESR) and the alkaline phosphatase level are elevated in half the affected individuals. However, a high level of suspicion, especially in the setting of fever with or without abdominal pain, should instigate a search for hepatic abscesses.

Immune Pathophysiology

Given the defects in ability to kill intracellular bacteria and fungi, pathogens, and possibly commensals alike, it is thought that the excessive granulomatous response is caused by delayed antigenic clearance or persistent infection. In this way, granulomata continue to multiply and grow while other inflammatory pathways that normally deal with the microbes or are induced by cytokines are activated. The end results of granulomatous inflammation are most evident in tissues rich in macrophages and reticuloendothelial cells, such as the gut lamina propria, liver, lymph nodes, and spleen.

Diagnosis

The symptoms and signs will dictate the initial diagnostic examinations. For diarrheal complaints, stool culture and examination for *C. difficile* toxin are required; in the setting of hypoalbuminemia, fractional fecal excretion of α_1-antitrypsin can detect

protein-losing enteropathy (>50 mg/24 hour) as a result of either diffuse mucosal inflammation or lymphangiectasia. For complaints of dysphagia, vomiting, or epigastric pain, upper endoscopy can help document macroscopic and microscopic involvement with granulomatous inflammation. Radiological studies using oral contrast can detect a narrowed lumen, stricturing, and motility and mucosal abnormalities of the esophagus and stomach but cannot provide histological confirmation. However, radiological imaging studies may be the primary diagnostic tools to evaluate obstructive symptoms from the small intestine, including barium small-bowel studies and CT or magnetic resonance enterography. CT and MR can show penetrating complications such as fistulization. Finally, to evaluate lower abdominal and perianal pain and rectal bleeding, anoscopy, colonoscopy, and pelvic CT or magnetic resonance imaging (MRI) will help diagnose granulomatous inflammation of the colon and its complications, including perirectal and perianal abscesses. In the case of hepatic abscess detection (generally 1 to 6 cm), CT, MRI, and ultrasonography have similar sensitivity (≈60%). Active abscesses appear to be solid, hypoechoic lesions on ultrasonography or postcontrast ring–enhancing lesions on CT and MRI.

The GI histological diagnosis hinges on the presence of noncaseating granulomata, both gross and microscopic. These granulomata are often seen against a background of acute inflammation (acute focal colitis, crypt abscesses, cryptitis) as well as chronic inflammation (lymphocytic infiltrate, Paneth cell metaplasia in the colon) that can be mild to severe. The histological picture may resemble Crohn disease except that the granulomata of CGD are well defined, often large collections of epithelioid histiocytes that can expand the mucosa (and even deform the overlying epithelium to make it look flattened in the case of the villous mucosa of the small intestine). Like Crohn disease, the inflammation can affect the three layers of the gut wall, but unlike in Crohn disease, CGD biopsy specimens also show prominent lipid-laden macrophages that have periodic acid-Schiff (PAS)–positive cytoplasmic granules.

Treatment

Current clinical practice for CGD includes using prophylactic antimicrobials to prevent infections, typically trimethoprim-sulfamethoxazole for bacterial and itraconazole for fungal infections. Some clinicians also use IFN-γ (subcutaneous, three-times-weekly dosing) to prevent infections, although this is not a universal practice. Obviously, discovery of infectious etiologies of diarrhea should be treated appropriately.

Once infections are ruled out and granulomatous inflammation of the GI tract is established, with or without complications, such as stricturing of the bowel, treatment with corticosteroids is indicated; beginning doses up to 1 mg/kg/day tapering over 12 to 20 weeks to maintenance doses of 2.5 to 5 mg every other day has been reported to induce rapid alleviation of symptoms. Use of sulfasalazine for colitis may have limited benefit in some patients. Isolated reports of successful use of cyclosporine and infliximab have indicated that these agents are best reserved for refractory cases because of the potential for infectious side effects. Similarly both granulocyte colony-stimulating factor (G-CSF) and granulocyte–macrophage colony-stimulating factor (GM-CSF) have been used with therapeutic benefit in GI complications of CGD because of their success in the granulomatous colitis of glycogen-1-β storage disease.[70] Although there are no data to suggest that IFN-γ worsens established disease, it also may not prevent it as greater than 40% of one cohort developed GI manifestations of granulomatous inflammation after starting it; in contrast, isolated reports have attributed alleviation of GI inflammation to IFN-γ. IL-1 receptor blockade, based on preclinical efficacy data in CGD, may offer limited success.[71]

Finally, it cannot be overstated that surgical drainage of complicating abscesses and resection of fibrotic or refractory strictures needs to be pursued, when indicated. Although there can be considerable postoperative complications because of ongoing fistula formation and wound breakdown, these problems may be managed by administration of corticosteroids. In addition, judicious use of endoscopic therapy to dilate narrowed esophageal or pyloric regions is an option for symptomatic strictures.

Treatment of hepatic abscesses has been shown to be amenable to nonsurgical management by adding corticosteroids to the antibiotic regimen. Empiric antibiotic coverage (at least to cover the most frequent isolate, *Staphylococcus aureus*) is generally appropriate although percutaneous aspiration of hepatic abscesses may be needed as well.[72]

KEY CONCEPTS

Chronic Granulomatous Disease

- Half of patients with chronic granulomatous disease (CGD) develop gastrointestinal (GI) involvement with granulomatous inflammation.
- Symptoms are caused by both obstructive complications of stricture formation and mucosal inflammation.
- A resemblance to Crohn disease is suggested by granulomatous inflammation, transmural bowel involvement, and underlying innate immune defect, but unlike in Crohn disease, there are also prominent skin and lung infections and periodic acid-Schiff–positive lipid-laden macrophages in the lamina propria.
- Complicated hepatic abscesses require collaborative medical and surgical intervention.

Gastrointestinal Complications Occurring in Other Primary Immunodeficiency States

Severe combined immunodeficiency (SCID) covers a wide phenotype of low-to-absent T cells, NK cells, and dysfunctional B cells reflecting the mechanisms of the genetic defect (see Chapter 32). Recurrent infectious diarrhea and thrush are typical GI conditions before a diagnosis of SCID is made in newborns and neonates. After allogeneic hematopoietic stem cell transplantation (HSCT), GI graft-versus-host disease (GvHD) is also a possibility. For many of the monogenic immunodeficiency diseases that confer susceptibility to GI inflammation, the success of HSCT to correct the secondary effect on the GI tract will depend entirely on whether the gene defect affects the myeloid or lymphoid cells primarily. If the gut stromal or epithelial cells also depend on normal function of the defective gene, then bone marrow transplantation will not likely relieve the GI effects.

The rare X-linked recessive Wiskott-Aldrich syndrome (*WAS*) (see Chapter 34) results from *WAS* gene mutations, a signaling protein largely restricted to hematopoietic cells, and leads to a syndrome of eczema, thrombocytopenia, and infections related to combined immunodeficiency (impaired antibody responses and T-cell function). A vexing aspect of the gut complication in this immunodeficiency is that patients may develop a noninfectious colitis resembling UC (confluent mucosal inflammation with ulceration, crypt abscesses, and no granulomata) with the bleeding exacerbated

by the thrombocytopenia.[73] Patients may respond to mesalamine drugs but use of steroids and immunosuppressives must be done cautiously because of increasing the infection risk. Successful bone marrow transplantation can treat the colitis as well.

Very early-onset IBD is a severe intestinal inflammation from rare mutations that affect IL-10 or its receptor. Patients present with colitis within weeks of birth that has features of Crohn disease. Diagnosis is made by genetic testing as well as by assay of peripheral blood mononuclear cells for absence of IL-10 production or absent IL-10 signaling demonstrated by decreased/absent induction of phospho-STAT3 following the addition of IL-10. X-linked lymphoproliferative syndrome 2 (*XIAP* deficiency) can present with very early-onset fistulizing perianal disease in up to 20% of patients.[74] The only cure is with HSCT.

The hyper-IgM syndrome (type 1) (see Chapter 33) is an X-linked condition resulting from mutations in CD40 ligand leading to defective antibody class switching (hence high [or normal] IgM with low IgA and IgG) and NK- and T-cell cytotoxicity. *Cryptosporidium* infectious diarrhea is seen, and sclerosing cholangitis, cirrhosis, and neoplasms of the hepatobiliary system complicate the course in teenagers and young adults. Rarely, noninfectious colitis has been reported. HSCT, preferably done before chronic complications occur, is the only approach with a potential for cure. IPEX (immune dysfunction/polyendocrinopathy/enteropathy/X-linked) syndrome is interesting in that it results from a defect in FOXP3 expression and function, thereby inhibiting the generation of Tregs (see Chapter 13). There is a lack of immune regulation, and enhanced responses, including autoimmunity, occur. The gut mucosa is a major site affected by inflammation, and this complication presents with a watery, sometimes bloody, diarrhea and malabsorption. Biopsies show villus blunting and atrophy in the small bowel and increased intraepithelial lymphocytes, lamina propria infiltration with T cells, eosinophils, and neutrophils elsewhere in the gut. Treatment requires HSCT, but the gut disease may be managed temporarily with corticosteroids and immunosuppressants.

Of note, colitis is a major side effect of the so-called checkpoint inhibitors that remove the regulation of the immune response in order to foster an anti-tumor T-cell response. Antibodies to cytotoxic T-lymphocyte antigen-4 (CTLA-4) and programmed cell death 1 (PD-1) can induce a colitis that can have penetrating ulcers; this unexpected side effect points to the importance of gut mucosal regulatory mechanisms for maintaining immune homeostasis.

ON THE HORIZON

Primary Immunodeficiency Diseases

Translation of monogenic immunodeficiency disease mechanisms, especially in very-early-onset inflammatory bowel disease (IBD), into knowledge of pathophysiology (genetic and acquired) of Crohn disease and ulcerative colitis for novel drug development.

REFERENCES

1. Neumann WL, Coss E, Rugge M, Genta RM. Autoimmune atrophic gastritis—pathogenesis, pathology and management. *Nat Rev Gastroenterol Hepatol*. 2013;10(9):529–541.
2. Lahner E, Norman GL, Severi C, et al. Reassessment of intrinsic factor and parietal cell autoantibodies in atrophic gastritis with respect to cobalamin deficiency. *Am J Gastroenterol*. 2009;104(8):2071–2079.
3. Nguyen TL, Dipaolo RJ. A new mouse model of inflammation and gastric cancer. *Oncoimmunology*. 2013;2(10):e25911.
4. Lahner E, Carabotti M, Annibale B. Treatment of *Helicobacter pylori* infection in atrophic gastritis. *World J Gastroenterol*. 2018;24(22):2373–2380.
5. Hall SN, Appelman HD. Autoimmune gastritis. *Arch Pathol Lab Med*. 2019;143(11):1327–1331.
6. Massironi S, Zilli A, Elvevi A, Invernizzi P. The changing face of chronic autoimmune atrophic gastritis: an updated comprehensive perspective. *Autoimmun Rev*. 2019;18(3):215–222.
7. Pimentel-Nunes P, Libanio D, Marcos-Pinto R, et al. Management of epithelial precancerous conditions and lesions in the stomach (MAPS II): European Society of Gastrointestinal Endoscopy (ESGE), European Helicobacter and Microbiota Study Group (EHMSG), European Society of Pathology (ESP), and Sociedade Portuguesa de Endoscopia Digestiva (SPED) guideline update 2019. *Endoscopy*. 2019;51(4):365–388.
8. Egan BJ, Holmes K, O'Connor HJ, O'Morain CA. *Helicobacter pylori* gastritis, the unifying concept for gastric diseases. *Helicobacter*. 2007;12(suppl 2):39–44.
9. Sepulveda AR, Patil M. Practical approach to the pathologic diagnosis of gastritis. *Arch Pathol Lab Med*. 2008;132(10):1586–1593.
10. Das S, Sarkar A, Ryan KA, et al. Brain angiogenesis inhibitor 1 is expressed by gastric phagocytes during infection with *Helicobacter pylori* and mediates the recognition and engulfment of human apoptotic gastric epithelial cells. *FASEB J*. 2014;28(5):2214–2224.
11. Kivrak Salim D, Sahin M, Koksoy S, et al. Local immune response in *Helicobacter pylori* infection. *Medicine (Baltimore)*. 2016;95(20):e3713.
12. Kronsteiner B, Bassaganya-Riera J, Philipson C, et al. Systems-wide analyses of mucosal immune responses to *Helicobacter pylori* at the interface between pathogenicity and symbiosis. *Gut Microbes*. 2016;7(1):3–21.
13. Kamboj AK, Cotter TG, Oxentenko AS. *Helicobacter pylori*: The past, present, and future in management. *Mayo Clin Proc*. 2017;92(4):599–604.
14. Neumeister P, Troppan K, Raderer M. Management of gastric mucosa-associated lymphoid tissue lymphoma. *Dig Dis*. 2015;33(1):11–18.
15. Caio G, Volta U, Sapone A, et al. Celiac disease: a comprehensive current review. *BMC Med*. 2019;17(1):142.
16. Christophersen A, Risnes LF, Dahal-Koirala S, Sollid LM. Therapeutic and diagnostic implications of T cell scarring in celiac disease and beyond. *Trends Mol Med*. 2019;25(10):836–852.
17. Malamut G, Verkarre V, Suarez F, et al. The enteropathy associated with common variable immunodeficiency: the delineated frontiers with celiac disease. *Am J Gastroenterol*. 2010;105(10):2262–2275.
18. Caminero A, Galipeau HJ, McCarville JL, et al. Duodenal bacteria from patients with celiac disease and healthy subjects distinctly affect gluten breakdown and immunogenicity. *Gastroenterology*. 2016;151(4):670–683.
19. Valitutti F, Cucchiara S, Fasano A. Celiac disease and the microbiome. *Nutrients*. 2019;11(10):2403.
20. Tang F, Sally B, Lesko K, et al. Cysteinyl leukotrienes mediate lymphokine killer activity induced by NKG2D and IL-15 in cytotoxic T cells during celiac disease. *J Exp Med*. 2015;212(10):1487–1495.
21. Castellanos-Rubio A, Fernandez-Jimenez N, Kratchmarov R, et al. A long noncoding RNA associated with susceptibility to celiac disease. *Science*. 2016;352(6281):91–95.
22. Malamut G, Cording S, Cerf-Bensussan N. Recent advances in celiac disease and refractory celiac disease. *F1000Res*. 2019;8:F1000.
23. Fiocchi C. Inflammatory bowel disease pathogenesis: where are we? *J Gastroenterol Hepatol*. 2015;30(suppl 1):12–18.
24. Ananthakrishnan AN, Kaplan GG, Ng SC. Changing global epidemiology of inflammatory bowel diseases: sustaining health care delivery into the 21st century. *Clin Gastroenterol Hepatol*. 2020;18(6):1252–1260.
25. Cosnes J, Cattan S, Blain A, et al. Long-term evolution of disease behavior of Crohn's disease. *Inflamm Bowel Dis*. 2002;8(4):244–250.
26. Powrie F, Leach MW, Mauze S, et al. Inhibition of Th1 responses prevents inflammatory bowel disease in SCID mice reconstituted with CD45RBhi $CD4^+$ T cells. *Immunity*. 1994;1(7):553–562.
27. Yen D, Cheung J, Scheerens H, et al. IL-23 is essential for T cell-mediated colitis and promotes inflammation via IL-17 and IL-6. *J Clin Invest*. 2006;116(5):1310–1316.

28. Reinisch W, de Villiers W, Bene L, et al. Fontolizumab in moderate to severe Crohn's disease: a phase 2, randomized, double-blind, placebo-controlled, multiple-dose study. *Inflamm Bowel Dis*. 2010;16(2):233–242.
29. Mannon PJ, Fuss IJ, Mayer L, et al. Anti-interleukin-12 antibody for active Crohn's disease. *N Engl J Med*. 2004;351(20):2069–2079.
30. Sandborn WJ, Gasink C, Gao LL, et al. Ustekinumab induction and maintenance therapy in refractory Crohn's disease. *N Engl J Med*. 2012;367(16):1519–1528.
31. Hueber W, Sands BE, Lewitzky S, et al. Secukinumab, a human anti-IL-17A monoclonal antibody, for moderate to severe Crohn's disease: unexpected results of a randomised, double-blind placebo-controlled trial. *Gut*. 2012;61(12):1693–1700.
32. de Lange KM, Moutsianas L, Lee JC, et al. Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease. *Nat Genet*. 2017;49(2):256–261.
33. Kuballa P, Huett A, Rioux JD, et al. Impaired autophagy of an intracellular pathogen induced by a Crohn's disease associated ATG16L1 variant. *PLoS One*. 2008;3(10):e3391.
34. Sarin R, Wu X, Abraham C. Inflammatory disease protective R381Q IL23 receptor polymorphism results in decreased primary $CD4^+$ and $CD8^+$ human T-cell functional responses. *Proc Natl Acad Sci U S A*. 2011;108(23):9560–9565.
35. Glassner KL, Abraham BP, Quigley EMM. The microbiome and inflammatory bowel disease. *J Allergy Clin Immunol*. 2020;145(1):16–27.
36. Franzosa EA, Sirota-Madi A, Avila-Pacheco J, et al. Gut microbiome structure and metabolic activity in inflammatory bowel disease. *Nat Microbiol*. 2019;4(2):293–305.
37. Mitsuyama K, Niwa M, Takedatsu H, et al. Antibody markers in the diagnosis of inflammatory bowel disease. *World J Gastroenterol*. 2016;22(3):1304–1310.
38. Nguyen NH, Singh S, Sandborn WJ. Positioning therapies in the management of Crohn's disease. *Clin Gastroenterol Hepatol*. 2020;18(6):1268–1279.
39. Barnes EL, Lightner AL, Regueiro M. Perioperative and postoperative management of patients with Crohn's disease and ulcerative colitis. *Clin Gastroenterol Hepatol*. 2020;18(6):1356–1366.
40. Murthy SK, Begum J, Benchimol EI, et al. Introduction of anti-TNF therapy has not yielded expected declines in hospitalisation and intestinal resection rates in inflammatory bowel diseases: a population-based interrupted time series study. *Gut*. 2020;69(2):274–282.
41. Loftus Jr. EV, Harewood GC, Loftus CG, et al. PSC-IBD: a unique form of inflammatory bowel disease associated with primary sclerosing cholangitis. *Gut*. 2005;54(1):91–96.
42. Ungaro R, Mehandru S, Allen PB, et al. Ulcerative colitis. *Lancet*. 2017;389(10080):1756–1770.
43. Fumery M, Dulai PS, Gupta S, et al. Incidence, risk factors, and outcomes of colorectal cancer in patients with ulcerative colitis with low-grade dysplasia: a systematic review and meta-analysis. *Clin Gastroenterol Hepatol*. 2017;15(5):665–674. e5.
44. Fuss IJ, Strober W. The role of IL-13 and NK T cells in experimental and human ulcerative colitis. *Mucosal Immunol*. 2008;1(suppl 1):S31–S33.
45. Heller F, Florian P, Bojarski C, et al. Interleukin-13 is the key effector Th2 cytokine in ulcerative colitis that affects epithelial tight junctions, apoptosis, and cell restitution. *Gastroenterology*. 2005;129(2):550–564.
46. Reinisch W, Panes J, Khurana S, et al. Anrukinzumab, an anti-interleukin 13 monoclonal antibody, in active UC: efficacy and safety from a phase IIa randomised multicentre study. *Gut*. 2015;64(6):894–900.
47. Beaudoin M, Goyette P, Boucher G, et al. Deep resequencing of GWAS loci identifies rare variants in CARD9, IL23R and RNF186 that are associated with ulcerative colitis. *PLoS Genet*. 2013;9(9):e1003723.
48. Cho JH, Brant SR. Recent insights into the genetics of inflammatory bowel disease. *Gastroenterology*. 2011;140(6):1704–1712.
49. Farh KK, Marson A, Zhu J, et al. Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature*. 2015;518(7539):337–343.
50. Mokry M, Middendorp S, Wiegerinck CL, et al. Many inflammatory bowel disease risk loci include regions that regulate gene expression in immune cells and the intestinal epithelium. *Gastroenterology*. 2014;146(4):1040–1047.
51. Danese S, Fiorino G, Peyrin-Biroulet L. Positioning therapies in ulcerative colitis. *Clin Gastroenterol Hepatol*. 2020;18(6):1280–1290. e1.
52. Gentile N, Yen EF. Prevalence, pathogenesis, diagnosis, and management of microscopic colitis. *Gut Liver*. 2018;12(3):227–235.
53. Kao KT, Pedraza BA, McClune AC, et al. Microscopic colitis: a large retrospective analysis from a health maintenance organization experience. *World J Gastroenterol*. 2009;15(25):3122–3127.
54. Stewart MJ, Seow CH, Storr MA. Prednisolone and budesonide for short- and long-term treatment of microscopic colitis: systematic review and meta-analysis. *Clin Gastroenterol Hepatol*. 2011;9(10):881–890.
55. Esteve M, Mahadevan U, Sainz E, et al. Efficacy of anti-TNF therapies in refractory severe microscopic colitis. *J Crohns Colitis*. 2011;5(6):612–618.
56. Shaheen NJ, Mukkada V, Eichinger CS, et al. Natural history of eosinophilic esophagitis: a systematic review of epidemiology and disease course. *Dis Esophagus*. 2018;31(8):doy015.
57. Davis BP, Rothenberg ME. Mechanisms of disease of eosinophilic esophagitis. *Annu Rev Pathol*. 2016;11:365–393.
58. Wen T, Stucke EM, Grotjan TM, et al. Molecular diagnosis of eosinophilic esophagitis by gene expression profiling. *Gastroenterology*. 2013;145(6):1289–1299.
59. Cheng E, Zhang X, Huo X, et al. Omeprazole blocks eotaxin-3 expression by oesophageal squamous cells from patients with eosinophilic oesophagitis and GORD. *Gut*. 2013;62(6):824–832.
60. Amininejad L, Charloteaux B, Theatre E, et al. Analysis of genes associated with monogenic primary immunodeficiency identifies rare variants in XIAP in patients with Crohn's disease. *Gastroenterology*. 2018;154(8):2165–2177.
61. Agarwal S, Cunningham-Rundles C. Gastrointestinal manifestations and complications of primary immunodeficiency disorders. *Immunol Allergy Clin North Am*. 2019;39(1):81–94.
62. Pulvirenti F, Pecoraro A, Cinetto F, et al. Gastric cancer is the leading cause of death in Italian adult patients with common variable immunodeficiency. *Front Immunol*. 2018;9:2546.
63. Shulzhenko N, Dong X, Vyshenska D, et al. CVID enteropathy is characterized by exceeding low mucosal IgA levels and interferon-driven inflammation possibly related to the presence of a pathobiont. *Clin Immunol*. 2018;197:139–153.
64. Mannon PJ, Fuss IJ, Dill S, et al. Excess IL-12 but not IL-23 accompanies the inflammatory bowel disease associated with common variable immunodeficiency. *Gastroenterology*. 2006;131(3):748–756.
65. Woodward JM, Gkrania-Klotsas E, Cordero-Ng AY, et al. The role of chronic norovirus infection in the enteropathy associated with common variable immunodeficiency. *Am J Gastroenterol*. 2015;110(2):320–327.
66. Washington K, Stenzel TT, Buckley RH, Gottfried MR. Gastrointestinal pathology in patients with common variable immunodeficiency and X-linked agammaglobulinemia. *Am J Surg Pathol*. 1996;20(10):1240–1252.
67. Pensieri MV, Pulvirenti F, Schiepatti A, et al. The high mortality of patients with common variable immunodeficiency and small bowel villous atrophy. *Scand J Gastroenterol*. 2019;54(2):164–168.
68. Matute JD, Arias AA, Wright NA, et al. A new genetic subgroup of chronic granulomatous disease with autosomal recessive mutations in p40 phox and selective defects in neutrophil NADPH oxidase activity. *Blood*. 2009;114(15):3309–3315.
69. Leiding JW, Freeman AF, Marciano BE, et al. Corticosteroid therapy for liver abscess in chronic granulomatous disease. *Clin Infect Dis*. 2012;54(5):694–700.
70. Wang J, Mayer L, Cunningham-Rundles C. Use of GM-CSF in the treatment of colitis associated with chronic granulomatous disease. *J Allergy Clin Immunol*. 2005;115(5):1092–1094.
71. Hahn KJ, Ho N, Yockey L, et al. Treatment with anakinra, a recombinant IL-1 receptor antagonist, unlikely to induce lasting remission in patients with CGD colitis. *Am J Gastroenterol*. 2015;110(6):938–939.
72. Lublin M, Bartlett DL, Danforth DN, et al. Hepatic abscess in patients with chronic granulomatous disease. *Ann Surg*. 2002;235(3):383–391.
73. Hsieh KH, Chang MH, Lee CY, Wang CY. Wiskott-Aldrich syndrome and inflammatory bowel disease. *Ann Allergy*. 1988;60(5):429–431.
74. Conrad MA, Kelsen JR. Genomic and immunologic drivers of very early-onset inflammatory bowel disease. *Pediatr Dev Pathol*. 2019;22(3):183–193.

76

Inflammatory Hepatobiliary Diseases

Benedetta Terziroli Beretta-Piccoli, Carlo Selmi, Michael P. Manns, and M. Eric Gershwin

Inflammatory hepatobiliary disease refers to chronic autoimmune diseases with predominant hepatic or biliary manifestations, without evidence of infectious diseases. Based on the target tissue, we can distinguish autoimmune hepatitis (AIH), targeting hepatocytes, from primary biliary cholangitis (PBC) and primary sclerosing cholangitis (PSC), both targeting the biliary tract. Cirrhosis and eventually liver failure are the common evolution of inflammatory hepatobiliary diseases, regardless of the target tissue, despite pathogenesis and therapeutics varying in between the spectrum of inflammatory hepatobiliary diseases.[1,2] In this chapter we aim to elucidate the main characteristics of AIH, PBC, and PSC, with particular interest for the clinical manifestations, pathogenesis, autoantibodies, and therapeutic options.

Immunoglobulin G4 (IgG4)-mediated cholangitis is a more recently reported clinical entity belonging to the group of immune-mediated hepatobiliary disorders, and will therefore also be briefly discussed in the present chapter.

AUTOIMMUNE HEPATITIS

Definition

AIH is defined as a chronic, inflammatory liver disease of unknown etiology, affecting all ages and races, and is characterized by female preponderance, elevated serum transaminase and immunoglobulin G levels, positive circulating autoantibodies, interface hepatitis at liver histology, and a swift response to corticosteroid treatment. If untreated, the disease invariably progresses over time to cirrhosis and eventually liver failure: the reported 5- and 10-year survival rates without treatment are 50% and 10%, respectively.[3]

Two types of AIH are distinguished based on the autoimmune serological profile: AIH type 1 is characterized by positive anti-nuclear (ANA) and/or anti-smooth muscle (SMA) antibodies, and AIH type 2 by anti-liver kidney microsomal type 1 (LKM1) and/or anti-liver cytosol (LC1) type 1 antibodies. Type 1 AIH can affect people of any age and sex, accounting for two-thirds of juvenile AIH cases. Type 2 AIH primarily affects young children and adolescents/young adults; onset later in life is virtually absent.

Epidemiology

Albeit an accurate estimate of AIH incidence and prevalence is not available, AIH type 1 incidence is supposed to be between 1.1 and 3 per 100,000 person-years in adults and between 0.23 and 0.83 in children, and AIH-1 prevalence between 16.9 and 35.9 in adults (being highest in Alaska) and 11.6 in children.[2,4] Therefore, AIH meets the definition of a rare disease.

AIH has a female preponderance, with a male:female ratio of 1:4, and a two-peak incidence during adolescence and at 30 to 45 years of age.[2] Type 2 AIH, which typically affects children and adolescents, is much rarer than type 1 AIH—the precise incidence and prevalence being largely unknown.

Risk Factors and Pathophysiology

Although the AIH etiology remains unknown, it results from the immune-mediated destruction of hepatocytes.[5] Both genetics and environmental risk factors contribute to the development of the disease. Susceptibility to AIH is strongly influenced by the large and highly polymorphic human leukocyte antigen (HLA) class II region (Chapter 5), whose protein products play a key role in the presentation of antigenic peptides to T cells.[5] There is a geographical variability of HLA alleles associated to AIH, DR3 (*DRB1*0301*), and DR4 (*DRB1*0401*) being found in 70% of AIH patients in Europe and North America, whereas *DRB1*0405* and *DRB1*0404* confer AIH susceptibility in populations living in Argentina, Mexico, and Japan. Of note, both *DRB1*0301* and *DRB1*0401* alleles code for the hexameric amino acid sequence LLEQKR at positions 67 to 72, whereby *DRB1*0405* and *DRB1*0404* allele encode a very similar hexameric sequence with arginine instead of lysine at position 7, suggesting that this sequence with a basic amino acid at position 71 is critical to confer susceptibility to AIH.[6]

Interestingly, *DRB1*0401* was found to be protective in juvenile AIH, whereas *DRB1*0301* is a risk allele also in children and adolescents.[6] Type 2 AIH is linked to possession of the *DRB1*0701* and *DRB1*0301* alleles. Moreover, HLA alleles have been linked to AIH clinical manifestations, as well as autoimmune serology, response to treatment, and prognosis.

Environmental factors have also been linked to AIH development, whereby the best investigated mechanism is molecular mimicry, a process leading to exaggerated immune response against self-components with structural homology to foreign antigens.[7] Supporting this hypothesis, it has been demonstrated that the hepatitis C virus (HCV) shares high amino acid sequence homology with the cytochrome P4502D6, which is the auto-antigenic target of anti-LKM1, and that up to 10% of HCV patients are seropositive for anti-LKM1. Nonviral environmental factors include medications, in particular antibiotics (nitrofurantoin and minocycline), statins, anti-tumor necrosis factor-α (TNF-α) agents (adalimumab and infliximab) and herbal remedies; however, the disease is different from classical AIH, and usually does not require long-term immunosuppressive treatment.

Therefore, a careful history on drug and herbal remedies intake is essential in patients presenting with suspected AIH.[8]

From an immunological standpoint, liver damage is initiated by the presentation of an autoantigenic peptide to a naïve CD4 T cell, which, depending on the cytokine milieu, will differentiate into a Th1, Th2, or Th17 cell (Chapters 10 and 11). These effector cells initiate a cascade of immune reactions: (a) Th1 cells secrete mainly interleukin (IL)-2 and interferon (IFN)-γ; IFN-γ is considered the main orchestrator of tissue damage since it stimulates CD8 cells, enhances the expression of HLA class I, induces the expression of HLA class II molecules on hepatocytes, and activates monocytes/macrophages, which in turn release IL-1 and TNF; (b) Th2 cells produce IL-4, IL-10, and IL-13, cytokines that induce the maturation of B cells into plasma cells, with consequent production of autoantibodies; (c) Th17 cells, which arise in the presence of transforming growth factor (TGF)-β and IL-6, produce IL-17, IL-21, IL-22, TNF, and chemokine ligand (CCL)-20.[2] Th17 cells, which play a central role in PBC, are being evaluated also in AIH as an increased number of Th17 cells have been reported in the peripheral blood and liver of patients with AIH compared to healthy controls.[1,9]

In healthy individuals, regulatory T cells (Treg), constituting 5% to 10% of peripheral CD4 T cells, are key players in limiting the ability of physiologically present autoreactive T cells to cause tissue damage. In AIH, Treg are numerically and functionally impaired, particularly during active disease phases. Therefore, therapeutic approaches aiming at restoring the number and the function of Treg are under investigation.

Clinical Presentation and Biochemical Features

While the majority of adult AIH patients present with mild, unspecific or no symptoms, an acute icteric hepatitis associated with hypergammaglobulinemia is the initial presentation in 20% to 30% of cases. Rarely, patients present with a complication of portal hypertension, and, very rarely, with acute liver failure.[10] Acute presentation is more frequent in children compared to adults.[2]

Some 20% of AIH patients have concomitant extrahepatic autoimmune diseases: therefore, presenting symptoms may be attributable to these diseases, which include, among others, Hashimoto thyroiditis, rheumatoid arthritis, lupus erythematosus, inflammatory bowel disease (IBD), alopecia, and psoriasis.[2,11]

Blood tests usually show marked elevation of aminotransferases; γ-glutamyl transferase is also frequently elevated, whereas alkaline phosphatase is generally normal in adults, but in children it is elevated due to bone growth. Direct bilirubin ranges from normal to markedly elevated with jaundice. Furthermore, serum globulins, particularly of the γ type, are commonly increased in AIH cases regardless of the histological stage.

At the time of diagnosis, approximately 30% of adult and pediatric patients have histological evidence of cirrhosis; however, when appropriately treated, only a small number of patients develop cirrhosis during follow-up if biochemical remission is achieved.[2,12,13] Cirrhosis at presentation is probably not associated with worse outcomes.[2,12] The occurrence of hepatocellular carcinoma (HCC) in patients with AIH is a rare event and only develops in long-standing cirrhosis.

The clinical course of AIH without treatment is burdened by high mortality, with 5- and 10-year survival rates estimated as 50% and 10%, respectively, but the use of corticosteroids has dramatically improved the course of the disease with a 10-year survival rate exceeding 90%.[8] The complications associated with AIH are similar to those of other progressive liver diseases as chronic hepatitis can evolve to cirrhosis and ultimately to HCC despite the use of immunosuppressive therapy, particularly if biochemical remission is not achieved.

Serum Autoantibodies

Serum autoantibodies are a key feature of AIH, being positive in up to 95% of the patients if tested according to international, dedicated guidelines issued in 2004 to overcome the lack of standardization (Table 76.1).[2,14] According to these guidelines, liver-relevant autoantibodies should be tested, at a screening level, by indirect immunofluorescence on triple rodent tissue: this methodology allows for the simultaneous detection of ANA, SMA, anti-LKM1, anti-LC1 and anti-mitochondrial antibody (AMA). The positivity cut-off is ≥1:40 in adults, while in children even lower titers are of clinical significance: that is, ≥1:20 for ANA and SMA, and ≥1:10 for anti-LKM1 and anti-LC1. ANA-positive sera should be further tested on HEp2 cells, allowing the detection of the nuclear pattern. Anti-neutrophil cytoplasmic antibody (ANCA) should be tested using human fixed neutrophils as an indirect immunofluorescence substrate. Autoantibody for which the target antigen has been identified are increasingly tested by molecular-based assays. The initial screening for liver-relevant autoantibodies should include, besides immunofluorescence on triple rodent tissue, a molecular-based assay for anti-soluble liver antigen (SLA), whose AIH specificity is as high as 98.9%, despite a relatively low sensitivity of about 30%.[2]

ANA in AIH patients most frequently produces a homogeneous pattern in immunofluorescence on HEp cells.[15] However, ANA is not specific for AIH, being not uncommon in patients with viral or other autoimmune liver diseases, with systemic/extrahepatic diseases, as well as in as many as 15% of healthy subjects, especially in older age groups.[16] Moreover, ANA positivity, immunofluorescence pattern, or titer does not reflect different AIH phenotypes and they cannot predict the disease's natural history.

Serum SMA is an autoantibody reacting with different proteins (actin, tubulin, vimentin, desmin, cytokeratins) of the cytoskeletal components. Its presence lacks AIH specificity, being detected in the serum of patients with other hepatic and extrahepatic autoimmune diseases, such as PSC or celiac disease, and in non-autoimmune liver diseases, including viral hepatitis C and E, and non-alcoholic steatohepatitis. However, when detected at high titers (>1:80), SMA is considered a sensitive marker for AIH-1, being found in up to 85% of cases.[15] In addition, the staining pattern in indirect immunofluorescence on rodent kidney tissue is of diagnostic help: staining of glomerula and tubuli, besides vessels, is more frequently observed in sera of AIH than non-AIH cases.[15]

Autoantibody against LKM1 is the serological hallmark of AIH-2. In indirect immunofluorescence on triple rodent tissue, it stains the proximal, larger renal tubuli and the hepatocellular cytoplasm, sparing stomach tissue. The 50-kDa autoantigen was identified as the cytochrome P450 2D6 (CYP2D6). Anti-LKM1 is detected also in serum of up to 13% of patients with chronic HCV infection, sharing target epitopes of CYP2D6 with the anti-LKM1 detected in AIH-2.[15] Importantly, anti-LKM1 titers correlate with disease activity of AIH-1, and should be used to monitor patients. The pathogenic role of anti-LKM-1 is still debated, despite the development of AIH-2 animal models following immunization with human CYP2D6 or with adenoviruses

TABLE 76.1 Most Relevant Features of Liver-Relevant Autoantibodies

Specificity	Known Antigenic Targets	Methods of Detection	Frequency	Diagnostic Role	Other Associated Liver Diseases	Comments
ANA	Homogeneous, speckled, nucleolar pattern: Chromatin, histones, cyclin A, ribonucleoproteins, double-/single-stranded DNA, SSA, SSB, Scl70, Smith Rim-like/membranous pattern: gp210, nucleoporin p62, lamin B receptor Multiple nuclear dots pattern: Sp100, promyelocytic leukemia protein, Sp140, small ubiquitin-related modifiers Centromeric pattern: CENP-A, CENP-B, CENP-C, CENP-D, CENP-E, CENP-F	IIF	AIH-1 and ASC: 75% PBC: 10%–65% PSC: 8%–77%	AIH: concomitant SMA confers 99% diagnostic specificity PBC: Rim-like/membranous and multiple nuclear dots patterns virtually diagnostic; Anti-centromere rarely present in isolation PSC: may suggest AIH overlap	Viral hepatitis DILI NAFLD Wilson disease HCC	AIH: Homogeneous pattern in ~75% of patients Speckled or nucleolar pattern in ~25% of patients PBC: Rim-like/membranous and, multiple nuclear dots associated with worse outcomes Anti-centromere associated with portal hypertensive phenotype
SMA	Filamentous actin Vimentin Desmin	IIF	AIH-1: 85% ASC: 75% PSC: up to 83% PBC: unknown	VG and VGT IIF patterns specific for AIH-1 Concomitant SMA: 99% diagnostic specificity	DILI NAFLD Viral hepatitis Wilson disease	V pattern in 20% of type 1 AIH patients Titers correlate with disease activity
Anti-actin antibody	Actin	Molecular-based assays	AIH-1: 60%	Specific for AIH-1 at high titers	DILI NAFLD Viral hepatitis Wilson disease	Sensitivity and specificity depend on the cut-off point. Less specific for AIH than the VG/VGT IIF pattern
Anti-LKM1	Epitopes of CYP2D6	IIF Molecular-based assays	AIH-2: up to 90%	Specific for AIH-2 in absence of hepatitis C	Hepatitis C	Titers correlate with disease activity Post-transplant reappearance predicts AIH-2 recurrence Atypical anti-LKM1 associated with de novo AIH
Anti-LC1	Formiminotransferase cyclodeaminase	IIF Molecular-based assays	AIH-2: 60%	Specific for AIH-2 in absence of hepatitis C	Hepatitis C	Only serological marker in 10%–30% of AIH-2 cases Titers correlate with disease activity
Anti-SLA	*O*-Phosphoseryl-tRNA (Sec) selenium transferase	Molecular-based assays	AIH-1 and AIH-2: up to 58% ASC: up to 41%	Highly specific for AIH. Low disease sensitivity when tested with commercial solid phase assays	Extremely rare in hepatitis C	Associated with worse outcomes
pANNA	Beta-tubulin isotype 5; HMG1; HMG2; other unknown autoantigens	IIF	AIH-1: 40%–96% PSC: 26%–94% ASC: up to 74%	Specific for AIH-1, PSC and IBD		May be the only serological marker in AIH-1 Absent in type II AIH IBD and sclerosing cholangitis must be excluded in pANNA-positive patients
AMA	E2 subunits lipoyl domains of PDC, OGDC and BCOADC, E3 binding protein of PDC	IIF Molecular-based assays	PBC: up to 95%	Highly specific of PBC	Acute liver failure PBC/AIH overlap	Predicts PBC if detected in absence of liver disease May be associated with histological PBC changes even in absence of biochemical cholestasis

AIH, Autoimmune hepatitis; *AMA*, anti-mitochondrial antibody; *ANA*, anti-nuclear antibody; *ASC*, autoimmune sclerosing cholangitis; *BCOADC*, branched-chain 2-oxo acid dehydrogenase complex; *CENP-C*, Centromere protein C; *CENP-D*, Centromere protein D; *CENP-E*, Centromere protein E; *CENP-F*, Centromere protein F; *DILI*, drug-induced liver injury; *HCC*, hepatocellular carcinoma; *HMG1*, high mobility group non-histone chromosomal protein; *IBD*, inflammatory bowel syndrome; *IIF*, indirect immunofluorescence; *LC1*, liver cytosol type 1; *LKM*, liver kidney microsomal; *NAFLD*, non-alcoholic fatty liver disease; *OGDC*, 2-oxoglutarate dehydrogenase complex; *pANNA*, perinuclear anti-neutrophil nuclear antibody; *PBC*, primary biliary cholangitis; *PDC*, pyruvate dehydrogenase complex; *PSC*, primary sclerosing cholangitis; *Scl70*, autoantibodies against topoisomerase I; *SLA*, soluble liver antigen; *SMA*, smooth muscle antibody; *SSA*, Sjögren's syndrome A antigen; *SSB*, Sjögren's syndrome B antigen *T*, tubulus; *V*, vessel; *G*, glomerulus; *VG*, vessels glomeruli; *VGT*, vessels glomeruli tubules.

delivering human CYP2D6 to hepatocytes. Finally, two other types of anti-LKM, giving slightly different staining patterns in indirect immunofluorescence, have been described in patients with ticrynafen-associated hepatitis (anti-LKM2, directed against CYP2C9) and in 19% of patients with type 2 AIH (anti-LKM3, directed against UGT1A), either alone or in combination with LKM1 antibodies. Anti-LKM3 is also detected in 13% of patients with chronic hepatitis delta infection. Its clinical use is hampered by detection only on primate or humane substrates.

Anti-soluble liver antigen (SLA) antibody is detectable by immunoblot and enzyme-linked immunosorbent assay, but not by immunofluorescence, and is directed against different epitopes of a UGA tRNA suppressor. Serum anti-SLA is occasionally found in patients with AIH who are negative for ANA, SMA, or anti-LKM and is cumulatively detected in 10% to 30% of cases of AIH-1 and -2. However, the assays used in clinical laboratories have low sensitivity; if tested by radioligand assays, which preserve conformational epitopes, anti-SLA are positive in 58% of AIH-1 and AIH-2, and in 41% of children with autoimmune sclerosing cholangitis (ASC).[15] Anti-SLA is the only autoantibody disease-specific autoantibody, having a specificity as high as 98.9%.[2] Its presence is also associated with a more aggressive disease course.

The anti-LC1 antibody is detected by indirect immunofluorescence in sera of up to 30% of patients with type 2 AIH and much less frequently in chronic hepatitis C. Importantly, however, anti-LC1 is the only detectable serological marker in 10% of AIH-2 cases. The LC1 autoantigen is the liver formiminotransferase cyclodeaminase, an enzyme involved in folate metabolism. Serum anti-LC1 antibody titer correlates with AIH-2 activity, and should be used to monitor patients.

Antibodies to neutrophil cytoplasmic antigens (ANCA) can be detected by indirect immunofluorescence on human neutrophil substrate, giving either A perinuclear (pANCA) or a cytoplasmic (cANCA) pattern. pANCA are frequently detected in sera from AIH-1 patients, although they give an atypical staining pattern; as opposed to pANCA, which give a cytoplasmic pattern on ethanol-fixed neutrophils, the perinuclear pattern of atypical pANCA is unaffected by the method of fixation. As a consequence, they are also known as perinuclear anti-neutrophil nuclear antibody (pANNA) or nuclear anti-neutrophil antibody (NANA). ANCA are absent in AIH-2, and are frequently detected also in PSC and IBD.

Histology

The role of liver histology in the diagnosis of AIH remains critical and all suspect cases should undergo a liver biopsy. In fact, although no typical feature can be sufficient to prove the diagnosis, histology remains the gold standard for grading and staging, and to rule out coexisting conditions, including overlap with PBC. Typical findings include portal hepatitis with mononuclear cell infiltrate and interface hepatitis (Fig. 76.1). Fibrosis is frequently observed; bridging or multiacinar necrosis indicates severe disease requiring prompt therapy. Importantly, the presence of granulomas, bile duct damage, or iron or copper accumulation should not be overlooked since these signs point towards alternative diagnoses. On the other hand, steatosis is a nonspecific and frequent finding that does not rule out AIH.[9]

Diagnostic Criteria

There is no single diagnostic test for AIH; therefore, diagnosis is based upon several indicative clinical, serological, biochemical, and histological findings. It is critical to exclude alternative causes of liver injury, particularly viral hepatitis. The International Autoimmune Hepatitis Group (IAIHG) has established, and subsequently simplified for clinical use, a set of diagnostic criteria with high sensitivity and specificity for the diagnosis (Table 76.2). Factors taken into account by the simplified scoring system include serum autoantibodies, liver histology, and exclusion of viral hepatitis.[17]

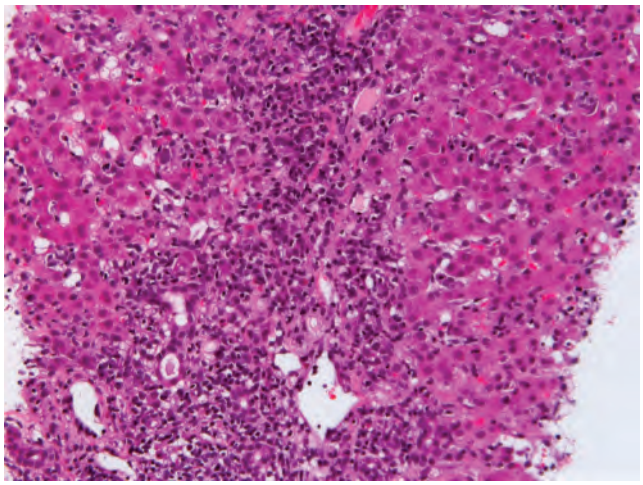


FIG. 76.1 Typical histological findings of autoimmune hepatitis showing marked lymphoplasmacellular infiltration of the portal tract with interface hepatitis. (Courtesy of Elisabetta Merlo.)

Therapy

Treatment of AIH aims at completely suppressing the hepatic inflammatory activity, which correlates with biochemical remission, defined as complete normalization of transaminase and IgG levels: patients achieving biochemical remission very rarely have histological progression.[18] Induction of remission is based on corticosteroids, in particular prednisone, while maintenance of remission is based on azathioprine, with or without low-dose prednisone, tailored to the individual risk–benefit profile. The initial prednisone dose is 0.5 to 1 mg/kg/day, but lower doses can be given in mild disease, and up to 100 mg/day intravenously should be used in case of severe disease. Steroids should be reduced at weekly intervals, under close monitoring of transaminase levels. A decrease in transaminase levels is almost universally achieved within 2 to 3 weeks; lack of response requires reconsideration of the diagnosis.[2] Once transaminase and bilirubin levels improve, azathioprine should be started, gradually reaching the dose of 1 to 2 mg/kg/day. Therapy should be maintained for at least 3 years, and an attempt of gradual withdrawal should be made only after stable biochemical remission. Relapses following therapy discontinuation are common since <20% of patients remain in sustained remission off treatment. Close monitoring of patients is required long term.

Mycophenolate mofetil is the drug of choice in case of azathioprine intolerance, with a reported response rate of 60% to 80%; however, it should be used with caution in fertile women due to its teratogenicity.

In case of insufficient response to initial treatment, azathioprine metabolites should be checked to assess both adherence and pharmacodynamics. Mycophenolate mofetil is rarely effective in patients not responding to azathioprine, and cyclosporin A or tacrolimus should be preferred in this situation. Infliximab and rituximab are reserved for very difficult patients and should be used only by specialized centers. Liver transplantation is the ultimate treatment for AIH patients presenting with acute liver failure not responding to steroids, or who

TABLE 76.2 Revised Diagnostic Scoring System Proposed by the International Autoimmune Hepatitis Study Group

Criteria	Points
Sex:	
Male	+2
Female	0
Ratio of ALP vs. AST/ALT:	
>3.0	+3
1.5–2.0	+2
1.0–1.5	+1
<1.0	0
Autoantibodies (ANA, SMA, LKM1) titer:	
>1:80	+3
1:80	+2
1:40	+1
<1:40	0
AMA:	
Positive	−4
Negative	0
Seropositivity for other autoantibodies	+2
Viral hepatitis markers:	
Negative	+3
Positive	−3
History of drug use:	
Yes	−4
No	+1
Average alcohol consumption (g/day):	
<25	+2
>60	−2
Presence of genetic factors (HLA, DR3, or DR4)	+1
Presence of other autoimmune disorders	+2
Liver histology:	
Interface hepatitis	+3
Predominant lymphocytic infiltrate	+1
Rosetting of liver cells	+1
None of the above	−5
Biliary changes	−3
Other changes	−3
Response to therapy:	
Complete	+2
Relapse	+3

A score >15 or >17 indicates a definite diagnosis of AIH pre- or post-treatment, respectively. Scores between 10–15 and 12–17 indicate a probable diagnosis, pre- or post-therapy, respectively.
ALP, Alkaline Phosphatase; *ALT*, alanine aminotransferase; *AST*, aspartate aminotransferase; *AMA*, Anti-mitochondrial autoantibodies; *ANA*, anti-nuclear; *DR3*, no expaned form; *DR4*, no expaned form; *HLA*, human leukocyte antigen; *LKM1*, anti-liver–kidney microsomal antibodies; *SMA*, anti-smooth-muscle antibodies.

progress to end-stage liver disease despite treatment (10% to 20% of the cases). Although liver transplantation for AIH is very successful, it should be noted that AIH may recur after transplant in up to 30% of the cases, rarely requiring a second transplant. Patients with AIH undergoing liver transplant have overall 5- and 10-year survival rates of 90% and 75%, respectively.

KEY CONCEPTS

- Autoimmune hepatitis (AIH) is a chronic inflammatory liver disease associated with high morbidity and mortality, especially due to the development of cirrhosis and liver failure.
- Autoantibodies represent a key diagnostic feature of AIH; the definition of an AIH subset is based on serum autoantibodies.
- Liver biopsy is mandatory to diagnose AIH, as histology remains the gold standard for grading and staging, and to rule out concomitant liver diseases.

CLINICAL PEARLS

- Serum transaminase levels can spontaneously improve in untreated autoimmune hepatitis (AIH): this should not delay treatment
- If autoantibodies are tested according to dedicated recommendations, up to 95% of AIH patients are positive for at least one autoantibody
- In children, overlap with sclerosing cholangitis is frequent, and magnetic resonance cholangiopancreatography should be part of the diagnostic work-up

THERAPEUTIC PRINCIPLES

- Corticosteroids represent the cornerstone of remission induction: they should be rapidly tapered under close monitoring of transaminase levels in order to minimalize side effects.
- Long-term maintenance therapy is based on azathioprine, or mycophenolate mofetil in intolerant patients.
- Dosage of steroids and azathioprine should be carefully tailored to the individual patient, taking into account disease severity, comorbidities, and side effects.
- Liver transplantation is very successful; however, autoimmune hepatitis may recur after transplant.

PRIMARY BILIARY CHOLANGITIS

Definition

PBC is defined as a chronic autoimmune attack to intrahepatic biliary epithelial cells. It is characterized by a striking female preponderance, biochemical cholestasis, pruritus, positive AMA, and progression to cirrhosis if left untreated.[1] PBC is absent in children, and the average age at diagnosis is within the 5th and 6th decade of life.

Nomenclature

In 2014, the term primary biliary cirrhosis has shifted towards PBC to correct the inaccuracy and remove the cirrhosis stigma as well the misunderstanding, disadvantages, and discriminations emanating from this misnomer for patients. This change was necessary after the dramatically improved PBC diagnosis, treatment, and prognosis, thanks to widely available AMA testing, and early treatment initiation with ursodeoxycholic acid (UDCA). Nowadays, two out of three PBC patients treated with UDCA have an expected survival equal to the general population, and only a minority will ever develop cirrhosis.[19]

Epidemiology

PBC is considered a rare disease: prevalence reported in Europe, North America, Australia, and Asia ranges from 19 to 402 cases per million inhabitants; incidence varies between 3.3 and 58 cases per million inhabitants per year.[20,21] Overall, the reported incidence and prevalence have increased in recent years, potentially due to more widely available AMA testing and increased awareness among clinicians, leading to diagnosis of mild, asymptomatic cases; however, a true increase in disease frequency cannot be excluded.

Pathogenesis

As in many autoimmune diseases, genetics, epigenetic, and environmental factors interact to lead to PBC, whose precise pathogenesis, however, remains elusive.

Genetic factors play a predominant role in PBC susceptibility, as suggested by the fact that the disease is more frequent in

relatives of affected individuals, and the term "familial PBC" has been coined to indicate families that have more than one case. In our experience, 6% of cases have a first-degree relative that is also affected, while AMA may be positive in first-degree relatives and offspring of patients with PBC without any sign of disease, thus indirectly suggesting the existence of a strong genetic predisposition.[1,21,22] Importantly, the concordance rate observed among monozygotic twins for PBC is 63%, which is among the highest reported in autoimmunity.[21] As for other autoimmune disorders, genetic factors are not limited to a single gene but are complex multi-gene traits.[21] HLA alleles have been identified by case-control studies, and later confirmed by a genome-wide association study, as the strongest link with PBC susceptibility.[21] Interestingly, PBC is associated with not only the HLA *DRB1*08* allele but also two protective alleles, *HLA DRB1*11* and *DRB1*13*. A wide array of non-HLA risk loci has also been identified, and, although the identified loci are different across studies, they involve the same immunological pathways, particularly antigen presentation and production of interleukin-12, and T- and B-cell activation.

In spite of these observations, genetic susceptibility alone is insufficient to explain the PBC pathogenesis. Other factors, including epigenetics and exposure to environmental factors, have been shown to play a key role in increasing the susceptibility to PBC. Urinary tract and vaginal infections are more frequent in PBC than in controls, with *Escherichia coli* being the main etiological agent. The mechanism linking infections and autoimmune biliary tract damage is probably molecular mimicry. Xenobiotics are foreign compounds, which may trigger an autoimmune response to self-proteins as they modify their molecular structures or complex to self- or nonself-proteins to generate neoantigens. In this view, our group has demonstrated that xenobiotic modification of the E2 subunit of the pyruvate dehydrogenase (PDC-E2) inner lipoyl domain can lead to loss of tolerance in genetically susceptible hosts.[23]

Epigenetic modifications could represent the link between genetic and environmental factors influencing the onset and evolution of PBC. For instance, DNA hypermethylation of a number of immunological genes on chromosome X has been identified in studies of discordant monozygotic twins and sister pairs, and decreased methylation of the CD40 ligand gene on chromosome X has been shown to correlate with serum IgM elevation, a typical feature of PBC. In addition, higher X chromosome monosomy has been demonstrated in PBC patients.[21] Differential expression of a wide range of micro-RNAs (mi-RNAs) in PBC compared to controls has also been reported, suggesting their potential role as novel biomarkers (Chapter 19). Moreover, a functional role of miRNA has robustly been demonstrated in PBC: upregulation of miR-506 correlates with decreased expression of the anion exchanger 2 protein—the main protein involved in biliary bicarbonate secretion. The bicarbonate-rich layer on the canalicular cholangiocyte membrane, named biliary bicarbonate umbrella, is crucial in protecting cholangiocytes from toxic bile acids, which in protonated form, can invade cells, leading to apoptosis and exposure of intracellular antigens, triggering an autoimmune attack. This cascade of events raises the fundamental question as to whether the disrupted biliary homeostasis in PBC is the cause or the consequence of the autoimmune attack. This question has to be addressed by future research.[1]

Diagnosis, Clinical Features, and Prognosis

The diagnosis of PBC is based on the presence of two out of three internationally accepted criteria: that is, detectable serum AMA (titer >1:40), increased cholestasis enzymes (i.e., alkaline phosphatase) for longer than 6 months, and a compatible or diagnostic liver histology.[24] Serum IgM is typically elevated in PBC cases, with no correlation with AMA titers or levels of other Ig subtypes.

Symptoms of PBC in the early phases, affecting more than half of the patients, are mainly fatigue and pruritus, while physical findings may include skin hyperpigmentation, hepatosplenomegaly, and xanthelasmas. Late-stage symptoms are those of all types of cirrhosis, including ascites, jaundice, hepatic encephalopathy, and upper digestive bleeding. However, portal hypertension may appear earlier in the disease course as compared to non-cholestatic diseases, owing to a pre-sinusoidal component; treatment of PBC-associated portal hypertension is not different from other chronic liver diseases. Fatigue is a nonspecific symptom affecting 70% of PBC patients, often overlooked, particularly in middle-aged women. The severity of fatigue is independent of the stage of PBC. Pruritus is more frequent in advanced-stage PBC, but may long precede jaundice onset, and typically worsens at night, following contact with wool, or in warm climates. Since both fatigue and pruritus are subjective, it is very important to collect information directly from patients, in a quantitative manner, in the form of "patient-reported outcomes." Dedicated, PBC-specific tools such as the PBC-40 questionnaire, are available, but must be validated locally.[1] Osteopenia and osteoporosis are more frequent in PBC patients when comparing to sex- and age-matched healthy individuals: therefore, screening with bone mineral density at diagnosis and during follow-up is advisable in all patients, as well as calcium and vitamin D supplementation, if needed. Hyperlipidemia affects up to 80% of patients with PBC. However, it is not accompanied by a proportionally increased incidence of cardiovascular events or atherosclerosis. This assumption has recently been challenged, and therefore clinicians should carefully evaluate cardiovascular risk of PBC patients.[24]

Extrahepatic Autoimmune diseases frequently overlap with PBC, most commonly Sjögren syndrome, Raynaud phenomenon, autoimmune thyroid disease, scleroderma, and systemic lupus erythematosus, while the prevalence of rheumatoid arthritis does not differ from controls.

Similarly to other types of cirrhosis, end-stage PBC can be complicated by the occurrence of hepatocellular carcinoma (HCC) and patients should be admitted to a surveillance program with ultrasound every 6 months.

The progression of PBC varies widely, and risk factors for a more aggressive course have been identified, including male sex, younger age at diagnosis, PBC-specific ANA positivity, lack of biochemical UDCA response (see below), and advanced fibrosis at diagnosis. An important message for patients is that UDCA initiation at an early disease stage is accompanied by 10-year survival rates similar to the general population.[1]

Serum Autoantibodies

AMA are highly specific for PBC and can be detected in up to 95% of patients when sensitive diagnostic methodologies based on recombinant antigens are used.[1] However, the gold standard method to detect AMA in clinical settings is still indirect immunofluorescence, because it allows the simultaneous detection of all liver-relevant autoantibodies, except anti-SLA. AMA is directed against components of the 2-oxoacid dehydrogenase family of enzymes within the mitochondrial respiratory chain, most frequently the E2- and E3-binding protein components of

the PDC and the E2 components of the 2-oxo-glutarate dehydrogenase and branched-chain 2-oxo acid dehydrogenase complexes. In all three antigens, epitopes contain the motif DKA, with lipoic acid covalently bound to the lysine (K) residue: lipoic acid is the immunodominant AMA epitope.[1] The pathogenic role of AMA is debatable, since no clinical correlation can be found and animal models developing serum AMA do not develop PBC-like liver lesions. Importantly, AMA may be present in the absence of biochemical cholestasis: in this setting, a liver biopsy may reveal PBC-compatible histology; in the absence of a liver biopsy or when histology is normal, patients deserve close follow-up, since clinically apparent disease develops over time in the majority of cases.[1]

Autoantibodies other than AMA can be found in the majority of PBC patients. ANA is present in 50% of PBC patients, the most common indirect immunofluorescence patterns on HEp2 cells being the "nuclear rim" and the "multiple nuclear dots" patterns. The target antigens of the nuclear rim pattern are gp210 and nucleoporin 62 (both within the nuclear pore complex of the nuclear envelope), whereas the multiple nuclear dots pattern corresponds mainly to antibodies against Sp100, promyelocytic leukemia protein, sp140, and small ubiquitin-like modifiers.[15] These two patterns are PBC specific, having the same diagnostic value as AMA: that is, being diagnostic of PBC when associated with a cholestatic biochemical profile. Therefore, the laboratory should always report the immunofluorescence pattern of ANA, and the clinician should be aware of the disease specificity of these two patterns. ANA-positive patients are more frequently AMA negative, possibly because of the lack of a masking effect of this latter antibody in indirect immunofluorescence. The pathogenic role of ANA in PBC remains to be investigated, although cross-sectional and longitudinal data demonstrate an association between PBC-specific ANA positivity, particularly the rim-like pattern, and a worse prognosis.[25] Finally, anti-centromere antibody, another ANA subtype, is typically detected in PBC patients with coexisting limited systemic sclerosis; very rarely, this serological specificity may be the only autoantibody in PBC.

Histology

PBC is characterized histologically by chronic non-suppurative destructive cholangitis, with bile duct loss (ductopenia) and granulomatous inflammation (Fig. 76.2). It is important to be aware that interface hepatitis, the typical histological finding of AIH, is universally detected in untreated PBC.

PBC histology has classically been assessed by two staging systems: by Scheuer in 1967, and Ludwig in 1978. They both recognize four stages, according to portal/periportal inflammation, ductular reaction, and fibrosis. A more recent staging system proposed by Nakanuma in 2010 includes three features more specific for PBC and correlating with disease progression.[1]

Therapy

The pharmacological treatment of PBC is based on drugs impacting the bile acid physiology, since immune-based therapies have failed to show effectiveness, despite the autoimmune basis of the disease, highlighting the key pathogenetic role of altered biliary homeostasis in PBC. UDCA is the first-line treatment, and is the only drug with a demonstrated beneficial effect on long-term transplant-free and overall survival.[24,25] The recommended dose is 13 to 15 mg/kg body weight/day. UDCA is a natural, non-toxic secondary bile acid, forming 1% to 3% of the bile acids pool in human bile. This percentage is increased to 40% in subjects treated with therapeutic UDCA doses.[26] One probable mechanism of UDCA action is therefore decreased bile toxicity; additional effects include stimulation of biliary bicarbonate secretion, immunomodulation, and antiinflammatory and anti-apoptotic effects.[1] UDCA is a safe drug; the only side effect is increased stool frequency. It should not be discontinued even during pregnancy and breast-feeding. The aim of UDCA treatment is biochemical response, defined according to different criteria based on bilirubin and alkaline phosphatase serum levels after 6, 12, or 24 months of UDCA treatment.[25] Patients achieving a biochemical response have better clinical outcomes and transplant-free survival. These criteria use thresholds of biochemical tests allowing dichotomization of patients in low- or high-risk groups for liver transplantation or death.[25] More recently, the GLOBE score and the UK-PBC score, which provide a continuous risk stratification, have been demonstrated to have excellent sensitivity and specificity for predicting

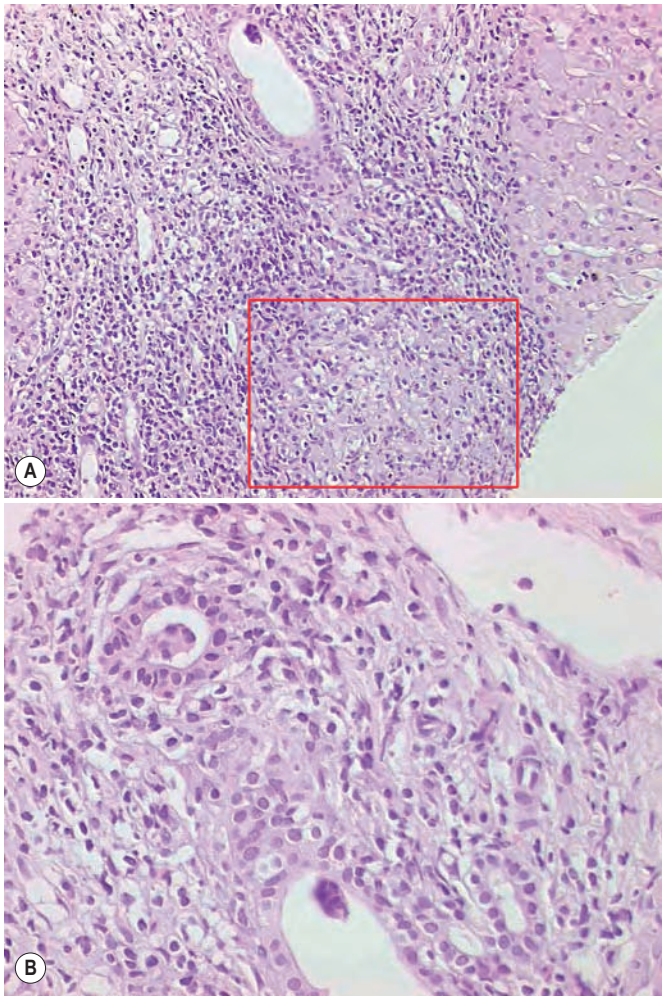


FIG. 76.2 Histological findings in early stages of primary biliary cholangitis: that is, non-suppurative destructive cholangitis (hematoxylin and eosin staining). (A) Mixed lymphocytic and plasma cell periductular inflammation with bile duct infiltration and granulomatous reaction *(square)*. (Magnification × 200.) (B) Detail of bile duct disruption with lymphocytic and plasmacellular periductular and intraepithelial infiltration. (Magnification × 200.)

transplant-free survival up to 15 years.[24] Recently, the GLOBAL PBC study group has refined the biochemical criteria associated with better outcomes, showing that bilirubin level <0.6 × upper limit of normal (ULN) and normal alkaline phosphatase level at 1 year are best in predicting the risk for liver transplantation or death.[27]

Of note, UDCA treatment has a beneficial impact on long-term survival even in patients without biochemical response, and should therefore be given lifelong to every PBC patient, with the exception of the rare patients who prove to be UDCA-intolerant.

Obeticholic acid (OCA), a synthetic analogue of the natural chenodeoxycholic acid, was approved in 2016 as second-line treatment for PBC patients with insufficient UDCA response or intolerance. OCA is an agonist of the nuclear receptor called farnesoid-X receptor (FXR); nuclear receptors are ligand-activated transcription factors that control transcription of a wide range of proteins involved in the immune system and metabolism. OCA is the first-in-class FXR agonist and has been shown to decrease bile acid synthesis and promote choleresis, although the precise downstream effects of FXR agonism in humans remain to be elucidated. OCA is associated with pruritus, which hampers its use in PBC patients already suffering from this symptom. In addition, dose reduction is needed in patients with advanced liver disease, and recent data suggest an increased risk of gallstones in patients treated with OCA. Newer FXR agonists, so-called nonsteroidal agonists, are being tested in PBC, and are expected to have fewer side effects.

Targeting the nuclear receptors known as peroxisome proliferator-activated receptors (PPARs) impacts bile acid synthesis and detoxification, and biliary phospholipid secretion. PPAR agonists include fenofibrate and bezafibrate, which have been on the market since the early 1970s to treat hyperlipidemia. Both drugs have been shown to improve biochemical cholestasis in PBC in small trials; these data have recently been confirmed in a large, placebo-controlled trial of bezafibrate as second-line treatment for PBC,[28] with a positive effect also on pruritus and hyperlipidemia. Of note, the primary endpoint of this trial was complete normalization of alkaline phosphatase, bilirubin, aminotransferases, albumin, and prothrombin index after 24 months. Efficacy on hard endpoints and long-term safety still need to be investigated. Importantly, fibrates use in PBC remains off-label.

Liver transplantation is the ultimate treatment for end-stage PBC, with survival rates of 92% and 85% at 1 and 5 years after transplant, respectively. Recurrence is common but rarely impacts on graft or patient survival; it is prevented by cyclosporine and UDCA.

KEY CONCEPTS

- Primary biliary cholangitis (PBC) is a chronic autoimmune cholestatic disease affecting more frequently women in the 5th–6th decades of life.
- Genetics play a strong role in PBC pathogenesis, as suggested by the familiar recurrence and the high concordance in monozygotic twins.
- Anti-mitochondrial antibodies (AMAs) are highly specific for PBC, and can be detected in nearly 100% of patients.
- Anti-nuclear antibody (ANA) giving a multiple nuclear dot or a nuclear rim staining pattern on HEp2 cells are PBC-specific, being of high diagnostic value in AMA-negative patients.

CLINICAL PEARLS

- Fatigue and pruritus represent the most frequently observed symptoms that may be disabling for patients.
- Primary biliary cholangitis (PBC) can be frequently associated with other autoimmune diseases: that is, Sjögren syndrome and systemic sclerosis.
- Clinically significant portal hypertension may occur in pre-cirrhotic stages.
- Osteopenia and osteoporosis are frequently observed in PBC patients.

THERAPEUTIC PRINCIPLES

- Ursodeoxycholic acid (UDCA; 13–15 mg/kg) is the cornerstone therapy for primary biliary cholangitis, improving long-term survival in all patients.
- Obeticholic acid improves biochemical cholestasis in patients intolerant to UDCA or with insufficient response; pruritus is a major safety issue of this drug.
- Fibrates, particularly bezafibrate, have been shown to be highly effective in improving cholestasis and pruritus in patients intolerant to UDCA or with insufficient response; data on long-term safety and efficacy are still lacking.

PRIMARY SCLEROSING CHOLANGITIS

Definition

PSC is a rare progressive cholestatic liver disease of unknown etiology characterized by inflammation, fibrosis, and destruction of intra- and extrahepatic bile ducts, by a strong association with inflammatory bowel disease (IBD), by male preponderance, and by high risk of cholangiocarcinoma and colorectal cancer.[29] It should be distinguished from sclerosing cholangitis related to other identifiable causes. Small-duct PSC, affecting 6% to 16% of PSC patients, is characterized by biochemical cholestasis and PSC typical histology but normal cholangiography. It has a more benign course but may progress to classical PSC over time.

Autoimmune sclerosing cholangitis is a pediatric nosological condition representing overlap of AIH with bile duct disease.[13] Its relationship with adult PSC is still debated.[13]

Epidemiology

The reported yearly incidence in Caucasians of PSC ranges between 0.91 and 1.3 per 100,000 inhabitants.[29] Its prevalence is approximately 10 to 15/100,000 in Northern Europe and the USA.[24] There is a North-South gradient, with 10 to 100 times lower prevalence in Southern Europe and Asia.

At variance from the vast majority of autoimmune diseases, PSC is more commonly diagnosed in men, with a female-to-male ratio estimated as 1:2, and generally between 30 and 40 years. However, recent data show a similar PSC prevalence in males and females, with a more benign course in females, suggesting underdiagnosis in asymptomatic female patients. Approximately 60% to 80% of Caucasian patients with PSC have coexisting IBD, most often ulcerative colitis.[29] Reported PSC prevalence in IBD patients varies between 5% and 7.5%.

Pathogenesis

The etiopathogenesis of PSC is unknown, despite evidence that (auto)immune-mediated mechanisms play a role, as suggested by the frequent association with IBD and other autoimmune

diseases, the presence of serum autoantibodies, and the reported HLA susceptibility associations. PSC autoantigens are unknown. Genetics play a role in determining PSC susceptibility; family studies have demonstrated that the PSC prevalence among first-degree relatives is 100 times higher compared to the unrelated population. The most significant PSC genetic risk factor is represented by HLA variations, suggesting a prominent role of adaptive immunity: positive associations have been reported with *HLA-B*08:01*, *HLA-DRB1*03:01, HLA-DRB1*13:01, HLA-DQA1*01:03*, and *HLA-DQA1*01:01*. The haplotypes *HLA-DRB1*04, DQB1*03:02,* and *HLA-DRB1*07:01* and *DQB1*03:03* confer protection.

Environmental factors contribute to PSC pathogenesis, although no definitive causal agent has been identified so far.[30] The close association with IBD suggests that the gut microbiome plays a pathogenic role in PSC. An altered microbiome has been reported in PSC, affecting not only the colon but also the oral cavity, the duodenum, and the bile.[31] An increased biliary concentration of toxic components, such as taurolithocholic acid, has also been observed in PSC patients, potentially due to abnormal bile acid microbial metabolism.[31]

Clinical Features and Diagnosis

The typical PSC patient is a 30- to 40-year-old male with IBD and elevated serum alkaline phosphatase and γ-glutamyltransferase serum levels. Of note, PSC may develop even after colectomy for IBD. In a subgroup of patients, IBD presents after the diagnosis of PSC, even after liver transplantation.

The diagnosis is based on typical magnetic resonance cholangiopancreatography (MRCP) bile duct changes; MRCP has been shown to have similar sensitivity and specificity as endoscopic retrograde cholangiopancreatography. The latter procedure should be performed only if an intervention is required. Typical MRCP findings are irregular bile ducts with a beading appearance due to multiple short strictures affecting most commonly both intrahepatic and extrahepatic bile ducts. However, singular intrahepatic involvement or, more rarely, isolated extrahepatic involvement are also encountered (Fig. 76.3). Liver biopsy is not part of the standard diagnostic work-up, but remains the key diagnostic procedure to diagnose small-duct PSC and overlap with AIH.

ANA and SMA are frequently positive in PSC patients; their presence requires exclusion of AIH overlap. Atypical pANCA are detected in serum of up to 94% of PSC patients; however, they lack disease specificity (as they are present also in IBD and AIH) and prognostic significance.

The histological hallmark of PSC is concentric, periductal, so-called onion-skin fibrosis (Fig. 76.4), which may be absent in early disease, though some degree of bile-duct damage is usually identified.

The genotype and phenotype of IBD associated with PSC is distinct from IBD without PSC, as it is characterized by pancolitis, backwash ileitis, and rectal sparing. An upper endoscopy and colonoscopy are recommended in every newly diagnosed PSC patient without known IBD, irrespective of gastrointestinal symptoms.

PSC symptoms are generally nonspecific and include right-upper-quadrant abdominal pain, jaundice, weight loss, pruritus, and fatigue; 40% to 50% of patients are asymptomatic at the time of diagnosis. At more advanced stages, symptoms include those of all types of decompensated cirrhosis or neoplasia; however, similarly to PBC, clinically significant portal hypertension may occur in pre-cirrhotic stages.

PSC is a progressive disease, leading to cirrhosis and liver failure in the vast majority of patients. However, mortality is mainly

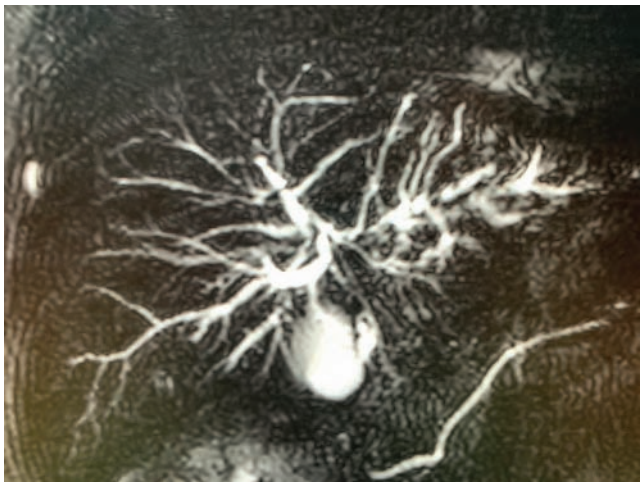


FIG. 76.3 Magnetic resonance cholangiopancreatography of intrahepatic biliary tree (sequence SENSE shimming (SSH), repetition time (TR) 5100, echo time (TE) 740, flip angle 90) showing intrahepatic bile ducts with marked caliber irregularities specially in the left liver lobe with narrowing and dilatation. (Courtesy of Mario Alerci.)

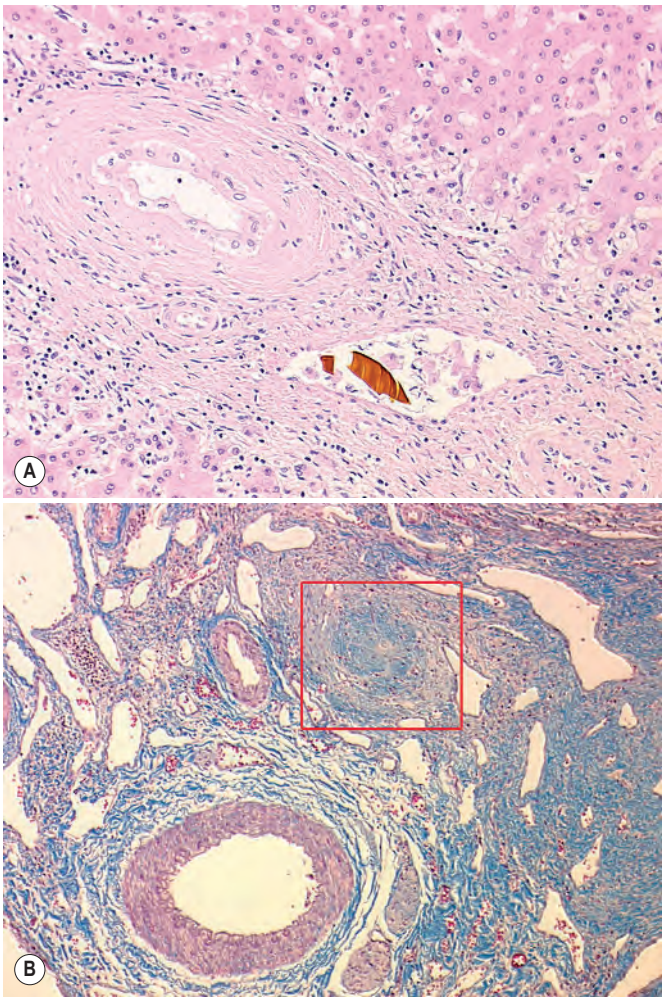


FIG. 76.4 Histological Findings in Primary Sclerosing Cholangitis. (A) Early disease and periductular fibrosis. (Magnification × 200, hematoxylin and eosin staining.) (B) Advanced disease with cirrhosis and bile duct substitution by fibrous scar *(square)*. (Magnification × 200, Masson staining.)

due to cholangiocarcinoma and colorectal cancer. Cholangiocarcinoma occurs in 20% of patients after a disease duration of 30 years, leading to the proposal of considering PSC as a pre-malignant condition. Cholangiocarcinoma should always be excluded in PSC patients with rapid deterioration. The diagnosis is often difficult in the early stages, since radiological distinction between malignancy and inflammatory/fibrotic lesions is inaccurate. Brush cytology may be of use, with variable sensitivity depending on lesion accessibility, as well as operator and pathologist experience. Cholangiocarcinoma surveillance with annual MRCP and contrast-enhanced liver magnetic resonance imaging coupled with serum CA19-9 is generally performed, despite the absence of evidence for survival advantage.

Annual colonoscopic surveillance of colorectal cancer is recommended in patients with concomitant IBD.

Therapy

There is no approved pharmaceutical therapy of PSC, since no drug has been demonstrated to have a beneficial effect on disease progression and survival.

UDCA at a dose of 10 to 20 mg/kg is widely prescribed based on a beneficial impact on biochemical cholestasis and decreased bile toxicity. High-dose UDCA (28 to 30 mg/kg/day) should be avoided since a high-quality study demonstrated a significant increase in clinical endpoints.

Endoscopic measures are indicated to treat bile duct strictures associated with worsening of symptoms and cholestasis, or bacterial cholangitis. Some studies have shown improved survival of patients undergoing endoscopic treatment of clinically significant strictures. Brush sampling should always be performed for cholangiocarcinoma surveillance.

Finally, PSC represents an important indication for liver transplantation since patients are commonly younger compared to other autoimmune liver diseases. Recurrence of disease is common and affects 20% to 40% of transplanted patients during prolonged follow-up.

KEY CONCEPTS

- Primary sclerosing cholangitis (PSC) is a rare, chronic autoimmune cholestatic disease that can affect all tracts of the biliary tree, including the extrahepatic bile ducts.
- PSC has a striking association with inflammatory bowel disease.
- PSC can be regarded as a pre-malignant condition, being complicated by cholangiocarcinoma in 20% of patients over a 30-year disease course.

CLINICAL RELEVANCE

- Primary sclerosing cholangitis (PSC) affects all ages.
- Diagnosis is based on the typical appearance of magnetic resonance cholangiopancreatography.
- Secondary forms of sclerosing cholangitis must be ruled out.
- Autoimmune sclerosing cholangitis is a pediatric clinical entity representing overlap of autoimmune hepatitis with sclerosing cholangitis.
- Small duct PSC is diagnosed in the presence of typical histology and normal imaging.

THERAPEUTIC PRINCIPLES

- Ursodeoxycholic acid (UDCA; 10–20 mg/kg) is widely prescribed, and is associated with improvement of biochemical cholestasis, without evidence of improved survival.
- Endoscopic procedures are indicated to treat stenosis.
- Primary sclerosing cholangitis represents an important indication for liver transplantation, despite disease recurrence in almost half of the cases.

OVERLAP/VARIANT SYNDROMES

A subgroup of PBC and PSC patients develop clinical, serological, histological, and radiological features of AIH and are classified as variant syndromes. This nomenclature has been proposed as being more appropriate compared to overlap syndromes, since the latter implies coexistence of different diseases in the same patient, while these patients should be considered as PBC or PBC with AIH manifestations.[32] In a minority of cases, AIH predates bile duct disease. Histology plays a central diagnostic role of the variant syndromes.

Treatment is empiric, and based on adding AIH treatment on top of UDCA in PBC and PSC patients, or, conversely, on adding UDCA in AIH patients with cholestatic features.

IMMUNOGLOBULIN 4-RELATED SCLEROSING CHOLANGITIS

IgG4-related sclerosing cholangitis is the bile tract manifestation of IgG4-related diseases and is characterized by inflammation and fibrosis leading to progressive stenosis and destruction of the bile ducts.[33] The condition affects more commonly males in their 60s. The clinical features include jaundice, pruritus, and biochemical cholestasis. Elevated serum IgG4 is found in up to 85% of the cases; the diagnosis can be aided by the IgG4/IgG1 ratio determined by polymerase chain reaction, since high serum IgG4 is present in 10% to 20% of PSC patients.[33,34] MRCP shows segmental or diffuse stenosis of the intra- and/or extrahepatic bile ducts with thickened walls. Typical histological findings include storiform fibrosis, lymphoplasmacytic infiltration, and obliterative phlebitis. Treatment is based on corticosteroids, which induce a remarkable improvement in clinical and biochemical manifestations, and should be started at the dosage of 0.6 mg/kg/day of prednisolone for 2 to 4 weeks, gradually tapered in 2 to 3 months. Long-term outcomes are excellent, despite 20% of relapse after corticosteroid discontinuation.

ON THE HORIZON

- Novel, more targeted therapeutic approaches are needed for autoimmune hepatitis, possibly including Treg adoptive therapy and low-dose interleukin (IL)-2. A better knowledge of the immunodominant epitopes could lead to the establishment of peptide immunotherapy.
- It should be possible to identify high-risk primary biliary cholangitis patients at diagnosis with novel biomarkers. Novel therapeutic approaches may include not only a combination of FXR+PPAR agonists and anti-IL-17 agents but also stage-specific treatments.
- The relationship of primary sclerosing cholangitis (PSC) with the pediatric condition known as autoimmune sclerosing cholangitis deserves further investigation and may lead to a better understanding of the PSC pathophysiology.

REFERENCES

1. Terziroli Beretta-Piccoli B, Mieli-Vergani G, Vergani D, et al. The challenges of primary biliary cholangitis: what is new and what needs to be done. *J Autoimmun.* 2019:102328. https://doi.org/10.1016/j.jaut.2019.102328.
2. Mieli-Vergani G, Vergani D, Czaja AJ, et al. Autoimmune hepatitis. *Nat Rev Dis Primer.* 2018;4:18017. https://doi.org/10.1038/nrdp.2018.17.
3. Liberal R, Krawitt EL, Vierling JM, et al. Cutting edge issues in autoimmune hepatitis. *J Autoimmun.* 2016;75:6–19. https://doi.org/10.1016/j.jaut.2016.07.005.
4. Hurlburt KJ, McMahon BJ, Deubner H, et al. Prevalence of autoimmune liver disease in Alaska Natives. *Am J Gastroenterol.* 2002;97:2402–2407. https://doi.org/10.1111/j.1572-0241.2002.06019.x.

5. Dyson JK, De Martin E, Dalekos GN, et al. Review article: unanswered clinical and research questions in autoimmune hepatitis—conclusions of the International Autoimmune Hepatitis Group Research Workshop. *Aliment Pharmacol Ther.* 2019;49:528–536. https://doi.org/10.1111/apt.15111.
6. Autoimmune hepatitis. *Nat Rev Dis Primer*. 2018;4:18018. https://doi.org/10.1038/nrdp.2018.18.
7. Vergani D, Mieli-Vergani G. Autoimmune manifestations in viral hepatitis. *Semin Immunopathol*. 2013;35:73–85. https://doi.org/10.1007/s00281-012-0328-6.
8. European Association for the Study of the Liver. EASL clinical practice guidelines: autoimmune hepatitis. *J Hepatol.* 2015;63:971–1004. https://doi.org/10.1016/j.jhep.2015.06.030.
9. Zhao L, Tang Y, You Z, et al. Interleukin-17 contributes to the pathogenesis of autoimmune hepatitis through inducing hepatic interleukin-6 expression. *PLoS One*. 2011;6:e18909. https://doi.org/10.1371/journal.pone.0018909.
10. Kamar N, Izopet J, Pavio N, et al. Hepatitis E virus infection. *Nat Rev Dis Primer*. 2017;3:17086. https://doi.org/10.1038/nrdp.2017.86.
11. Terziroli Beretta-Piccoli B, Invernizzi P, Gershwin ME, et al. Skin manifestations associated with autoimmune liver diseases: a systematic review. *Clin Rev Allergy Immunol.* 2017;53(3):394–412. https://doi.org/10.1007/s12016-017-8649-9.
12. Di Giorgio A, Hadzic N, Dhawan A, et al. Seamless management of juvenile autoimmune liver disease: long-term medical and social outcome. *J Pediatr.* 2020;218:121–129.e3. https://doi.org/10.1016/j.jpeds.2019.11.028.
13. Mieli-Vergani G, Vergani D, Baumann U, et al. Diagnosis and management of paediatric autoimmune liver disease: ESPGHAN Hepatology Committee position statement. *J Pediatr Gastroenterol Nutr*. 2018;66(2):345–360. https://doi.org/10.1097/MPG.0000000000001801.
14. Vergani D, Alvarez F, Bianchi FB, et al.; International Autoimmune Hepatitis Group. Liver autoimmune serology: a consensus statement from the committee for autoimmune serology of the International Autoimmune Hepatitis Group. *J Hepatol.* 2004;41:677–683. https://doi.org/10.1016/j.jhep.2004.08.002.
15. Terziroli Beretta-Piccoli B, Mieli-Vergani G, Vergani D. The clinical usage and definition of autoantibodies in immune-mediated liver disease: a comprehensive overview. *J Autoimmun.* 2018;95:144–158. https://doi.org/10.1016/j.jaut.2018.10.004.
16. Selmi C, Ceribelli A, Generali E, et al. Serum antinuclear and extractable nuclear antigen antibody prevalence and associated morbidity and mortality in the general population over 15 years. *Autoimmun Rev*. 2016;15:162–166. https://doi.org/10.1016/j.autrev.2015.10.007.
17. Hennes EM, Zeniya M, Czaja AJ, et al.; International Autoimmune Hepatitis Group. Simplified criteria for the diagnosis of autoimmune hepatitis. *Hepatol Baltim Md.* 2008;48:169–176. https://doi.org/10.1002/hep.22322.
18. Muratori L, Muratori P, Lanzoni G, et al. Application of the 2010 American Association for the Study of Liver Diseases criteria of remission to a cohort of Italian patients with autoimmune hepatitis. *Hepatol Baltim Md*. 2010;52:1857; author reply 1857–1858. https://doi.org/10.1002/hep.23924.
19. Beuers U, Gershwin ME, Gish RG, et al. Changing nomenclature for PBC: from "cirrhosis" to "cholangitis". *J Hepatol*. 2015;63:1285–1287. https://doi.org/10.1016/j.jhep.2015.06.031.
20. Lleo A, Colapietro F. Changes in the epidemiology of primary biliary cholangitis. *Clin Liver Dis.* 2018;22:429–441. https://doi.org/10.1016/j.cld.2018.03.001.
21. Tanaka A, Leung PSC, Gershwin ME. The genetics and epigenetics of primary biliary cholangitis. *Clin Liver Dis.* 2018;22:443–455. https://doi.org/10.1016/j.cld.2018.03.002.
22. Gershwin ME, Selmi C, Worman HJ, et al.; USA PBC Epidemiology Group. Risk factors and comorbidities in primary biliary cirrhosis: a controlled interview-based study of 1032 patients. *Hepatol Baltim Md.* 2005;42:1194–1202. https://doi.org/10.1002/hep.20907.
23. Tanaka T, Zhang W, Sun Y, et al. Autoreactive monoclonal antibodies from patients with primary biliary cholangitis recognize environmental xenobiotics. *Hepatol Baltim Md.* 2017;66(3):885–895. https://doi.org/10.1002/hep.29245.
24. Lindor KD, Bowlus CL, Boyer J, et al. Primary biliary cholangitis: 2018 practice guidance from the American Association for the Study of Liver Diseases. *Hepatol Baltim Md.*2018;69(1):394–419. https://doi.org/10.1002/hep.30145.
25. European Association for the Study of the Liver. EASL Clinical Practice Guidelines: The diagnosis and management of patients with primary biliary cholangitis. *J Hepatol.* 2017;67:145–172. https://doi.org/10.1016/j.jhep.2017.03.022.
26. Beuers U, Trauner M, Jansen P, et al. New paradigms in the treatment of hepatic cholestasis: from UDCA to FXR, PXR and beyond. *J Hepatol.* 2015;62:S25–S37. https://doi.org/10.1016/j.jhep.2015.02.023.
27. Murillo Perez CF, Harms MH, Lindor KD, et al. Goals of treatment for improved survival in primary biliary cholangitis: treatment target should be bilirubin within the normal range and normalization of alkaline phosphatase. *Am J Gastroenterol.* 2020;115(7):1066–1074. https://doi.org/10.14309/ajg.0000000000000557.
28. Corpechot C, Chazouillères O, Rousseau A, et al. A Placebo-controlled trial of bezafibrate in primary biliary cholangitis. *N Engl J Med*. 2018;378:2171–2181. https://doi.org/10.1056/NEJMoa1714519.
29. Chapman MH, Thorburn D, Hirschfield GM, et al. British Society of Gastroenterology and UK–PSC guidelines for the diagnosis and management of primary sclerosing cholangitis. *Gut*. 2019;68:1356–1378. https://doi.org/10.1136/gutjnl-2018-317993.
30. Dyson JK, Beuers U, Jones DEJ, et al. Primary sclerosing cholangitis. *Lancet Lond Engl.* 2018;391:2547–2559. https://doi.org/10.1016/S0140-6736(18)30300-3.
31. Liwinski T, Zenouzi R, John C, et al. Alterations of the bile microbiome in primary sclerosing cholangitis. *Gut*. 2020;69:665–672. https://doi.org/10.1136/gutjnl-2019-318416.
32. Weiler-Normann C, Lohse AW. Variant syndromes of autoimmune liver diseases: classification, diagnosis and management. *Dig Dis Basel Switz.* 2016;34:334–339. https://doi.org/10.1159/000444472.
33. Tanaka A. Immunoglobulin G4-related sclerosing cholangitis. *J Dig Dis*. 2019;20:357–362. https://doi.org/10.1111/1751-2980.12789.
34. Alvarez F, Berg PA, Bianchi FB, et al. International Autoimmune Hepatitis Group Report: review of criteria for diagnosis of autoimmune hepatitis. *J Hepatol*. 1999;31:929–938.

77

Lymphoid Leukemias

Sarah Elitzur, Shai Izraeli, Dina Ben-Yehuda, and Moshe E. Gatt

Leukemias are a group of hematologic malignant clonal diseases arising in bone marrow that present with differing clinical and laboratory features. The focus of this chapter is on acute lymphoblastic leukemia (ALL), which is the most common leukemia of lymphoid precursors, and on chronic lymphocytic leukemia (CLL), which is the most common leukemia of mature lymphoid cells in adults. A special emphasis is given to immunologic aspects of both diseases.

ACUTE LYMPHOBLASTIC LEUKEMIA

ALL manifests with the clonal proliferation and accumulation of malignant lymphoid progenitors. ALL can be viewed as a developmental disease of the lymphoid system because it often arises as a "developmental accident" during normal fetal lymphopoiesis. Studies of chromosomal translocations in ALL cells have identified key genes involved in normal lymphopoiesis and hematopoiesis. Conversely, basic studies of the development of the immune system and the immune receptors have provided important tools for the diagnosis and management of ALL. These achievements in basic and clinical research have led to the remarkable transformation of ALL from a uniformly fatal disease several decades ago to a disease that is curable in more than 85% of children. However, adults remain a challenge.[1]

Epidemiology and Etiology

ALL is the most common malignancy of childhood. In contrast, ALL accounts for less than 20% of leukemias in adults. In developed countries, the incidence of ALL peaks at 2 to 5 years of age. The low age peak is characteristic of affluent societies.[2]

Most ALLs are sporadic, with less than 5% associated with hereditary or constitutional syndromes. For example, children with Down syndrome have approximately 20-fold increased risk of ALL. Other disorders associated with increased risk are rare inherited genomic instability syndromes, such as ataxia-telangiectasia, Bloom, and Li-Fraumeni syndromes. Similarly, ALL is more common in patients with other congenital immunodeficiencies, such as X-linked agammaglobulinemia, immunoglobulin A (IgA) deficiency, and common variable immunodeficiency (CVID). Some cases of CVID associated with leukemia are caused by germline mutations in *IKZF1*, which encodes the lymphoid transcription factor Ikaros.[3] Germline mutations in additional hematopoietic and B-cell developmental genes, such as *ETV6, RUNX1*, and *PAX5*, also predispose to ALL.[4]

Studies of leukemia in identical twins have shed light on the etiology of childhood ALL. Although ALL is not hereditary, there is markedly increased risk of leukemias in identical twins. If leukemia occurs in one identical twin, the other twin generally has a 10% to 20% chance of developing the disease. This phenomenon has promoted the hypothesis that at least two genetic hits are required for the development of ALL[2] (Fig. 77.1). The first occurs during fetal lymphopoiesis and results in clonal proliferation of a preleukemic clone. Intrauterine metastasis of such a preleukemic clone from one twin to the other via their shared placental circulation is responsible for the concordant leukemia. Additional genetic hits in the preleukemic cells occur after birth and are required for the development of full-blown leukemia. The initial findings in identical twins with leukemia have been extended to sporadic ALL; in at least 70% of patients, the preleukemic clone can be detected molecularly in the neonatal blood samples collected after birth (known as Guthrie cards). More recently, careful molecular analysis of the cord blood of normal infants has demonstrated that the occurrence of a preleukemic clone carrying a leukemia-defining chromosomal translocation is relatively common. However, only 1% of children born with such a preleukemic clone will develop leukemia, implying the impracticality of a molecular screen for the early diagnosis of childhood ALL.

KEY CONCEPTS

Environmental Factors in the Epidemiology of Childhood Acute Lymphoblastic Leukemia (ALL): Roles for Infection and Immunity?

"Common" B-cell precursor ALL at the preschool age is the most common type of ALL in the suburban regions of affluent countries. The causes for this phenomenon are unknown. A popular hypothesis suggests a modified immune response to delayed infections during infancy.

The causes of the relatively rare postnatal leukemogenic genetic hits are unknown. Although environmental agents such as ionizing radiation and chemical mutagens have been implicated in the induction of ALL, almost all cases lack discernible etiologic factors. Because the risk of B-cell precursor ALL during early childhood is markedly increased by higher socioeconomic status and a suburban style of living in which the exposure of children to infectious pathogens is typically delayed beyond the neonatal period, Greaves hypothesized that many childhood cases are the consequence of an abnormally late immunologic response to common infections. One proposed mechanism is that growth inhibitory factors, such as interferon or transforming growth factor-β (TGF-β), secreted during this immune response provide a survival advantage to a preleukemic clone, setting the stage for additional leukemogenic mutations. A more direct suggestion for the involvement of infection (or the response to infection) in the pathogenesis of ALL has been provided; PAX5 heterozygous mice developed preB-cell ALL only after exposure to common mouse pathogens.[5]

Immunologic and Molecular Classification of ALL

Immunologic Classification

The subtypes of ALL are usually identified by their immunophenotype, which tends to resemble the lymphoid developmental stage in which the leukemic cell was arrested (Table 77.1).

B-Cell Precursor Leukemias

B-cell precursor ALLs are the most common childhood leukemias. **ProB-cell ALL** is characterized by expression of CD19 and CD34 without CD10. This is the most common leukemia of infants, which contains rearrangements of the *MLL*(*KMT2A*) gene on chromosome 11q23. It is associated with a poor outcome. Leukemic blast cells of **early preB-cell ALL** resemble normal B-lymphoid cell precursors. They express CD19, CD22, and CD79a. CD10 and terminal deoxynucleotidyl transferase (TdT) are detectable in 90% of cases. CD34 is detected in greater than 75% of cases. Early preB-cell ALL is the most prevalent type of ALL and is thus often called "common ALL."

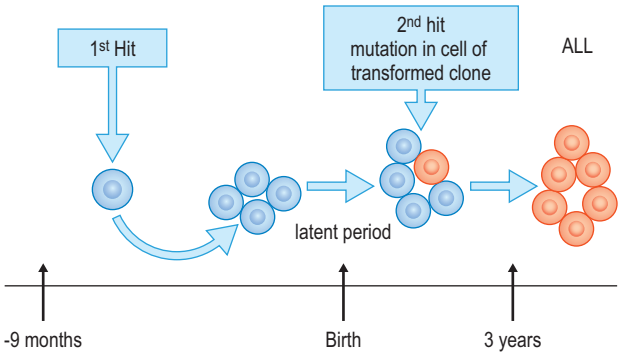


FIG. 77.1 A Model for the Development of Childhood Acute Lymphoblastic Leukemia (ALL). The first acquired genomic hit (e.g., chromosomal translocation or change in chromosomal copy number) occur during fetal hematopoiesis and results in clonal proliferation of a preleukemic clone. This event is common occurring in up to 1:20 children. Additional genetic aberrations occurring after birth are required for the development of ALL. These events are rare and are estimated to occur in approximately 1% of the children born with a preleukemic clone [Greaves, 2006 #1482].[5]

CLINICAL PEARLS

Mature Versus Precursor B-Cell Acute Lymphoblastic Leukemia

- Because the treatment of mature B-cell leukemia (the leukemic form of Burkitt lymphoma) is vastly different from the treatment of B-cell precursor ALL, it is critical to distinguish between the two.
- Burkitt leukemia is characterized by mature B-cell phenotype and by the presence of chromosomal translocations involving the *cMyc* gene (see Chapter 78).

Mature B-cell ALL is the leukemic form of Burkitt lymphoma (see Chapter 78). Because treatment is dramatically different from that for B-cell precursor ALL, this subtype must be specifically ruled out as part of the immunophenotypic evaluation of ALL. Mature B-cell ALL cells express surface Ig μ heavy chains in association with either κ or λ light chains.

T-Cell ALL

In affluent countries, T-ALL occurs in 10% to 15% of children with ALL. It is more prevalent in nonaffluent countries, probably as a reflection of a lower incidence of the common B-lineage early childhood peak. T-ALL is also more common in adults.

T-cell ALL cells express surface CD7 and cytoplasmic CD3 (cCD3) antigens on their surface; more than 90% of T lymphoblasts express CD2, CD5, and TdT. T-cell receptor (TCR) proteins are heterogeneously expressed in T-lineage ALL. In approximately two-thirds of cases, membrane CD3 and TCR proteins are absent. In half these cases, however, TCR proteins (TCRβ, TCRα, or both) are present in the cytoplasm. When membrane CD3 and TCR chains are expressed, the αβ form of the TCR predominates. Only a minority of cases express TCR γδ proteins.

A genetically and immunophenotypically distinct form of T-ALL originating from early T-cell precursors (ETPs) has been recognized. It is characterized by low or absent CD1a and CD5 coupled with the expression of at least one myeloid marker.[6]

Genetic and Molecular Classification

Virtually every leukemic cell contains acquired alterations in multiple genes. These alterations often manifested by gross numeric or structural aberrations that frequently define a specific clinical subtype of ALL. Common and/or clinically significant genetic aberrations that are typically found in ALL are summarized in Table 77.2.

TABLE 77.1 Immunophenotypic Classification of Acute Lymphoblastic Leukemia

	Leukocyte Antigen Expression (% of Cases Positive)										Frequency (%)	
Subtype	CD19	cCD22	CD79a	CD10	CD7	CD2	cCD3	clg μ	slg μ	slg κ/λ	Children	Adults
Pre-preB	100	>95[a]	>95	0	0	0	0	0	0	0	5	10
Early preB	100	>95[a]	>95	95	5	<5	0	0	0	0	60–65	50–55
PreB	100	100[a]	100	>95	0	0	0	100	0	0	20–25	10
Transitional preB	100	100[a]	100	50	0	0	0	100	100	0	1–3	?
B	100	100[a]	100	50	0	0	0	>95	>95	>95	2–3	4
PreT	<5	0	0–20	45	100	0	100	0	0	0	1	5
T	<5	0	0–20	45	100	95	100[a]	0	0	0	10–15	15–20

c, Cytoplasmic; *clg μ*, cytoplasmic immunoglobulin μ chain; *slg κ/λ*, surface immunoglobulin κ or λ chains; *slg μ*, surface immunoglobulin μ chain.
[a]Detectable on the cell-surface membrane in some cases.

TABLE 77.2 Frequencies of Major, Clinically Important Genetic Aberrations in Childhood and Adult Acute Lymphoblastic Leukemia

Genetic Aberration	Children	Adults
B-Cell Lineage		
Hyperdiploidy (>50 Chromosomes)	30%	9%
Hypodiploid (<45 chromosomes)	1%	2%
Amplified 21q	2%	2%
TEL-AML1 (t12;21)	25%	3%
MLL rearrangements	9%	13%
BCR-ABL	4%	33%
E2A-PBX1	5%	4%
"Ph like" including CRLF2	8%	25%
MYC rearrangements	2%	5%
T-Cell Lineage		
Notch1 mutations	60%	70%
TAL1 (SCL) cluster	58%	33%
HOX11 (TLX1) cluster	3%	33%
HOX11L2 (TLX3) cluster	20%	5%
LYL1 cluster	12%	37%
MLL-ENL	2%	2%
NUP214-ABL	6% (?)	

Numeric Chromosomal Aberrations

Deviation from the normal chromosomal modal number is called *aneuploidy* and is the most common chromosomal aberration in cancer. **High hyperdiploid ALL** (Fig. 77.2A), containing 50 to 60 chromosomes, is the most common type of B-lineage ALL in children and is associated with approximately 90% cure rate. Typically there is an excess of specific chromosomes, most commonly chromosomes 6, 10, 14, 17, 18, 21, and X. **Hypodiploid ALL** (see Fig. 77.2B), which contains less than 45 chromosomes, is much rarer and is associated with a very poor prognosis.

Structural Genetic Aberrations

Chromosomal translocations can be divided into two general subtypes. The first results from a translocation of an oncogene into the proximity of a strong regulatory region, resulting in its marked overexpression. Often these translocations are mediated by the V(D) J recombination machinery (see Chapter 4) and can be therefore viewed as an unfortunate developmental "accident" caused by the physiologic lymphocyte-specific genomic instability that is necessary for the creation of the diversity required to recognize novel antigens. Examples of these translocations include the activation of the *MYC* oncogene by t(8;14) translocation in Burkitt lymphoma and of the *SCL* (*TAL1*) gene by t(1;14) or by rearrangement with the *STIL* gene on chromosome 1p32 in T-cell ALL. The second type of translocation creates a novel fusion protein consisting of the genes that participate in the chromosomal translocation. The mechanisms that underlie these translocations remain unclear. Most of the translocations characteristic of B-cell precursor ALL are of this type, for example the *t*(12; 21) fusing of the *TEL* (*ETV6*) gene on chromosome 12 with the *AML1* (*RUNX1*) gene on chromosome 21.

Amplification and deletion of a small chromosomal region is another type of structural aberration that is often detected in leukemias. For example, deletions of the *INK4A* locus or *PAX5* are commonly detected in T- and B-cell precursor ALLs, respectively.

Oncogenic-activating mutations in ALL are reported with an increasing frequency. For example, the Notch pathway, which plays a role in T-cell development, is activated by acquired mutations in more than 60% of T-ALL.

Many of the genes modified by chromosomal translocations, amplification, deletions, or point mutations function in normal lymphoid or hematopoietic development (Table 77.3). The acquired aberrations promote malignant transformation by either overexpression or dysregulated expression (in the wrong cell or in the wrong developmental stage) of the developmental gene. For example, loss-of-function mutations in the IL7Rα receptor cause $T^{-}B^{+}NK^{+}$ severe combined immunodeficiency (see Chapter 34), and activating mutations are found in 10% of T-cell ALL.[7] Conversely, the acquired genetic aberration may block the normal developmental function of the involved gene(s). Good examples are genomic deletions or inactivating mutations in B-cell differentiation genes, such as *PAX5, EBF*, or *IKZF1*, detected in approximately 50% of B-lineage ALLs.

Molecular Subtypes of Clinical Relevance

(See Table 77.2.)

B-Lineage ALL

Hyperdiploid ALL and *TEL/AML1* gene translocation comprise the majority of "common ALL" leukemias typical of young children and are rare in adults with ALL. Both are associated with an extremely good prognosis. In contrast, the much less common hypodiploidy (less than 45 chromosomes) and the intrachromosomal amplification of chromosome 21 (iAMP21) are associated with a poor prognosis. The *MLL(KMT2A)* gene located on chromosome 11q23 is involved in fusion translocations with more than 80 different partner genes. The most common translocation in ALL fuses the *MLL* with the *AF4* gene on chromosome 4. It is characteristic of infant leukemia and is associated with a poor prognosis.

Another aberration with a poor prognosis is the B-cell receptor (BCR)-ABL1 fusion protein created by the *t*(9;22) "Philadelphia chromosome." Its frequency is low in children (3% to 5%) and much higher in adults (at least 25%). Outcome of patients with Philadelphia-ALL has dramatically improved after the addition of tyrosine kinase inhibitors to chemotherapeutic treatment.

A novel subgroup of ALLs has been termed "Philadelphia-like ALL" because of a gene expression profile that is similar to BCR-ABL1, while lacking the classic fusion sequence.[8,9] These leukemias are characterized by activation of either the ABL or the JAK kinase pathway and are caused by many types of fusion genes or aberrantly expressed receptors, such as ABL1 and ABL2, PDGFRB, CSF1R, EPOR, JAK2, and CRLF2. Their proper diagnosis is important because of the potential therapeutic effect of kinase inhibition.

Deletions of the *IKZF1* gene, encoding the B-cell transcriptional factor Ikaros, are common in both Philadelphia and in Philadelphia-like ALL and are associated with poor prognosis.

T-Lineage ALL

Although multiple genetic and molecular subtypes of T-cell ALL have been recently described, their clinical significance is presently unclear, except ETP-ALL, which has been associated with poor prognosis. Most of the genetic aberrations in T-cell ALL result in the abnormal expression of transcription factors. Examples include SCL (TAL1), which forms a complex with

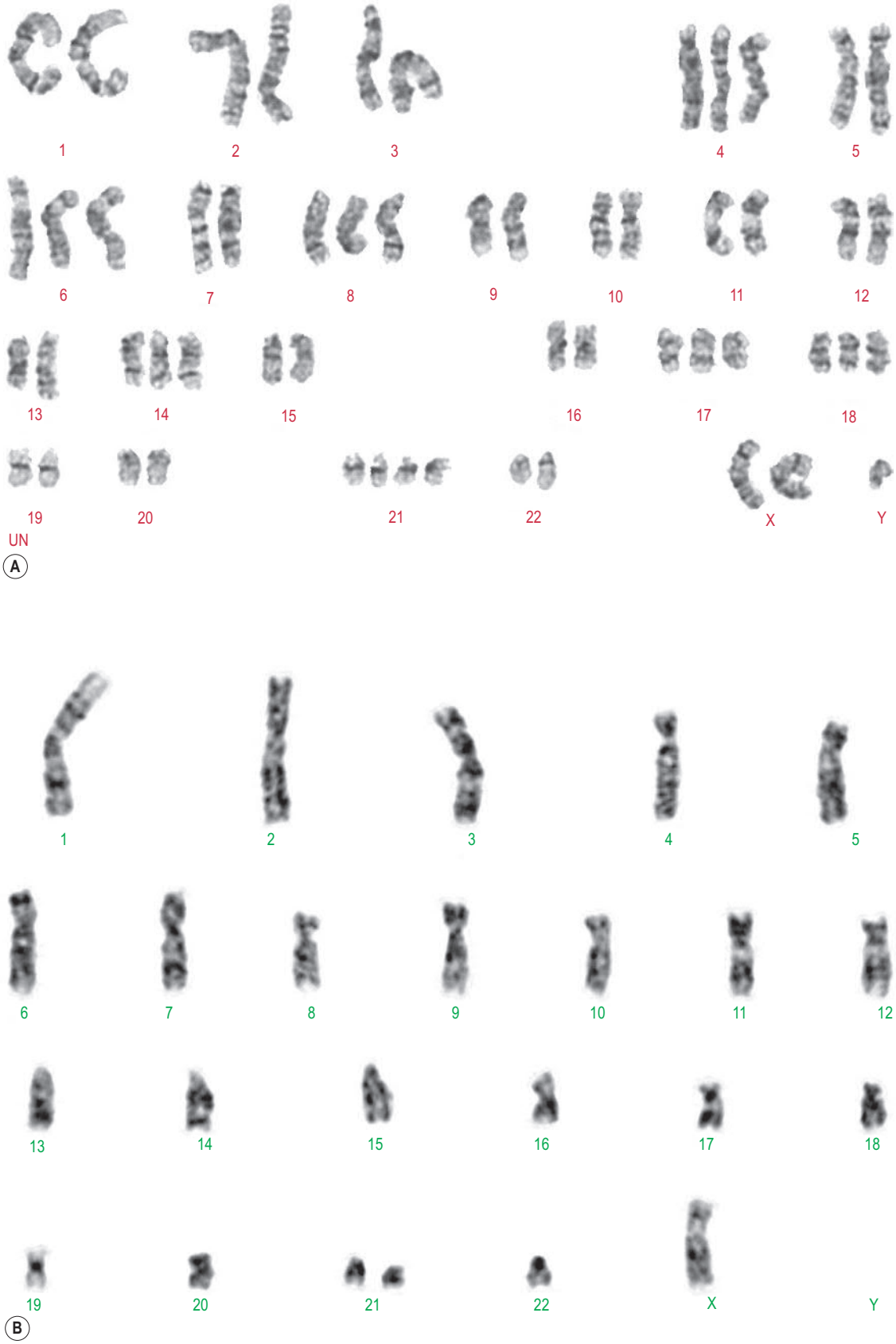


FIG. 77.2 Continued

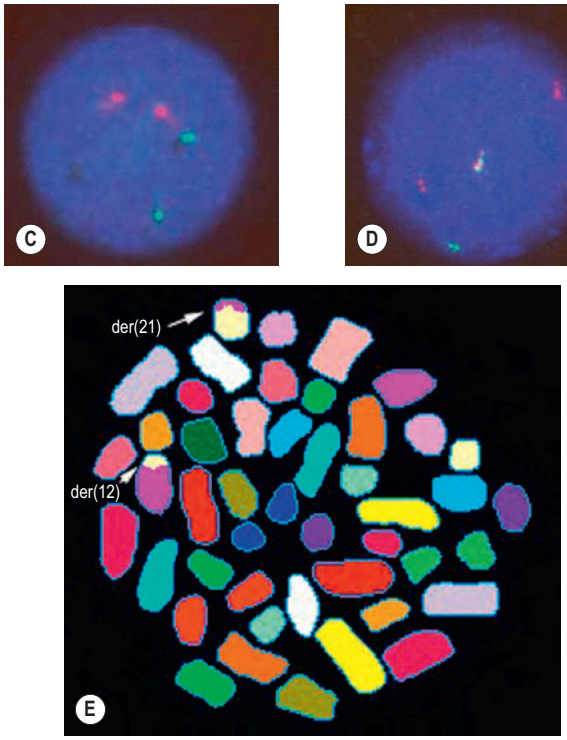


FIG. 77.2 Chromosomal Aberrations in Acute Lymphoblastic Leukemia (ALL). (A) A typical karyotype of hyperdiploid ALL. (B) A hypodiploid karyotype (pictures courtesy of B. Stark and D. Betts). Panels (C) and (D) display interphase fluorescence in situ hybridization with probes to the AML1 (RUNX1) gene on chromosome 21 *(red)* and to the TEL (ETV6) gene on chromosome 12 (green). (C) normal cell and (D) leukemic cell with fusion TEL-AML1 translocation *(arrow)*. Panel E display the same translocation on metaphase chromosomes using a molecular cytogenetics technique called Spectral Karyotyping *(arrows)*. Classical cytogenetics analysis often misses this translocation. (Courtesy of L. Trakhtenbrot).

TABLE 77.3 Example of Hematopoietic Genes Involved in the Pathogenesis of Leukemia

Gene(s) Names	Normal Hematopoietic Development	Leukemic Involvement
SCL (TAL1)	Hemangioblast specification Erythropoiesis and megakaryopoiesis	T-ALL
LMO1/2	Similar to SCL	T-cell ALL
NOTCH1	T lymphocytes	T-cell ALL
HOX11	Spleen	T-cell ALL
E2A	T and B lymphocytes	BCP-ALL
PAX5	B lymphocytes	BCP-ALL, B-NHL
SLP-65	B lymphocytes	BCP-ALL
TEL	Bone marrow hematopoietic stem cells	BCP-ALL, T-cell ALL rarely myeloid malignancies
RUNX1 (AML1, CBFA2)	Definite hematopoiesis Megakaryopoiesis and T lymphocytes	BCP-ALL, AML (M0-M1) Hereditary FPD/AML
CBFB	Same as RUNX1	AML (M4e)
C/EBP 1-3	Myeloid cells	AML (M1, M2)
PU.1	Myeloid and lymphoid stem cells	AML
GATA1	Erythropoiesis, megakaryopoiesis, and mast cells	AML (M7) associated with trisomy 21
FLT3	Hematopoiesis and lymphopoiesis	AML and ALL
MLL	Hematopoiesis stem cells	AML and ALL
IL7R	T lymphocytes	ALL

ALL, acute lymphoblastic leukemia; AML, acute myeloid leukemia; BCP-*ALL*, B-cell precursor acute lymphoblastic leukemia; B-*NHL*, B-cell non-Hodgkin lymphoma; *FPD*, familial platelet disorder; *T-ALL*, T-cell acute lymphoblastic leukemia

LMO1 or LMO2, HOX11L2 (TLX3) or HOX11 (TLX1), LYL1, and MYB. Other abnormalities include MLL-ENL fusion and amplification of the *ABL* oncogene. Cooperating mutations are activating mutations in NOTCH1 or interleukin-7R (IL-7R) and inactivating mutations in the E3 ligase FBW7 or the phosphatase PTEN. In several treatment protocols, T-cell ALL was considered to have a less favorable prognosis compared with B-cell precursor ALL, but outcome has improved considerably following therapy intensification through the past decades.

Clinical Features

The clinical signs and symptoms of leukemia relate to the replacement of bone marrow cells by leukemic blasts and to the infiltration of extramedullary sites. Pallor, fatigue, petechiae, bleeding, or fever may be caused by cytopenias. Bone pain and arthralgias, the onset of a limp and refusal to walk, and even frank arthritis are not uncommon. The musculoskeletal symptoms are sometimes confused with osteomyelitis or juvenile rheumatoid arthritis, which may delay diagnosis. Splenomegaly, hepatomegaly and lymphadenopathy are also common at presentation. Uncommonly, central nervous system (CNS) involvement may present as headache or cranial nerve palsies. Overt testicular leukemia manifests as painless testicular enlargement. Mediastinal involvement, common in T-cell ALL, may cause dyspnea and superior vena cava syndrome.

Clinical laboratory findings often include anemia and thrombocytopenia. Approximately 20% of children present with leukocyte counts greater than 50,000/μL. Importantly,

approximately 40% of children have leukocyte counts of less than 10,000/μL, and leukemic blasts may or may not be seen on peripheral smears. Therefore the diagnosis of leukemia may occasionally be missed in routine automated blood count. Elevated serum lactate dehydrogenase activity, hyperuricemia, and hyperphosphatemia are common in patients with a large leukemic cell burden.

The diagnosis of ALL is established by bone marrow examination. The normal bone marrow contains less than 5% blasts. A minimum of 25% lymphoblasts on differential examination of the bone marrow aspirate is necessary for the diagnosis of ALL. Most children with ALL have a hypercellular marrow with blasts constituting 60% to 100% of the nucleated cells.

Although overt CNS leukemia is relatively rare, submicroscopic CNS involvement is present at diagnosis in at least half the patients in the absence of any neurologic symptoms. Thus CNS-directed therapy is routinely included in ALL therapy.

CLINICAL PEARLS

Acute Lymphoblastic Leukemia and Rheumatoid Disorders

- ALL can mimic juvenile idiopathic arthritis (JIA; see Chapter 54) and other musculoskeletal disorders.
- Because leukemic blasts may be absent from the peripheral blood, bone marrow examination should be considered in any child with JIA, especially prior to commencing steroid therapy.

As many as 10% of children with ALL are first evaluated at pediatric rheumatology clinics. Fever, arthralgias, arthritis, or a limp accompanied by anemia, mild splenomegaly, and lymphadenopathy frequently can be confused with juvenile idiopathic arthritis (see Chapter 54) or osteomyelitis. These patients may be treated with antibiotics and antiinflammatory agents for several weeks to months before the diagnosis of ALL is finally made. Bone marrow examination should be seriously considered in such patients.

Special Diagnostic Tests

The classification and risk stratification of patients in clinical ALL trials are based on detailed immunophenotyping and genotyping analysis.[10] Classical cytogenetics by karyotyping may be performed to determine the chromosomal complement of the leukemic cells (see Fig. 77.2). In addition, clinically relevant structural and numeric chromosomal aberrations can be detected with the use of commercially available fluorescence *in situ* hybridization (FISH) probes (see Fig. 77.2). Sequencing-based assays that use DNA or RNA of sets of genes accurately detect mutations and rearrangements.

The elucidation of the human genome and the invention of the genomic technologies are in the process of revolutionizing the diagnostics of leukemias. New methodologies of next generation sequencing (NGS) are likely to transform both our understanding of leukemia biology and the diagnostic approach. Routine use of NGS mutation panels and copy number genomic arrays is likely to replace routine cytogenetic analysis in the near future.

Principles of Therapy[1]

ALL therapy consists several treatment phases. Typical remission-induction regimens include a glucocorticoid (prednisone or dexamethasone), vincristine, asparaginase, optional use of an anthracycline, and intrathecal chemotherapy. Rates of complete remission (CR) currently range from 97% to 99% in children and from 75% to 90% in adults. However, remission is not cure, and relapses will occur in the absence of additional therapy. Following remission, treatment includes several months of intensive combination chemotherapy, designed to consolidate remission and prevent systemic and CNS relapses. Repeated courses of methotrexate are an integral part of contemporary ALL therapy. Delayed intensification, or reinduction, is a subsequent phase of therapy that uses therapeutic agents similar to those administered during induction. Patients then receive prolonged low-intensity maintenance therapy with daily mercaptopurine and weekly methotrexate.

CNS-directed therapy is an integral part of ALL treatment. Cranial irradiation has dramatically improved cure rates among patients with ALL but is associated with significant deleterious long-term effects, including neurocognitive effects, endocrinopathies, and an increased risk of secondary CNS tumors. Consequently, CNS irradiation has been limited to smaller patient subgroups over time. Intrathecal chemotherapy and administration of methotrexate remain cornerstones of CNS prophylaxis. It is highly recommended that children and adults with ALL be treated in specialized centers as part of prospective clinical prospective studies. These clinical trials have led to the dramatic improvement in outcome that has been achieved over the past several decades. Optimal supportive care is required to mitigate the life-threatening toxicities of leukemia and its treatment.

Allogeneic hematopoietic stem cell transplantation (allo-HSCT) is generally reserved for relapsed or refractory leukemia and for patients with a very-high-risk leukemia that demonstrates a slow response to therapy or is highly likely to relapse.

Prognostic Factors

Modern therapy for ALL is based on adjustment of the intensity of therapy to the risk assessment of the relapse hazard (Table 77.4). Several clinical parameters of prognostic significance have been described, mainly age at diagnosis and leukocyte count. In children, age at presentation between 1 to 9 years and a leukocyte count less than 50×10^9/L are favorable prognostic factors. Females fare somewhat better than males. The prognostic significance of the major genetic aberrations has been described earlier.

TABLE 77.4 Major Prognostic Factors in Acute Lymphoblastic Leukemia[a]

Prognostic Factor	Good Prognosis	Worse Prognosis
Age at diagnosis	Age 1 year to <10 years (children)	<1 year; >10 year (children) >60 years (adult)
Peripheral blood WBC	<50,000 cells/μL	>100,000 cells/μL
Response to therapy	Early response to therapy Negative MRD at the end of induction	Slow response to therapy. High MRD
Genetic abnormalities	Hyperdiploidy (>50 chr.); *TEL/AML1 (ETV6/RUNX1)*	*BCR/ABL MLL/AF4* Hypodiploidy <45 chr.

chr., Chromosomes; *MRD*, minimal residual disease; *WBC*, white blood cells.
[a]The most important prognostic factor is the *treatment protocol*. Thus the prognostic significance of various clinical and laboratory variables may differ between protocols. Here, the significant parameters common to most studies are listed.

The most significant prognostic factor is the initial response to therapy. Rapid clearance of leukemic cells from blood or bone marrow confers a favorable prognosis. The level of minimal residual disease (MRD) after the induction of clinical remission has emerged as a powerful tool for gauging treatment response and predicting outcome.

Where Immunology Meets Oncology—Minimal Residual Disease

Modern treatment protocols have led to morphologic CR in the majority of patients, which is defined as less than 5% blasts in bone marrow examination. If treatment is discontinued at that stage, in most patients the disease will eventually relapse. Prospective clinical studies have shown that ALL should be treated for at least 2 years. These facts indicate that at the completion of remission induction not all clonogenic malignant lymphoblasts have been destroyed, even though most of the patients are in clinical and morphologic remission. Indeed, by this criterion, patients may have as many as 10^{10} undetectable neoplastic cells when in remission. Because, by definition, leukemic cells must constitute at least 1% to 5% of the nucleated cells in bone marrow to be detected by microscopic examination, morphologic examination is clearly inadequate for evaluation of the quality of remission in patients with ALL. Therefore more sensitive techniques for the detection of rare leukemic cells are required. This is the rationale behind the incorporation of modern techniques of MRD detection into treatment protocols of childhood ALL.

During the past two decades, two general methodologies have been developed for the sensitive detection of submicroscopic residual leukemic cells. These methodologies could not have been developed without elucidation of the scientific basis of the developmental phenotype of immune cells (see Chapters 7 and 9) and of the elaborate process of Ig gene rearrangements (see Chapter 4).

The most widely studied DNA-based MRD methodology is based on the identification of clonospecific rearrangements of Ig genes or TCRs (Ig/TCR-PCR).[11] This approach exploits the physiologic process of somatic rearrangement of Ig and TCR gene loci that occurs during the early differentiation of B cells and T cells. Thus any single T or B lymphocyte carries a unique rearrangement that is not shared by any other lymphoid cell. Because leukemia is clonal (i.e., it originates from one lymphoid cell), all of the leukemic cells of a particular person carry the same Ig and/or TCR rearrangements. Because leukemic cells are genetically unstable, they often (>90% of the cases) carry multiple rearrangements, a fact that facilitates the usefulness of using these rearrangements as a clonal marker for MRD detection. The major advantages of this technique are the exquisite sensitivity (at least 10^{-5}), reliability, reproducibility, and its applicability to greater than 90% of children with ALL. The application of NGS techniques lowers the costs and complexity of this approach.

Current strategies for flow cytometric detection of MRD rely on combinations of leukocyte markers that do not normally occur in cells of the peripheral blood and bone marrow. Such leukemia-associated phenotypes can be identified by multiple color staining techniques (Fig. 77.3). Flow cytometric analysis of these immunophenotypes allows the detection of one leukemic cell among 10^{-4} or more normal cells. The advantages of flow cytometry for MRD detection are adequate

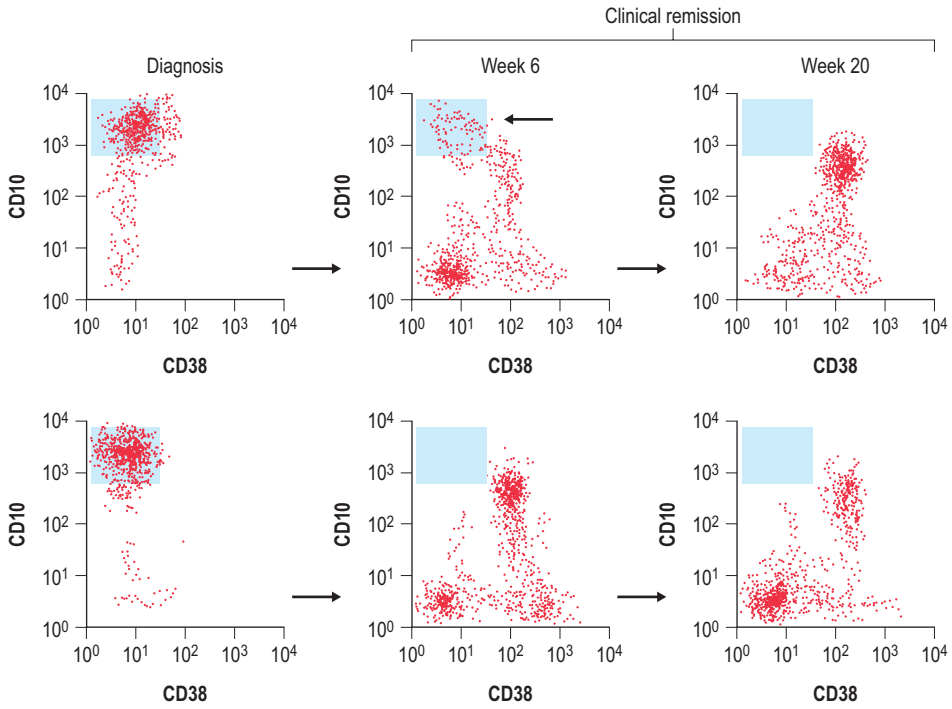


FIG. 77.3 Different Kinetics of Leukemia Cytoreduction Revealed by Minimal Residual Disease (MRD) Studies with Flow Cytometry. The left panels illustrate the leukemia-specific immunophenotype ($CD10^+$, $CD38^-$) determined at diagnosis in two children with acute lymphoblastic leukemia. This phenotype is not found in normal bone marrow. Bone marrow samples were collected during clinical remission from both patients. In one patient *(top panels)*, 0.04% of mononuclear cells expressed the leukemia-specific phenotype at week 6. MRD was undetectable by week 20. In the other patient *(bottom panels)*, a profound remission (MRD <0.01%) was achieved by week 6 and maintained at week 20.

sensitivity, and the presence of immunophenotyping facilities in most major centers that facilitate timely performance of the analysis on fresh cells at a reasonable cost.

Many ALL clinical studies have revealed strikingly similar results. Rapid clearance of leukemic blasts to less than 10^{-4} cells within the first 2 to 4 weeks of therapy is detected in approximately 40% of children with ALL and is associated with an extremely good prognosis. Conversely, the presence of greater than 0.1% blasts after 2 or 3 months of therapy defines a very-high-risk group. A prospective study involving 3184 children with ALL confirmed the strength of PCR MRD over genetic classification as a prognostic marker.[12] Currently, MRD studies have been incorporated into most treatment protocols. Patients with high MRD are stratified into a high-risk arm and receive more intensive chemotherapy.

Course and Prognosis

Current survival rates are approximately 90%. Between 15% to 20% of children will relapse. Cure rates after relapse are significantly lower. Prognostic factors include time to relapse (shorter time is worse), immunophenotype (T-cell immunophenotype is worse), and site of relapse (bone marrow disease is associated with a worse prognosis than extramedullary disease). MRD following relapse treatment is also a significant prognostic factor. Although some relapses can be treated with chemotherapy alone, many ultimately require HSCT. Genomic analysis has discovered marked subclonal heterogeneity in ALL.[13] Most of the relapses apparently arise from a minor subclone present at diagnosis. The identification and specific targeting of these resistant cells is a major future challenge.

Treatment Sequelae

Treatment-related mortality during induction and later, following remission, occurs in 2% to 4% of children with ALL, mostly of infectious causes. As cure rates improve, treatment-related mortality accounts for a higher percentage of total deaths. Certain subgroups are at higher risk for chemotherapy-associated death, such as infants, adolescents and young adults, patients with certain genetic predisposition syndromes, and those who receive more intensive therapy.

Many survivors of childhood ALL do not develop significant long-term effects. Factors that determine a child's risk for late effects include the type of treatment administered, age at leukemia treatment, and various host pharmacogenomic factors. A significant late effect of contemporary therapy is osteonecrosis, which occurs in 5% to 10% of patients, more commonly in adolescents and females. Osteonecrosis may be severely debilitating. It often involves major joints and may lead to joint collapse, requiring total joint replacement. Other long-term effects include obesity and the metabolic syndrome, secondary malignancies, neurocognitive impairments and, rarely, cardiotoxicity. Long-term follow-up to monitor for possible late effects is essential. Improvements in treatment have resulted in minimizing certain adverse effects, but further research is needed concerning the long-term effects of contemporary therapy.[14]

Where Immunology Meets Oncology—Immunotherapy of B-linage acute lymphoblastic leukemia (B-cell precursor ALL)

Despite tremendous progress, ALL remains a significant cause of childhood cancer-related mortality, due to both disease relapse and treatment-related toxicity. Acute adverse effects of chemotherapy may be life-threatening, and long-term side effects often impair survivors' quality of life. Treatment outcomes are still unsatisfactory for certain patient subgroups, such as infants, adolescents, young adults, and those with unfavorable genomic alterations. Survival rates in patients with relapsed or refractory ALL remain poor, especially in patients with early bone marrow relapse, in patients with multiple relapses, and in those who relapse after prior allo-HSCT. Certain subgroups, such as patients with Philadelphia ALL or those with "Ph-like" ALL, may benefit from molecularly targeted agents, but the majority lack a suitable target. This led to the introduction of novel immune-based therapies into the treatment of relapsed/refractory ALL (Fig. 77.4). These modalities gained US Food and Drug Administration (FDA) approval after demonstrating improved remission rates compared with chemotherapy or showing long-term benefits compared with historical controls.

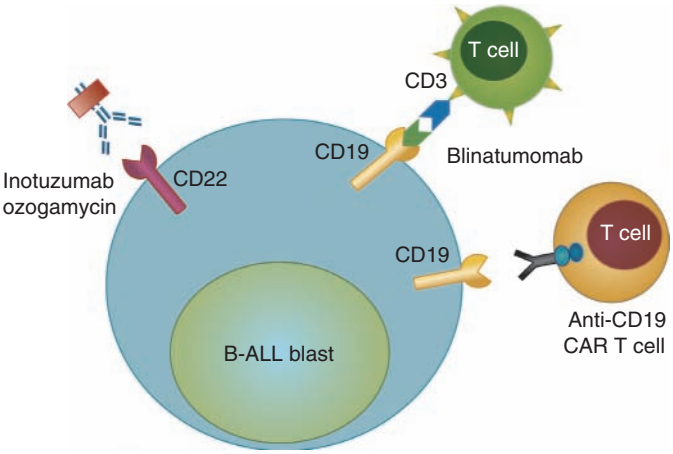


FIG. 77.4 Immunotherapy in B-ALL. Immunotherapeutic approaches include the anti-CD22 antibody-drug conjugate inotuzumab ozogamicin, the bispecific T-cell engager anti-CD19 antibody blinatumomab and chimeric antigen receptor T cells.

Inotuzumab Ozogamicin

CD22 is expressed on more than 90% of B-ALL cells and therefore serves as an attractive target for immunotherapy. Inotuzumab is a monoclonal antibody to CD22 linked to calicheamicin, a cytotoxic agent. Upon binding to CD22, the complex is rapidly internalized. Calicheamicin is released in the acidic environment of the lysosomes, with subsequent DNA damage and cell death. In adults with relapsed/refractory B-ALL, treatment with inotuzumab resulted in higher rates of remission and MRD negativity and better overall survival and quality of life, compared with chemotherapy.[15]

Inotuzumab has been shown to be effective also in combination with chemotherapeutic agents. Inotuzumab has been shown to be safe and effective in the pediatric population.

Higher rates of hepatotoxicity have been reported following inotuzumab therapy, mainly due to the development of sinusoidal obstruction syndrome in patients who subsequently underwent HSCT.

Blinatumomab

Blinatumomab is a bispecific T-cell engager (BiTE) construct designed to direct cytotoxic T cells towards CD19, an early B-lineage restricted antigen expressed on almost all leukemic cells in patients with B-ALL. Blinatumomab consists of two

connected single-chain variable antibody fragments: one binds to CD19 expressed on B cells, and the other to the TCR/CD3 complex, thus forming an immunologic complex which triggers apoptosis of the CD19-positive leukemia cell. Blinatumomab therapy has resulted in significantly higher remission rates compared with standard-of-care chemotherapy in adult ALL patients, along with significantly longer overall survival.[16] Single agent blinatumomab effectively eradicated MRD, leading to improved survival in responders compared with MRD nonresponders. Blinatumomab is safe and effective in the pediatric population. In pediatric patients in first ALL relapse, blinatumomab proved superior to intensive chemotherapy. Patients who received blinatumomab were more likely to attain MRD negativity and to proceed to HSCT, with considerably less toxicity and an improved event-free survival.

Following the success of blinatumomab treatment in the relapsed/refractory B-ALL setting, current pediatric clinical trials (e.g., the international AIEOP-BFM ALL 2017 study, *NCT03643276*) are incorporating blinatumomab into the first line treatment of B-ALL treatment, as an alternative to highly intensive and toxic chemotherapy for high-risk subgroups, and in addition to conventional chemotherapy for medium-risk patients, in an effort to reduce relapse rates.

Chimeric Antigen Receptor T Cells

Chimeric antigen receptor T cells (CARs) are a form of adoptive immunotherapy designed to modify T cells to recognize specific proteins expressed by cancer cells. Most CAR products use autologous T cells, which are harvested by leukapheresis and then reengineered to contain a receptor protein that combines antigen-binding and T-cell activating functions. CARs are then expanded for clinical use and infused back to the patient. Upon binding the target cell, the linked T-cell signaling domain activates the cytotoxic machinery, subsequently causing the death of the antigen-expressing cell. Most CARs developed for clinical trials include a primary CD3-ζ signaling domain and a CD28 or 4-1BB secondary costimulatory domain. The majority of CARs designed for clinical development have targeted the B-cell associated CD19 surface antigen and were used in B-cell malignancies, including B-cell leukemias and lymphomas in adults and children. Dramatic clinical responses and high remission rates have been observed as a result of CAR therapy of B-ALL.[17]

Unique and significant toxicities are associated with blinatumomab and CAR therapy, mainly cytokine release storm after cytotoxic T-cell activation, which can be life threatening, and neurotoxicity of varying severity. Hypogammaglobulinemia develops as a result of B-cell aplasia, and monitoring of immunoglobulin levels is recommended.

Current and future advances in immunotherapy have the potential to transform childhood ALL treatment. Yet growing experience with these treatment modalities has revealed that several mechanisms of resistance may arise. Major causes of treatment failure are loss of T-cell persistence following CAR therapy and loss of target-antigen expression on leukemic cells. Further research will reveal the precise role of these agents in pediatric ALL therapy and strategies will need to be developed to overcome these obstacles.

Future Perspectives

The high cure rates in patients with ALL (nearly 90% of children; 40% of adults) attest to the steady progress that has been made in treating this disease. A further increase in cure rates will require efforts to maximize the efficacy and minimize the toxicity of current therapy. Advanced genomic technologies carry the promise of discovering the full spectrum of leukemogenic pathways and the identification of targets for new therapies. An exciting development is the advent of immunotherapy. The role of immunotherapy in the treatment of both newly diagnosed and relapsed patients with ALL will be more precisely defined in the near future.

ON THE HORIZON

Translational Research of Acute Lymphoblastic Leukemia

- Whole genomic analyses will discover all leukemia genetic abnormalities, enabling further personalization of treatment.
- Novel therapies targeting specific leukemia-associated abnormalities will be implemented.
- Immunotherapy utilizing antibodies directed against lymphoid antigens and conjugated either to toxins or to T-cell engaging molecules or modified T cells will be routinely implemented in the treatment of ALL.

CHRONIC LYMPHOCYTIC LEUKEMIA

CLL is an often-indolent lymphoproliferative neoplasm of mature peripheral circulating B cells. It is the most common leukemia in adults living in countries in the Western hemisphere. CLL originates from a clonal lymphoid evolved mature stem cell that can be identified by its distinct B-cell Ig gene rearrangement. Both clinically and at the molecular level, it is a heterogeneous disease. Some patients have an indolent course, whereas other have a more rapid and aggressive disease. During its progression, CLL may be associated with significant immune deficiencies and autoimmune phenomena that can complicate its course and treatment. These abnormalities may be profound and thus alter the nature of the disease. They are attributed to the clonal nature of its B-cell origin. Less than 5% of patients have T-cell CLL. Understanding the molecular pathways involving BCR moieties enables focusing of novel and targeted chemoimmunotherapy.

Epidemiology

The extensive use of automated peripheral blood lymphocyte counts has increased the rate of diagnosis of asymptomatic patients with CLL. The incidence rate of CLL increases logarithmically from age 35 years, with a median age at the time of diagnosis of 65 years. There is a male predilection, and the disease appears to have geographic and ethnic variations in incidence. In the United States, CLL is uncommon in people of Asian descent. Because many of these patients may never require tissue diagnosis or inpatient treatment, cases among them are not likely to be recorded in a tumor registry, thus making the true annual incidence of the disease higher than previously thought (6.8 per 100,000 population).[18] With sensitive techniques, a monoclonal population of B lymphocytes that is indistinguishable in immunophenotype from CLL cells may be found in the blood of 3.5% of persons older than 40 years of age.[19]

The presence of B lymphocytes less than 5000/μL showing clonality is defined as monoclonal B-lymphocytosis (MBL). MBL is an indolent disorder that may progress to frank CLL at a rate of 1% to 2% per year, and almost all patients with CLL

begin with MBL. This state is divided into low or high MBL (less than or greater than 500/μL clonal lymphocytes in the peripheral blood, respectively). Low MBL is more abundant and rarely progresses to frank CLL, whereas high MBL may progress at the aforementioned rate. Some patients with MBL will experience autoimmune phenomena, as later described for CLL. However, the importance of the MBL clone is not fully understood and does not necessarily imply a future progression to CLL.[20]

The etiology of CLL is still unknown. However, as with other forms of malignancy, there is increasing evidence for the role of inherited factors in its development. Family surveys show a genetic predisposition in first-degree relatives, who also demonstrate increased susceptibility to other lymphoproliferative disorders, including other lymphomas.[21] Anticipation, the phenomenon of earlier onset, and a more severe phenotype in successive generations have been reported in families of patients with CLL.

Pathogenesis and the Biology of Leukemic Lymphocytes

CLL is now viewed as two related entities, both of which originate from B lymphocytes but differ in their activation state and in the maturation state of the cellular subgroup.[19.]

Normal B lymphocytes mature in bone marrow (see Chapter 7). In the process, they undergo rearrangement of Ig V(D)J gene segments to create the code for an Ig molecule that serves as the BCR for antigen (see Chapter 4). When an antigen of adequate affinity engages the receptor, the cell enters a germinal center located in a lymphoid follicle. There, as a centroblast, it rapidly divides and the V domains of its Ig undergo somatic hypermutation. Cells with receptors that have enhanced antigen-binding affinity proliferate in the presence of the antigen, whereas centrocytes with receptors that no longer bind the antigen (or bind autoantigens) are normally eliminated. Once the centrocytes are selected, they become plasma or memory B cells (Fig. 77.5).

The sequence and structure of Igs expressed by CLL cells from various patients have been found to be similar, suggesting a common pathogenetic antigen, such as an autoantigen-provoking clonal expansion. These findings suggest that antigenic stimulus plays a role in the promotion of CLL proliferation.

CLL may originate from a clone with few or no V domain mutations or from a more mature clone where these domains are hypermutated. These differences in the extent of V domain mutation suggest differing entities with two different developmental histories. Both originate from antigen-stimulated mature B lymphocytes. However, in the CLL with mutated Ig V domains, the proliferating B cell may have traversed the germinal centers, whereas in CLL with unmutated Ig V domains, the malignant B cell may derive from a naïve, pre–germinal center B cell.

The mutational status of the V domains strongly correlates with prognosis in that patients with an unmutated clone have a much worse prognosis compared with patients with mutated clones. These patients may differ also in their association with specific genetic aberrations. The 11q22-23 (the ataxia-telangiectasia mutated gene) or 17p13 (the *p53* gene) deletions are associated with poor outcome and with an unmutated V domain profile.

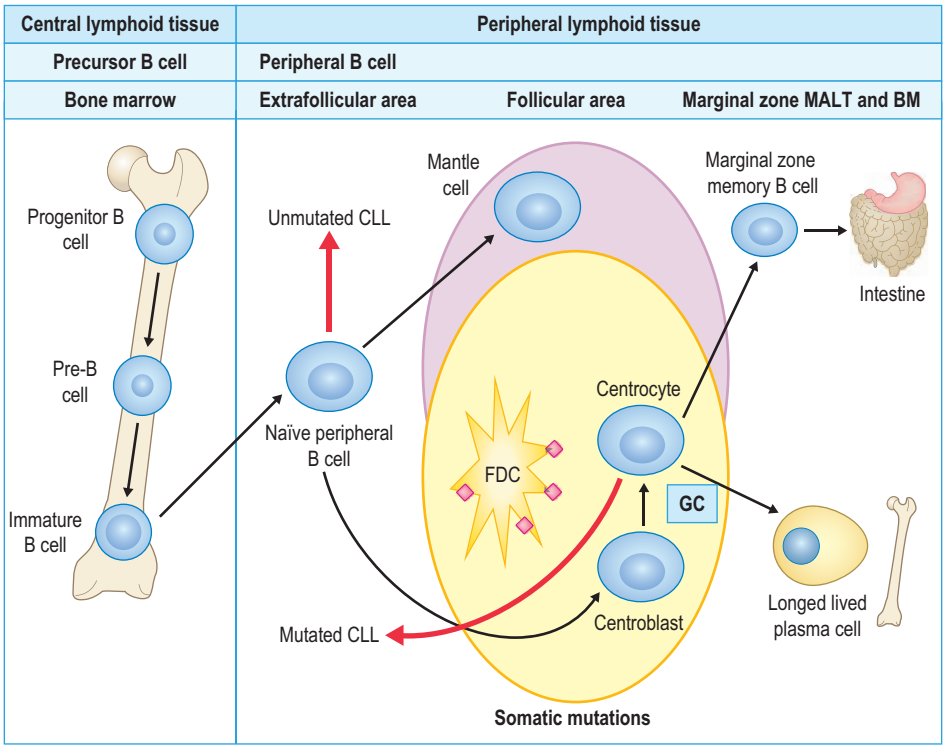


FIG. 77.5 **B-Cell Lymphoproliferative Disorders.** These diseases are related to different stages of normal B cell development. Maturation starts at the bone marrow. The more mature cells exit to the peripheral blood but do not experience the somatic hypermutation in the germinal center (GC), hence giving rise to the unmutated CLL form. When the B cells mature and differentiate within the GC into the hypermutated centrocytes, they will pathologically become the mutated form of CLL. Other forms of B cell neoplasms may arise at other stages. Earlier, as pro- and pre-B cells to give rise to acute lymphoblastic leukemia; intermediate stages to mantle zone lymphoma; late and more mature stages to marginal zone lymphoma and plasma cell–derived multiple myeloma.

These genes regulate apoptosis and resistance to chemotherapy. The 13q14 deletion or a normal karyotype is associated with a mutated profile and better prognosis. These chromosomal aberrations have independent prognostic significance (unrelated to the mutational status).

Surface membrane antigens typically found on B-cell CLL cells include CD19, CD21, and CD23. Expression of membrane IgM, IgD, and CD79b is reduced and is thus consistent with a phenotype matching that of mature, activated B lymphocytes. The pathologic features of biopsy specimens of lymph nodes are those of a small lymphocytic lymphoma (SLL).[19] The co-expression of CD5, a T cell–associated antigen, is a phenotypic characteristic and part of the disease defining criteria. $CD5^+$ B cells can be found in the peripheral blood of normal adults, suggesting that specific subsets of $CD5^+$ B cells from the mantle zone may be the normal counterparts of B-cell CLL. The low expression of the BCR is the hallmark of CLL cells and contributes to impairments in the activation of the cell following BCR stimulation.

BCR signaling in CLL has been extensively studied and has shed considerable light on the pathogenesis of the disease, thus aiding in the design of targeted therapies. The BCR is a multimeric complex containing the antigen-specific surface Ig and the two membrane-bound signal transduction elements CD79A and CD79B. Binding of antigen to the BCR induces activation of upstream kinases, which, in turn, activate other kinases by means of their cytoplasmic moieties, including SYK and the Src kinase. These kinases further activate the Bruton tyrosine kinase (BTK), PI3K, and other downstream pathways, including phospholipase C gamma 2 (PLC-γ2), calcium signaling, protein kinase C (PKC), nuclear factor (NF)-κB signaling, mitogen-activated protein kinases (MAPKs), and nuclear transcription.[22] The upstream kinases BTK and PI3K are currently targeted by specific agents, and other kinases and pathways are being explored for novel therapies (Fig. 77.6).

The CD38 surface molecule supports B-cell interactions and differentiation. Under certain circumstances, CD38 also augments signaling of BCR, delivering signals that regulate the apoptosis of B cells. Expression of CD38 correlates with expression of unmutated V domains and suggests a bad prognosis. Another molecule that influences the BCR is the ζ-associated protein 70 (ZAP70). High levels of this receptor associated protein tyrosine kinase (usually found in T and natural killer [NK] cells but not in normal B cells) are detected in the majority of unmutated CLLs and correlate with a poor prognosis (Fig. 77.7).

The microenvironment may play a role in the pathogenesis of CLL. Interactions with stromal cells rescue CLL cells from apoptosis *in vitro*. Activated T cells support the growth of CLL cells; cytokines, such as IL-4 and vascular endothelial growth factor (VEGF), and chemokines, such as CXCL12, support the expansion of CLL clones. When compared with blood-derived cells, expression profiling of CLL cells in the patient's lymph nodes showed BCR and nuclear factor (NF)-κB activation promoting cell proliferation, thus indicating the strong effects of the tumor microenvironment.[22] Mesenchymal stromal cells are also commonly found in secondary lymphatic tissues of patients with CLL. There they provide survival and migration signals to CLL cells, as well as protection from apoptosis and modulation of antigen presentation.[22,23] Endothelial cell and follicular dendritic cells (FDCs) may also play a role.

Interaction with and evasion from the normal immune system has been shown to be of significance.[22] The number of circulating T cells, oligoclonal in both the CD4 and the CD8 compartments, is increased. The function of these T cells is impaired, and they express exhaustion markers, including programmed death protein 1 (PD-1), at higher levels. Accordingly, CLL cells express high levels of PD-1 ligand (PD-L1). In addition, NK cells have reduced effector activities[22] that are associated with low expression levels of the activating receptors NK-cell p30-related protein (NKp30) and NK group 2 member D (NKG2D). Together, these findings provide an explanation for ability of CLL cells to evade immune-mediated destruction.

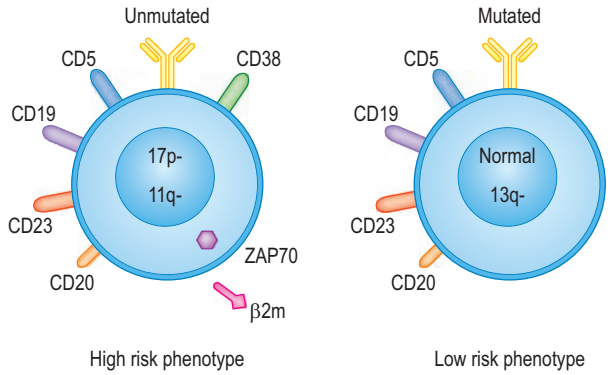


FIG. 77.6 Risk Markers and Stratification in Chronic Lymphocytic Leukemia (CLL). CLL is one disease with two different entities. It commonly expresses the B-cell surface antigens CD19 and CD23, low CD20, with the co-expression of the pan T-cell antigen CD5. The unmutated genotype correlates to worse prognosis, featuring positive CD38 surface antigen, intracellular ZAP 70, chromosomal aberrations as 17p– and 11q–, and high levels of soluble β2-microglobulin. This as opposed to the better prognosis associated with the mutated form of CLL, lacking these phenotypic features, and no chromosomal aberrations or the 13q–.

CLL has been classically characterized by the accumulation of mature B cells that evade apoptosis, with high levels of the BCL2 antiapoptotic protein. Contradicting this dogma is the measurement of CLL kinetics that has shown CLL cells to proliferate at a high dynamic rate of up to 1% of the clone per day.[24] In addition, proliferation-related genes (e.g., *c-MYC* and *E2F1*) have been found to be upregulated, especially in the unmutated CLLs.[23] This finding suggests that CLL is not solely an accumulative disease but also has a proliferative element.

Clinical Features of CLL

CLINICAL PEARLS

Chronic Lymphocytic Leukemia: Clinical Manifestations

- Absolute blood lymphocytosis >5000 /mm^3 sustained over a period of 4 weeks (to exclude transient lymphocytosis related conditions as viral infections)
- At least 30% lymphocytes in a normocellular or hypercellular marrow
- Phenotypically monoclonal lymphocytes that express mature B-cell markers (e.g., CD19) and CD5
- Morphologically mature-appearing lymphocytes
- Most patients have some degree of lymphadenopathy on physical examination
- Progression and the need for treatment depend on lymphocyte doubling time, bulky symptomatic disease, anemia, thrombocytopenia, autoimmune phenomena

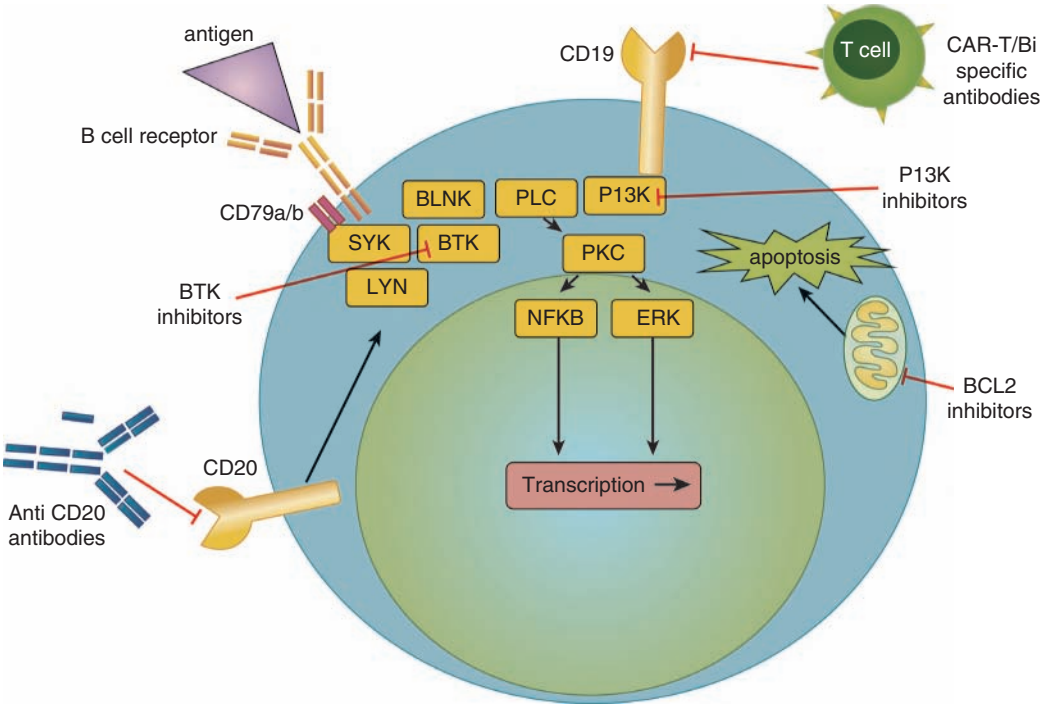


FIG. 77.7 Chronic Lymphocytic Leukemia (CLL) Signaling and Targeted Therapy. B-cell receptor (BCR) signaling in CLL has been extensively studied and helps the design of targeted therapies. Antigen binding by the BCR induces the formation of a signaling complex that is initiated by the cytoplasmic tails of CD79A and CD79B, further recruiting spleen tyrosine kinase, which is followed by the activation of B cell linker protein, phospholipase Cγ2, and Bruton tyrosine kinase (BTK). Inhibition of BTK by tyrosine kinase inhibitor ibrutininb and second-generation acalabrutinib and zanubrutinib interferes with this signaling and causes cell death. The tyrosine-protein kinase LYN also directly phosphorylates BTK. BCR signaling via the PI3K pathway is promoted further, and the positive BCR co-receptor CD19 contributes to the activation of the PI3K pathway and to the induction of survival. CD19 is also the target of new antibodies and the main target of activated chimeric antigen receptor T cells. Further downstream signaling of the pathway are responses of protein kinase C, the RAS-MAPK pathway, and nuclear factor-κB. The signaling response induces transcription. Other activators of cellular activity of survival like CD20, which is inhibited by first- (rituximab, ofatumumab) and second- (obinutuzumab) generation anti-CD20 antibodies. Promotion of apoptotic pathways are done via the BCL2 inhibitor venetoclax.

The clinical diagnosis of CLL requires an absolute lymphocytosis with a threshold of greater than 5000×10^9/L mature-appearing lymphocytes in the blood smear, persistence of the lymphocytosis for greater than 4 weeks, and a distinct immunophenotype (as described earlier). Approximately a quarter of patients with CLL are asymptomatic at diagnosis. The clinical characteristics at presentation include lymphadenopathy (87%), splenomegaly (54%), hepatomegaly (14%), high white blood cell (WBC) count, anemia, and thrombocytopenia (20%). Very high WBC counts are rare but may develop along the course of the disease. Hyperleukocytosis, which causes leukostasis and necessitates emergency treatment, is extremely rare. Other organs that are involved include other lymphoid tissues and rarely solid organs or skin.

A prognostic evaluation of the patient with CLL begins with study of the patient's blood and bone marrow. Lymph node biopsy is not necessary but, if performed, may reveal the diagnosis of SLL, which is considered a different manifestation of the same disease. The differential diagnosis includes other low-grade lymphoproliferative disorders, such as a leukemic phase of lymphoma and mantle cell lymphoma (usually negative for CD23); hairy cell leukemia (CD5 and CD21 negative and CD103 and CD25 positive); T-cell leukemia (other T-cell markers, such as CD3, CD4, and CD7); and prolymphocytic leukemia (PLL), which is distinguished by morphologically immature-appearing cells and the presence of FMC7 and CD79b on the cell surface. T-cell CLL is rare (<5%).

The oldest staging systems of CLL risk stratification rely on measurement of disease bulk as reported by Rai[25] and Binet[26,27] (Table 77.5). These staging systems are very useful for identifying patients who will need treatment at the time of diagnosis, but not for predicting who will subsequently need treatment. Approximately 33% of patients never require treatment and have a long survival. In another third, an initial indolent phase is followed by disease progression. The remaining third of patients exhibit an aggressive disease at the onset and need immediate treatment.

The most powerful predictors for rapid and aggressive progression are the mutational status of the V domains and chromosomal abnormalities, as described earlier (see Fig. 77.7; see Table 77.5). In contrast to genomic aberrations and serum markers, such as CD38, the mutational profile has the advantage of remaining constant during disease evolution. NGS has identified new genetic markers that have altered prognosis for patients with CLL.[27] *TP53, ATM, NOTCH1*, and *SF3B1* mutations indicate a worse prognosis.

TABLE 77.5 Major Factors Associated With Prognosis of Chronic Lymphocytic Leukemia (CLL)[4–7,9,30,31,33,34]

	Definition	Median Survival (Years)	Used in Clinical Practice
Rai stage 0	Leukocytosis	12.5	Yes
1	Leukocytosis and lymphadenopathy	8.4	
2	Lymphocytosis plus hepatosplenomegaly	6.9	
3	Lymphocytosis plus anemia (<11 g/dL)	1.5	
4	Lymphocytosis plus thrombocytopenia (<100 000 × 10^9/L)	1.5	
Binet	Leukocytosis and lymphadenopathy	Age-matched	Yes
Stage A	Lymphadenopathy of more than two involved areas	7	
Stage B	Anemia or thrombocytopenia	2	
Stage C			
β_2 microglobulin	Normal	9.7	Yes
	Elevated	4.5	
CD 38	>30%	2.9 to <10 years	Yes
Early-stage CLL	<30%	9 to >26 years	
Chromosomal aberrations (FISH analysis)	17p–	2.7	Sometimes
	11q–	6.5	
	Trisomy 12	9.5	
	Normal karyotype	9.3	
	13q–	11.1	
Zeta-associated protein 70 (ZAP-70)	>20%	7.5	If available
Early-stage CLL	<20%	Not reached	
Mutational status	Unmutated	5.7 to <9.9 years	If available
	Mutated	10.2 to >24 years	

FISH, Fluorescence in situ hybridization.

Treatment

THERAPEUTIC PRINCIPLES

Chronic Lymphocytic Leukemia Is Incurable, but It Is Possible to Ameliorate Symptoms

- High rates of complete remissions and encouraging survival curves for high-risk patients
- Watch and wait approach in the case of asymptomatic patients with early-stage CLL
- Supportive care (immunoglobulins, blood supplements and erythropoietin, treatment of infections)
- Steroids with or without chemotherapy, especially for autoimmune phenomena
- Conventional chemotherapy (alkylating agents, such as chlorambucil and cyclophosphamide; purine analogues, such as fludarabine and cladribine; other lymphoma regimens; bendamustine)
- Targeted therapy with or without chemotherapy (i.e., rituximab, ofatumumab, obinutuzumab (all anti-CD20 antibodies) and alemtuzumab (anti-CD52 antibody)
- B-cell receptor pathway inhibitors—BTK (ibrutinib), PI3K-d (idelalisib), and BCL2 antagonists (venetoclax)
- Allogeneic (mostly with reduced intensity) stem cell transplantation

Because CLL remains an incurable tumor, treatment may be delayed and the patient monitored until becoming symptomatic (see Table 77.4). Indications for therapeutic intervention include the development of symptoms, a worsening anemia and/or thrombocytopenia, autoimmune cytopenias, progressive splenomegaly, progressive lymphadenopathy, or a lymphocyte doubling time of 6 months or less. No prospective data yet exist to support the early treatment of asymptomatic patients with adverse prognostic features. However, this group warrants close monitoring.

Conventional chemotherapy was the mainstay of treatment. Novel targeted therapy gradually replaces this approach to the verge of "chemo-free" treatment paradigm.[28] Chlorambucil, alone or combined with corticosteroids, has been the most commonly used drug. It is advantageous in relieving symptoms, even in advanced disease. However, several randomized controlled trials have failed to demonstrate improved survival, and hardly any patient achieves CR. Purine analogues (most commonly fludarabine) with or without cyclophosphamide have been shown to induce higher response rates, with some patients achieving CR. Bendamustine is a novel agent that contains both alkylating properties and a purine-like benzimidazole ring. It has been shown to be effective in CLL and less toxic than the fludarabine and cyclophosphamide combination.[29]

The anti-CD20 mAb rituximab as a single agent has shown limited efficacy in CLL, possibly because of the weak receptor expression on CLL cells. However, in combination with chemotherapy, particularly with fludarabine and cyclophosphamide or bendamustine,[29] and to other novel targeted agents,[28] discussed later, it appears to act synergistically and to achieve high rates of response, including molecular CR and prolongation of disease-free and overall survival. A newer fully human mAb targeting CD20 (ofatumumab) has a very potent effect as a single agent in both rituximab-naïve and rituximab-treated patients. It targets a different epitope on the CD20 antigen. A second-generation anti-CD20 antibody, obinutuzumab, was added to chlorambucil and to venetoclax, a BCL2 inhibitor

(discussed later) in the treatment of the older adult population. When compared with chlorambucil, alone or in combination with rituximab and chlorambucil, obinutuzumab was shown to be superior and relatively safer.[28,29] Other invetigational agents showing efficacy include lenalidomide (an immune modulator related to thalidomide), albeit with an interestingly initial flair-up phenomenon of disease-enlarged lymph nodes.

A major breakthrough in treatment was achieved through drugs targeting the BCR pathway. A BTK inhibitor, ibrutinib, has shown an impressive progression-free survival in relapsing patients and as first line treatment, especially for patients with resistant or high-risk 17p-deleted CLL. It is currently considered as standard of care especially for high-risk patients. Interestingly, at treatment initiation, it causes a steep rise in peripheral blood lymphocyte counts, most likely due to translocation of nodal lymphocytes to the blood. This is actually correlated to a better and prolonged response.[28,30] Currently, second-generation BTK inhibitors such as acalabrutinib and zanubrutinib with fewer side effects are being tested.[28,30,31] In combination with rituximab, another pathway inhibitor, the PI3Kδ inhibitor idelalisib had an efficacy superior to rituximab alone in frail, older patients who were relapsing. However, it has been linked to severe toxicities, and second-generation kinases such as duvelisib and copanlisib are expected to be used.[28,30,31]

Furthermore, clinical trials with small molecules targeting BCL2, venetoclax, have shown deep and sustained responses, both in the upfront first line and in the relapsed refractory patients. Combined with anti-CD20 obinutuzumab and rituximab it allows the possibility to stop treatment after 12 to 24 months of therapy.

However, resistance due to specific mechanisms (e.g., BTK, PLCg2, and BCL2 mutations) is emerging, indicating the need to develop time-limited combinations and novel agents combinations.[28,30,31] In addition, under investigation are various novel agents targeting other surface molecules (CD23, CD79b, and other CD20 inhibitors), pathway inhibitors (SYK, PI3K, second-generation BTK, and others), and immune-checkpoint inhibitors. Some immunologic novel approaches include expanded autoreactive activated T cells (Chimeric antigen receptor [CAR] T-cells), which is showing promise. Unfortunately, the majority of patients with CLL do not attain durable responses.[31,32]

Allo-HSCT is the only curative treatment for CLL. Allo-HSCT relies on myeloablative doses of chemoradiotherapy, which makes the treatment unacceptably risky for the majority of patients with CLL. In nonmyeloablative, or reduced-intensity, approaches, rates of engraftment are similar to fully ablative conditioning regimens but with lower rates of early toxicity. Early evidence suggests that the graft-versus-leukemia (GvL) effect is present. Patients with deletion 17p (p53 involvement) with an extremely poor prognosis are physically fit candidates for allo-HSCT,[28,29] even as newer agents (i.e., BTK inhibitors, PI3K inhibitors, and BCL2 inhibitors) are showing high efficacy in this high-risk population. Studies involving autologous transplantation and high-dose chemotherapy for CLL have a limited survival advantage.

With treatment advances, the overall survival of patients with CLL has improved across the globe as a result of targeted therapy, even in high-risk and in patients with relapsed and refractory CLL. Thus classic prognostication (as presented in Table 77.5) is becoming less accurate as treatment improves.

Immunologic Aspects of CLL

CLINICAL PEARLS

Chronic Lymphocytic Leukemia: Immunologic Manifestations

- Panhypogammaglobulinemia
- A monoclonal immunoglobulin peak, usually of the immunoglobulin M type
- Downregulation of T-cell function and aberrant cytokine production
- Defects in the complement system
- High risk of recurrent infections—encapsulated bacteria and opportunistic infections
- Autoimmune-associated phenomena:
  - Autoimmune hemolytic anemia
  - Autoimmune thrombocytopenia
  - Pure red cell aplasia and autoimmune neutropenia
  - Other autoimmune disorders (myositis, vasculitis, pemphigus vulgaris, acquired angioedema, glomerulonephritis)

CLL is characterized by multiple immune deficiencies and autoimmune phenomena. It is reasonable to hypothesize that immune incompetence and autoimmunity are two sides of the same coin.

The Pathophysiologic Rationale

CLL cells secrete TGF-β, which is a potent inhibitor of B-cell proliferation. They also release high levels of circulating IL-2 receptor, which downregulates the T-helper (Th) cell function. Unlike normal cells, activated B cells, both B-cell CLL cells and anergic normal B cells, fail to present soluble antigen and alloantigens. Moreover, the T cells in patients with CLL often demonstrate profound abnormalities of their antigen receptor (TCR) repertoire and appear dysfunctional in terms of cytokine secretion.[33] This cytokine imbalance may be the cause of upregulation of the BCL-2 antiapoptotic activity. T-cell dysfunction could also explain the higher incidence of autoimmune complications, such as autoimmune hemolytic anemia (AIHA), among patients receiving purine analogues therapy, which induces T-cell depletion. Certain T-cell subsets appear to prevent the development of autoreactive B cells. When these are absent (e.g., after treatment with purine analogues), autoreactive B-cell clones may easily emerge and expand.

Immunologic Deficiencies

Patients with CLL are extremely sensitive to a number of infectious agents. A monoclonal Ig peak, usually of the IgM type, is found in 5% of patients with CLL, and a small amount of a monoclonal component can be identified in the serum or urine of 60% of patients. Hypogammaglobulinemia occurs in at least 60% of B-cell CLL cases and may include all three classes (IgG, IgA, and IgM). The pathogenesis of hypogammaglobulinemia in B-cell CLL is poorly understood because this phenomenon is rare in other B-cell malignancies except multiple myeloma. Low Ig levels correlate with recurrent infections of encapsulated organisms. In patients who receive intravenous immunoglobulin (IVIG), there is a decrease in the incidence of major bacterial infections.

Infections are a major cause of morbidity and mortality in patients with CLL. Impaired humoral and cellular immunities, defects in the complement systems, and variable neutropenia, depending on marrow infiltrates, all contribute to the high rate

of infections. Opportunistic infections are initially uncommon as the result of the relative preservation of cellular immunity early in the disease. Infection risk increases following purine analogue therapy because of the side effects of myelosuppression and marked lymphopenia with T-cell depletion. The addition of rituximab, the anti–B cell marker CD20 antibody, to nucleoside analogue-based therapy does not appear to increase the risk of early or late infections but may increase the rate of neutropenia. Active immunization with vaccines is hampered by the patient's inability to generate or retain a long and significant immune response. The advances in therapy with newer agents show high efficacy (i.e., BTK inhibitors, PI3K inhibitors, and BCL2 inhibitors) but also more atypical bacterial, fungal, and protozoal infections.[31]

Autoimmune Phenomena

Autoimmune-associated features are common in CLL. These manifestations primarily affect hematopoietic cells. For example, the most common known cause of AIHA is CLL.[34] Positive result of direct antiglobulin test (direct Coomb test) has been reported to be as high as in 7% to 35% of patients with CLL, and AIHA itself occurs in 10% to 25% of patients during the course of their disease, twice as often in patients with unmutated genes as in those with mutated ones. Autoantibodies against red blood cells (RBCs) are warm-reactive polyclonal IgG. They are not secreted by the malignant clone but rather by normal B cells.[34] Cold agglutinins are rare. AIHA is thought to arise from the imbalance among lymphocyte subsets, contributed to by therapy, resulting in the emergence of the autoimmune clone. It is usually observed in advanced stages of the disease, correlates with a poor prognosis, and has a close relationship with the activity of the CLL. After therapy, the autoimmune antibodies may remit in 70% of the treated patients.

Idiopathic thrombocytopenic purpura (ITP) is observed in approximately 2% to 3% of cases and presents as increased megakaryocytes in bone marrow. It should be distinguished from immune thrombocytopenia induced by marrow infiltration, which is very common (in up to 50% of patients at presentation).[34] Two-thirds of patients with CLL-associated ITP also have AIHA (Evan syndrome). Pure red cell aplasia (PRCA) and autoantibodies against neutrophils are only rarely observed but are part of the CLL-related autoimmunity repertoire. Interestingly, the addition of cyclophosphamide represses the emergence of fludarabine-induced AIHA, but the latter may also be seen with other chemotherapies (i.e., bendamustine).[35] Autoimmune phenomena in patients treated with purine analogues (mostly fludarabine-related) are of a more severe nature.

Other rare entities are reported as paraneoplastic autoimmune disorders with connective tissue disease manifestations, such as polymyositis, dermatopolymyositis, and focal myositis or as vasculitis, pemphigus vulgaris, and acquired angioedema. These autoimmune disorders are related to T-cell dysfunction and may be associated with purine analogue treatment. Paraneoplastic pemphigus also occurs in patients with CLL and may be triggered by chemotherapy and radiotherapy. Glomerulonephritis and nephrotic syndrome are seldom reported but, when present, are related to different mechanisms, such as cryoglobulins and antineutrophil cytoplasmic antibodies (ANCAs).

Therapy of autoimmune phenomena includes high-dose steroids and disease control.[34] In patients refractory to or relapsing after steroid therapy, more aggressive treatment is warranted. High-dose Igs offers transient amelioration in some patients. Splenectomy or splenic irradiation, cytotoxic agents, or cyclosporine may represent valid rescue approaches. In cases where AIHA has been triggered by fludarabine, further exposure is hazardous. Rituximab may be an alternative agent for the treatment CLL-associated autoimmune diseases, including rare autoimmune phenomena, such as pemphigus and PRCA.

In line with the introduction of novel agents, the possible role of these drugs in inducing or exacerbating autoimmune phenomena and, on the other hand, of treating them still needs to be elucidated and with most is anecdotally reported.[36] PI3K inhibitors may cause various autoimmune complications and therefore theoretically may cause or exacerbate these in CLL. Although some reports connect ibrutinib with autoimmune complications, most evidence supports its efficacy in treating or at least improving such autoimmune complications.

Other Malignancies

Second malignancies (hematologic and solid tumors) are not uncommon in CLL. The most common hematologic malignancy is the Richter transformation to diffuse large B-cell lymphoma, which occurs in approximately 5% of patients, as well as other high-grade lymphoproliferative diseases. Dermatologic tumors, such as basal cell carcinoma, are the most frequent of the solid tumors encountered in patients with CLL, and these malignancies are more likely to be locally aggressive and metastatic.[29] The pathogenesis of these second cancers is not fully understood, and although disease-related genetic factors (i.e., 17p deletion, notch mutation) are a major determinant, it is probably multifactorial and includes Epstein-Barr virus (EBV) infection and BCR configuration to respond to multiple autoantigens and immune/inflammatory stimuli present in the microenvironment.[29,37]

CONCLUSIONS

CLL is a common indolent lymphoid neoplasm with a wide clinical heterogeneity. It is suspected and diagnosed more commonly because of routine blood tests. Diagnosis is made with simple immunophenotyping. Cytogenetics and molecular diagnostic techniques are needed to determine the prognosis. The complications of CLL appear to be unique to this neoplasm and are part of a failing immune system with T-cell and B-cell dysregulation causing both deficiencies and predisposing patients to recurrent infections and autoimmune diseases. New molecular and protein markers are key to the current novel effective targeted therapies approach for the treatment of these patients.

REFERENCES

1. Pui CH, Yang JJ, Hunger SP, et al. Childhood Acute Lymphoblastic Leukemia: Progress Through Collaboration. *J Clin Oncol*. 2015;33:2938–2948.
2. Greaves M. Infection, immune responses and the aetiology of childhood leukaemia. *Nat Rev Cancer*. 2006;6:193–203.
3. Kuehn HS, Boisson B, Cunningham-Rundles C, et al. Loss of B Cells in Patients with Heterozygous Mutations in IKAROS. *N Engl J Med*. 2016;374:1032–1043.
4. Gocho Y, Yang JJ. Genetic Defects in Hematopoietic Transcription Factors and Predisposition to Acute Lymphoblastic Leukemia. *Blood*. 2019
5. Martin-Lorenzo A, Hauer J, Vicente-Duenas C, et al. Infection Exposure is a Causal Factor in B-cell Precursor Acute Lymphoblastic Leukemia as a Result of Pax5-Inherited Susceptibility. *Cancer Discov*. 2015;5:1328–1343.
6. Coustan-Smith E, Mullighan CG, Onciu M, et al. Early T-cell precursor leukaemia: a subtype of very high-risk acute lymphoblastic leukaemia. *The Lancet Oncol*. 2009;10:147–156.

7. Tal N, Shochat C, Geron I, Bercovich D, Izraeli S. Interleukin 7 and thymic stromal lymphopoietin: from immunity to leukemia. *Cellular and molecular life sciences: CMLS*. 2014;71:365–378.
8. Roberts KG, Li Y, Payne-Turner D, et al. Targetable kinase-activating lesions in Ph-like acute lymphoblastic leukemia. *N Engl J Med*. 2014;371:1005–1015.
9. Izraeli S. Beyond Philadelphia: 'Ph-like' B cell precursor acute lymphoblastic leukemias - diagnostic challenges and therapeutic promises. *Current Opinion Hematol*. 2014;21:289–296.
10. O'Connor D, Enshaei A, Bartram J, et al. Genotype-Specific Minimal Residual Disease Interpretation Improves Stratification in Pediatric Acute Lymphoblastic Leukemia. *J Clin Oncol*. 2017 JCO2017740449.
11. Szczepanski T, Orfao A, van der Velden VH, San Miguel JF, van Dongen JJ. Minimal residual disease in leukaemia patients. *Lancet Oncol*. 2001;2:409–417.
12. Conter V, Bartram CR, Valsecchi MG, et al. Molecular response to treatment redefines all prognostic factors in children and adolescents with B-cell precursor acute lymphoblastic leukemia: results in 3184 patients of the AIEOP-BFM ALL 2000 study. *Blood*. 2010;115:3206–3214.
13. Anderson K, Lutz C, van Delft FW, et al. Genetic variegation of clonal architecture and propagating cells in leukaemia. *Nature*. 2011;469:356–361.
14. Dixon SB, Chen Y, Yasui Y, et al. Reduced Morbidity and Mortality in Survivors of Childhood Acute Lymphoblastic Leukemia: A Report From the Childhood Cancer Survivor Study. *J Clin Oncol*. 2020;38:3418–3429.
15. Kantarjian HM, DeAngelo DJ, Stelljes M, et al. Inotuzumab Ozogamicin versus Standard Therapy for Acute Lymphoblastic Leukemia. *N Engl J Med*. 2016;375:740–753.
16. Kantarjian H, Stein A, Gokbuget N, et al. Blinatumomab versus Chemotherapy for Advanced Acute Lymphoblastic Leukemia. *N Engl J Med*. 2017;376:836–847.
17. Maude SL, Laetsch TW, Buechner J, et al. Tisagenlecleucel in Children and Young Adults with B-Cell Lymphoblastic Leukemia. *N Engl J Med*. 2018;378:439–448.
18. Zent CS, Kyasa MJ, Evans R, Schichman SA. Chronic lymphocytic leukemia incidence is substantially higher than estimated from tumor registry data. *Cancer*. 2001;92:1325–1330.
19. Chiorazzi N, Rai KR, Ferrarini M. Chronic lymphocytic leukemia. *N Engl J Med*. 2005;352:804–815.
20. Strati P, Shanafelt TD. Monoclonal B-cell lymphocytosis and early-stage chronic lymphocytic leukemia: diagnosis, natural history, and risk stratification. *Blood*. 2015;126:454–462.
21. Goldin LR, Bjorkholm M, Kristinsson SY, Turesson I, Landgren O. Elevated risk of chronic lymphocytic leukemia and other indolent non-Hodgkin's lymphomas among relatives of patients with chronic lymphocytic leukemia. *Haematologica*. 2009;94:647–653.
22. Ten Hacken E, Burger JA. Microenvironment interactions and B-cell receptor signaling in Chronic Lymphocytic Leukemia: Implications for disease pathogenesis and treatment. *Biochim Biophys Acta*. 2015;1863:401–413.
23. Herishanu Y, Perez-Galan P, Liu D, et al. The lymph node microenvironment promotes B-cell receptor signaling, NF-kappaB activation, and tumor proliferation in chronic lymphocytic leukemia. *Blood*. 117:563–574.
24. Messmer BT, Messmer D, Allen SL, et al. In vivo measurements document the dynamic cellular kinetics of chronic lymphocytic leukemia B cells. *J Clin Invest*. 2005;115:755–764.
25. Rai KR, Sawitsky A, Cronkite EP, Chanana AD, Levy RN, Pasternack BS. Clinical staging of chronic lymphocytic leukemia. *Blood*. 1975;46:219–234.
26. Binet JL, Lepoprier M, Dighiero G, et al. A clinical staging system for chronic lymphocytic leukemia: prognostic significance. *Cancer*. 1977;40:855–864.
27. Nabhan C, Raca G, Wang YL. Predicting Prognosis in Chronic Lymphocytic Leukemia in the Contemporary Era. *JAMA Oncol*. 2015;1:965–974.
28. Scheffold A, Stilgenbauer S. Revolution of Chronic Lymphocytic Leukemia Therapy: the Chemo-Free Treatment Paradigm. *Curr Oncol Rep*. 2020;22:16.
29. Stilgenbauer S, Furman RR, Zent CS. Management of chronic lymphocytic leukemia. *Am Soc Clin Oncol Educ Book*. 2015:164–175.
30. Brown JR. How I treat CLL patients with ibrutinib. *Blood*. 2018;131:379–386.
31. Iovino L, Shadman M. Novel Therapies in Chronic Lymphocytic Leukemia: A Rapidly Changing Landscape. *Curr Treat Options Oncol*. 2020;21:24.
32. Rhodes JM, Schuster SJ. Chimeric Antigen Receptor T Cells in Chronic Lymphocytic Leukemia: Are We Any Closer to a Cure? *Cancer J*. 2019;25:436–441.
33. Scrivener S, Goddard RV, Kaminski ER, Prentice AG. Abnormal T-cell function in B-cell chronic lymphocytic leukaemia. *Leuk Lymphoma*. 2003;44:383–389.
34. Hamblin TJ. Autoimmune complications of chronic lymphocytic leukemia. *Semin Oncol*. 2006;33:230–239.
35. Goldschmidt N, Gural A, Ben-Yehuda D, Gatt ME. Short communication: bendamustine-related hemolytic anemia in chronic lymphocytic leukemia. *Cancer Chemother Pharmacol*. 2013;72:709–713.
36. Vitale C, Montalbano MC, Salvetti C, et al. Autoimmune Complications in Chronic Lymphocytic Leukemia in the Era of Targeted Drugs. *Cancers (Basel)*. 2020:12.
37. Rossi D, Gaidano G. Richter syndrome: pathogenesis and management. *Semin Oncol*. 2016;43:311–319.

78

Lymphomas

Stefania Pittaluga, Ivo M.B. Francischetti, Joo Y. Song, and Elaine S. Jaffe

The classification of malignant lymphomas continues to undergo revision based on insights gained through the application of immunologic and molecular techniques as well as the application of these discoveries to individualized therapeutic approaches. Early classifications were based on morphologic characteristics of the neoplastic elements; however, with increasing knowledge of the complexity of the immune system, a more functional approach was sought. Differentiation schemes provided a useful starting point for understanding lymphomas (Fig. 78.1). High-throughput genomic studies have been applied to lymphomas to define their molecular signatures with the aim of improving the understanding of oncogenic pathways and their clinical implications. These studies have led to new prognostic and diagnostic tools associated with the emergence of more targeted therapies.[1]

KEY CONCEPTS

Lymphoma

- Classification consists of a list of individual disease entities defined by morphologic, immunophenotypic, genetic, and clinical features.
- Neoplastic cells are linked to the postulated normal counterpart, when possible.
- Histologic grade should be applied within individual diseases.
- Clinical factors for individual patients, as measured by the International Prognostic Index (IPI) and gene expression profiling, are useful in predicting clinical outcome.

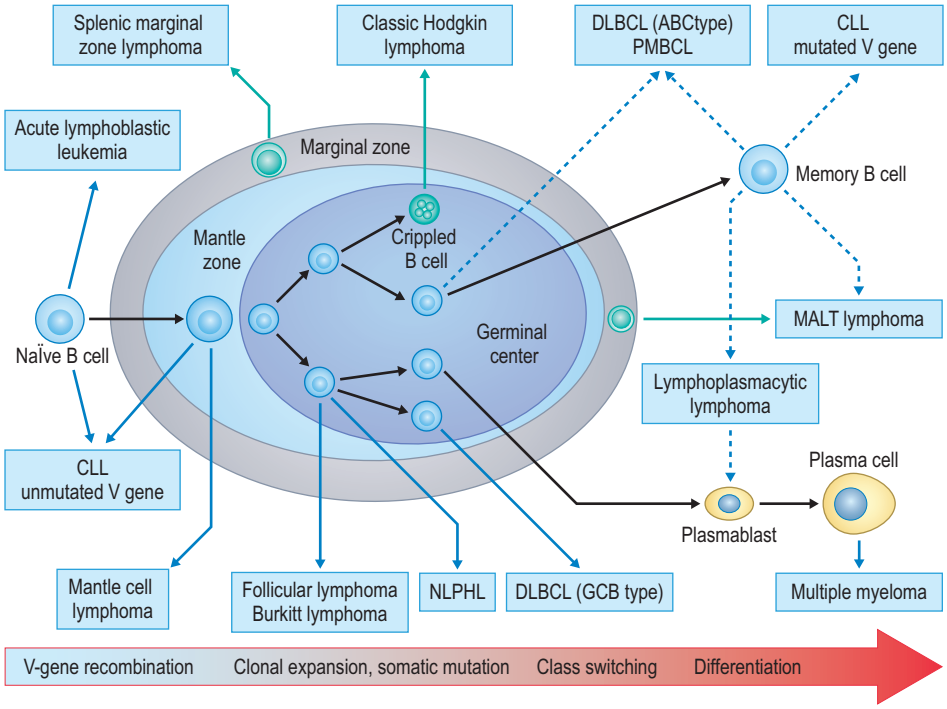


FIG. 78.1 Normal B-Cell Differentiation in Relation to a Secondary B Follicle, Mutational Stages of the Immunoglobulin Genes, and Cellular Counterparts for B-Cell Lymphomas. Simplified version of B-cell development indicates points at which V-gene recombination, clonal expansion, and somatic mutations occur in relation to a secondary B follicle. B-cell lymphomas are related to different stages of B-cell differentiation and function. *ABC type*, Activated B-cell-like type; *CLL*, chronic lymphocytic leukemia/lymphoma; *DLBCL*, diffuse large B-cell lymphoma; *GCB type*, germinal center B-cell–like type; *MALT lymphoma*, marginal zone B-cell lymphoma of mucosa-associated lymphoid tissue (MALT) type; *NLPHL*, nodular lymphocyte-predominant Hodgkin lymphoma; *PMBCL*, primary mediastinal B-cell lymphoma.

The guiding principles of the World Health Organization (WHO) classification of neoplasms of the hematopoietic and lymphoid tissues were published in 2001 and updated in 2008 and 2016.[2] These classifications identified individual diseases based on an integration of morphologic, immunophenotypic, genetic, and clinical features. The application of gene expression profiling (GEP) in lymphomas has generated distinct molecular "signatures" for a variety of disease entities, either corresponding more closely to different stages of lymphoid differentiation or offering insights into mechanisms of neoplastic transformation. The advent of whole-genome sequencing contributes additional insights into pathogenetic mechanisms. A latest revision of the WHO classification of lymphoid neoplasms (Table 78.1) was published clarifying the diagnosis of lesions at the very early stages of lymphomagenesis, refining diagnostic criteria for some entities, and detailing the expanding genetic/molecular landscape of numerous lymphoid neoplasms and their clinical correlates.[3]

TABLE 78.1 WHO Classification of Tumors of Hematopoietic and Lymphoid Tissues (2016)

B-Cell Neoplasm

Precursor B-cell lymphoblastic leukemia/lymphoma
- B-lymphoblastic leukemia/lymphoma, not otherwise specified
- B-lymphoblastic leukemia/lymphoma with recurrent genetic abnormalities

Mature B-Cell Neoplasm

Chronic lymphocytic leukemia/small lymphocytic lymphoma
Monoclonal B-cell lymphocytosis[a]
B-cell prolymphocytic leukemia
Splenic B-cell marginal zone lymphoma
Hairy cell leukemia
Splenic B-cell lymphoma/leukemia, unclassifiable
- *Splenic diffuse red pulp small B-cell lymphoma*
- *Hairy cell leukemia-variant*

Lymphoplasmacytic lymphoma
Waldenström macroglobulinemia
Monoclonal gammopathy of undetermined significance (MGUS), IgM[a]
μ heavy-chain disease
γ heavy-chain disease
α heavy-chain disease
Monoclonal gammopathy of undetermined significance (MGUS), IgG/A[a]
Plasma cell myeloma
Solitary plasmacytoma of bone
Extraosseous plasmacytoma
Monoclonal immunoglobulin deposition diseases[a]
Extranodal marginal zone lymphoma of mucosa-associated lymphoreticular tissue (MALT) lymphoma
Nodal marginal zone lymphoma
- Pediatric nodal marginal zone lymphoma

Follicular lymphoma
- In situ follicular neoplasia[a]
- Duodenal-type follicular lymphoma[a]
- Pediatric-type follicular lymphoma[a]
- *Large B-cell lymphoma with IRF4 rearrangement*[a]
- Primary cutaneous follicle-center lymphoma

Mantle cell lymphoma
- In situ *mantle cell neoplasia*[a]

Diffuse large B-cell lymphoma (DLBCL), NOS
- Germinal-center B-cell type[a]
- Activated B-cell type[a]

T-cell/histiocyte-rich large B-cell lymphoma
Primary DLBCL of the central nervous system
Primary cutaneous DLBCL, leg type
EBV+ DLBCL, NOS[a]
EBV+ mucocutaneous ulcer[a]
DLBCL associated with chronic inflammation
Lymphomatoid granulomatosis
Primary mediastinal (thymic) large B-cell lymphoma
Intravascular large B-cell lymphoma
ALK positive large B-cell lymphoma
Plasmablastic lymphoma
Primary effusion lymphoma
HHV-8–positive DLBCL, NOS[a]
Burkitt lymphoma
Burkitt-like lymphoma with 11q aberration[a]
High-grade B-cell lymphoma, with MYC and BCL2 and/or BCL6 rearrangements[a]
B-cell lymphoma unclassifiable, with features intermediate between diffuse large B-cell lymphoma and classic Hodgkin lymphoma

T-Cell Neoplasm

Precursor T-cell lymphoblastic leukemia/lymphoma

Mature T-Cell and NK-Cell Neoplasms

T-cell prolymphocytic leukemia
T-cell large granular lymphocytic leukemia
Chronic lymphoproliferative disorder of NK cells
Aggressive NK leukemia
Systemic EBV+ T-cell lymphoma of childhood[a]
Hydroa vacciniforme-like lymphoproliferative disorder[a]
Adult T-cell leukemia/lymphoma
Extranodal NK/T-cell lymphoma, nasal type
Enteropathy-associated T-cell lymphoma
Monomorphic epitheliotropic intestinal T-cell lymphoma[a]
Indolent T-cell lymphoproliferative disorder of the GI tract[a]
Hepatosplenic T-cell lymphoma
Subcutaneous panniculitis-like T-cell lymphoma
Mycosis fungoides
Sézary syndrome
Primary cutaneous CD30+ positive T-cell lymphoproliferative disorders
- Primary cutaneous anaplastic large-cell lymphoma
- Lymphomatoid papulosis

Primary cutaneous γ/δ T-cell lymphoma
Primary cutaneous CD8 positive aggressive epidermotropic cytotoxic T-cell lymphoma
Primary cutaneous acral CD8+ T-cell lymphoma[a]
Primary cutaneous CD4-positive small/medium T-cell lymphoproliferative disorder[a]
Peripheral T-cell lymphoma, NOS
Angioimmunoblastic T-cell lymphoma
Follicular T-cell lymphoma[a]
Nodal peripheral T-cell lymphoma with T_{FH} phenotype[a]
Anaplastic large-cell lymphoma, ALK positive
Anaplastic large-cell lymphoma, ALK negative[a]
Breast implant-associated anaplastic large-cell lymphoma[a]

Hodgkin Lymphoma

Nodular lymphocyte predominant Hodgkin lymphoma
Classic Hodgkin lymphoma
Nodular sclerosis Hodgkin lymphoma
Lymphocyte-rich classic Hodgkin lymphoma
Mixed cellularity classic Hodgkin lymphoma
Lymphocyte-depleted classic Hodgkin lymphoma

EBV, Epstein-Barr virus; *GI*, gastrointestinal; *HHV*, human herpes virus; *NK*, natural killer; *NOS*, not otherwise specified; *WHO*, World Health Organization.
[a]Refers to changes from the 2008 classification.
Note: More common entities are underlined. Provisional entities in italics.
From Swerdlow SH, Campo E, Harris NL, et al. *WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues*. Lyon, France: International Agency for Research on Cancer; 2017.

This chapter focuses on the classification of neoplasms derived from mature B cells, T cells, and natural killer (NK) cells, with emphasis on malignant lymphomas. Special attention is devoted to the impact of clinical features (e.g., age and anatomic site) on disease definition and a greater appreciation of early events in neoplastic transformation. These early lesions can sometimes be detected in otherwise healthy individuals, a situation that may or may not progress to overt lymphoma or leukemia. These early entities appear to carry fewer genetic aberrations compared with the conventional forms of the disease, which perhaps explains their indolent clinical behavior.[4]

CLINICAL PEARLS

Indolent Lymphomas

- Natural history: survival measured in years.
- Least sensitive to therapy.
- Good response to low-dose oral alkylating agents, radiotherapy, and steroids, but not curable.
- Higher response rate and complete remission with combination of standard chemotherapy and anti-CD20 monoclonal antibody.
- Gene expression profiling can help identify patients who might benefit from high-dose chemotherapy and autologous stem cell transplantation, a potentially curative modality.

MATURE B-CELL NEOPLASMS

KEY CONCEPTS

Somatic Mutation in Relation to Normal B-Cell Development

- *Premutational stage:* circulating naïve B cells (immunoglobulin [Ig]M$^+$/D$^+$) before antigen exposure
- *Stage of somatic mutation, clonal expansion, and isotype switch:* at the germinal center
- *Postmutational stage:* selected B cells move to the periphery (post–germinal center), to the recirculating pool (memory B cells), or undergo terminal differentiation (plasma cells)

Lymphoplasmacytic Lymphoma

Lymphoplasmacytic lymphoma (LPL) is a disease of adult life, usually presenting with generalized lymphadenopathy, constitutional symptoms, and splenomegaly.

Histologically, it consists of a diffuse proliferation of small lymphocytes (many with plasmacytoid features) and plasma cells, with or without immunoglobulin (Ig)-filled intranuclear inclusions (Dutcher bodies). An increased number of mast cells and iron-laden macrophages can be seen. Although many B-cell neoplasms occasionally show maturation to plasmacytoid or plasma cells with cytoplasmic Ig, the term LPL should be restricted to tumors lacking features of other well-defined entities, such as chronic lymphocytic leukemia (CLL) or mantle cell lymphoma (MCL), which occasionally can manifest plasmacytoid differentiation. Many but not all patients with LPL have clinical evidence of Waldenström macroglobulinemia (WM), based on the detection of an IgM monoclonal gammopathy of any concentration and associated with bone marrow involvement by LPL (Chapter 79).

In LPL, cells have surface and cytoplasmic Ig, usually IgM (but usually lacking IgD), and express B cell–associated antigens (CD19, CD20, CD22, CD79a). They are generally negative for CD5 and cyclin D1, distinguishing LPL from CLL and MCL, respectively. CD25, CD10, or CD11c may be weakly expressed in some cases. The postulated normal counterpart is thought to be a postfollicular medullary cord B cell, based, in part, on the presence of somatic mutations in the Ig heavy-chain and light-chain variable region genes.

The identification of *MYD88* L265P mutation (found in approximately 90% of WM cases) is a reliable marker supporting a diagnosis of LPL.[5] This mutation is also found in a significant proportion of IgM, but not IgG or IgA, monoclonal gammopathy of undetermined significance (MGUS) cases (Chapter 79). Cases with the *CXCR4^{WHIM}* mutations exhibit a resistance to ibrutinib.

Mantle Cell Lymphoma

MCL usually presents in adults (median age 62; variable male predominance) with advanced-stage disease—involving lymph nodes, Waldeyer ring lymphoid tissue, spleen, bone marrow, and peripheral blood. Gastrointestinal (GI) tract involvement is common, often associated with lymphomatous polyposis. Retrospective studies have shown a poor prognosis (median survival 3 to 5 years), with a high relapse rate following initial remission. MCL is composed of small lymphoid cells with slightly irregular nuclear contours, finely clumped chromatin, and scant cytoplasm. Specifically, blastoid and pleomorphic variants have been associated with a more aggressive clinical course. The postulated normal counterpart is a CD5$^+$ "naïve" surface IgM$^+$ and surface IgD$^+$ B cell, found in peripheral blood and the mantles of reactive follicles.

MCL is characterized by t(11;14)(q13;q32)—involving cyclin D1 (*CCND1*) and the *IGH* genes; an overexpression of *CCND1* is believed to be essential in the pathogenesis. Rare variants negative for *CCND1* but with similar immunomorphology and GEP have also been identified.[6] Half of the *CCND1* expression/rearrangement-negative forms have *CCND2* translocations, often with *IGK* or *IGL* as a partner locus.[7] SOX11 is overexpressed in most *CCND1*-positive and *CCND1*-negative cases.[8] Additional alterations involving other cell cycle regulatory proteins (RB, p53, CDK inhibitors) have been described in the more aggressive forms of MCL.

In situ mantle cell neoplasia (ISMCN) (previously "in situ" MCL) represents a clonal proliferation of cyclin D1–positive cells restricted to the inner mantle cuffs in an otherwise reactive lymph node/lymphoid tissue and is usually an incidental finding. Some cases will eventually progress to overt MCL[2]; however, the risk of progression is difficult to ascertain because the number of reported cases is small. The more recently identified non-nodal variant is characterized by a leukemic phase without nodal disease but often long-standing splenomegaly. These cases develop from IgHV-mutated SOX11-negative B cells, carry t(11;14) with few additional chromosomal abnormalities, and lack expression of SOX11.[9]

The proliferation rate based on Ki67 positivity has been considered of prognostic relevance. More recently, GEP using genes involved in cell cycle progression and DNA synthesis has identified a proliferation signature that defines different prognostic groups and shows some correlation with cytologic subtype such as the blastoid variant.[10]

The treatment approach for newly diagnosed patients with MCL depends on their eligibility for stem cell transplantation (SCT). The incorporation of rituximab into chemotherapeutic regimens has become standard of care. The application of the Bruton tyrosine kinase (BTK) inhibitors might change treatment paradigms by obviating the need for transplantation in

younger patients and chemotherapy in older patients. Patients with the non-nodal variant do not appear to require aggressive chemotherapy.[9]

Follicular Lymphoma

Follicular lymphoma (FL) is the most common subtype of non-Hodgkin lymphoma (NHL) in the United States and accounts for approximately 45% of all newly diagnosed cases. It has a peak incidence in the fifth and sixth decades and is rare younger than the age of 20 years, and both sexes are equally affected. Most patients have stage 3 or 4 disease at diagnosis with generalized lymphadenopathy and bone marrow involvement. Approximately 10% of patients have circulating malignant cells.

FL is composed of varying proportions of follicle center–type cells, centrocytes, and centroblasts, representing the proliferative component. According to the WHO classification, all low-grade FL are combined into a single category, grade 1/2—with an overall predominance of centrocytes and fewer than 15 centroblasts per high-power field (hpf). Grade 3 (with >15 centroblasts/hpf) is further subdivided into 3A and 3B, based on the presence or absence of background centrocytes. FL represents the neoplastic counterpart of the reactive germinal center cells; intraclonal heterogeneity with high numbers of somatic and ongoing Ig mutations can be detected in the neoplastic cells, as in the normal counterparts. Biologically, grade 3B is more closely related to diffuse large B-cell lymphoma (DLBCL) compared with FL.

The vast majority of FL (approximately 90%) is associated with a t(14;18) involving rearrangement of the *BCL2* gene. This translocation appears to result in constitutive expression of BCL2 protein, inhibiting apoptosis in lymphoid cells. The cells of FL accumulate and are at risk for secondary mutations, which may be associated with histologic progression. It is thought that the *BCL2* translocation occurs at a very early stage of B-cell development, during Ig gene rearrangement. Mutations in *BCL2* can lead to loss of BCL2 protein in the presence of the translocation; fluorescence in situ hybridization (FISH) studies for t(14;18) in these cases are informative. Grade 3B FL is less commonly associated with *BCL2* translocation and carries genetic aberrations more commonly seen in DLBCL. The neoplastic cells in FL have a mature B-cell phenotype with expression of the B-cell antigens (CD19, CD20, and CD22). Surface Ig is positive, most commonly with IgM expression, but IgG or IgA can also be seen in many cases. CD10 and BCL6 are positive, but CD5 is usually negative. Mutation analysis of FL has shown recurrent mutations in genes involved in epigenetic regulation and chromatin modification. Transformation of FL to DLBCL is a common event, and, although there is no unique mutation to predict transformation, genes commonly mutated include *TP53*, *EZH2*, *MYC*, *CCND3, MLL3*, and *CARD11*.[11]

FL is indolent but is still incurable with available therapeutic modalities. The occurrence of *BCL2* translocation at a very early stage of B-cell development might contribute to the difficulty in eradicating the neoplastic clone with chemotherapy. Clinical parameters have been used to develop prognostic indexes (e.g., the Follicular Lymphoma International Prognostic Index). Recently, much has also been learned about the mutational landscape of FL. Mutations in *CREBBP, MLL2*, and *EZH2* have been shown as extremely common early events and may be potential therapeutic targets.[12] The natural history of the disease is associated with histologic progression in cellular composition and pattern (Fig. 78.2).

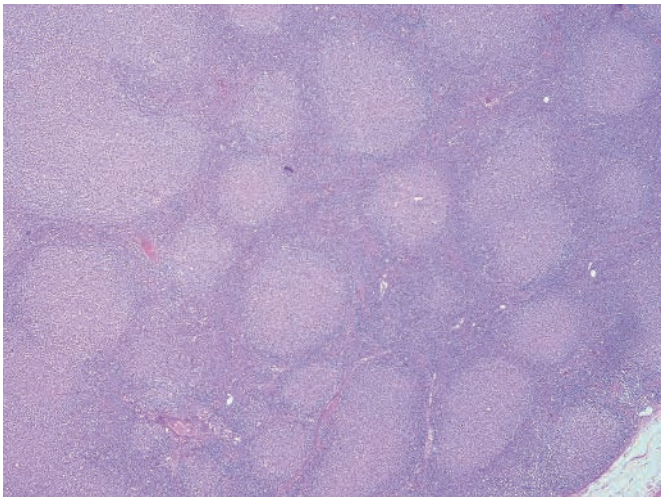


FIG. 78.2 Follicular Lymphoma. The neoplastic follicles are similar in size and are partially surrounded by lymphoid cuffs. In contrast to reactive germinal centers, they lack polarization and tingible body macrophages ("starry-sky pattern").

The 2008 WHO classification had recognized variants of FL: namely, *pediatric FL, GI tract FL*, and other *extranodal FL*. Pediatric-type FL (now a definite entity in the 2016 WHO classification) is more common in males and presents as localized nodal disease. It typically has an expansile serpiginous pattern, with relatively monotonous, medium-sized, blastoid cellular composition. *BCL2, BCL6*, and *MYC* rearrangements are lacking. Complete remissions may be obtained with either surgical excision or local radiation therapy. Some studies have raised the possibility that pediatric-type FL might be a "benign clonal proliferation with low malignant potential."[13,14] A recent study showed that pediatric-type FL had a low genetic complexity with frequent mutations in *TNFRSF14* and *MAP2K1*, the latter leading to ERK pathway activation. Of note, mutations commonly seen in adult FL such as *EZH2* and other histone-modifying genes were not seen or rare.[15]

The WHO classification recognizes "FL in situ" (now termed *in situ follicular neoplasia* [ISFN]) as a distinctive lesion. It should be distinguished from partial involvement of FL. ISFN shows involvement of germinal centers by CD10 and BCL2-positive cells carrying t(14;18) in an otherwise reactive lymph node, which is typically an incidental finding. Patients have a very low risk of progression, but ISFN may be detected in association with other forms of B-cell lymphoma, necessitating additional clinical assessment.[16] Fewer chromosomal abnormalities are noted in IFSN in comparison with partial or overt FL.[17] Patients lacking evidence of FL at staging have a low risk of developing the disease; this phenomenon appears to represent the tissue counterpart of circulating clonal B cells carrying t(14;18) as detected in healthy individuals. A higher level of circulating t(14;18)-positive lymphocytes ($>10^{-4}$ of total cells) indicates a higher risk for FL.

The 2016 WHO classification recognizes GI tract FL as duodenal-type FL because these can also be seen elsewhere in the GI tract and have features overlapping with ISFN and extranodal mucosa-associated lymphoid tissue (MALT) lymphoma.[17] These present as small mucosal polyps or nodules, and most cases are discovered incidentally on endoscopy. The lesions are usually low grade and have an indolent clinical course, and most

patients have been managed without therapy. Local recurrences in the intestine may occur but spread beyond the small intestine is rare.

Mucosa-Associated Lymphoid Tissue Lymphoma

Most lymphomas of marginal zone derivation present in extranodal sites and are part of the spectrum of MALT lymphomas. MALT lymphomas occur most frequently in the stomach, lung, thyroid gland, salivary gland, and lacrimal gland. Other less common sites of involvement include the orbit, breast, conjunctiva, bladder, kidney, and thymus. MALT lymphomas are characterized by a heterogeneous cellular composition, including centrocyte-like cells, monocytoid B cells, small lymphocytes, and plasma cells. In most cases, large transformed cells are uncommon, but reactive germinal centers are nearly always present. Historically, the distinction from reactive lesions has been problematic. Clonality can be established based on light-chain restriction or molecular studies. Follicular colonization by the neoplastic cells can simulate FL. The clinical course is usually quite indolent, but MALT lymphomas tend to relapse in other MALT-associated sites. MALT lymphomas of the salivary gland, thyroid gland, and mediastinum (thymus) are usually associated with a history of autoimmune diseases, predominantly Sjögren syndrome.

MALT lymphomas have several recurring cytogenetic abnormalities, including t(11;18)(q21;q21), t(1;14)(p22;q32), t(14;18)(q32;q21), t(3;14)(q27;q32), and t(3;14)(p14.1;q32); these abnormalities are observed with variable frequency, often depending on the anatomic site.[18] Although several genes are involved in these translocations, at least three of them—t(11;18), t(1;14), and t(14;18)—share a common pathway, leading to the activation of nuclear factor κB (NF-κB). By genome-wide DNA profiling integrated with GEP, differences were detected among the three main types of marginal zone lymphoma (MZL) (MALT, nodal and splenic), lending support to the current WHO classification that separates these three entities.[19] The translocation t(11;18)(q21;q21) is associated exclusively with low-grade extranodal MALT and is not detected in cases with concurrent low-grade and high-grade tumors, or in primary extranodal large-cell lymphomas—raising doubts whether these primary extranodal lymphomas are, in fact, related to low-grade MALT. The term *extranodal marginal zone lymphoma* of MALT type applies only to low-grade MALT; the term *high-grade MALT* should not be used for extranodal large B-cell lymphomas in a MALT site.

There is a strong association between chronic infection with *Helicobacter pylori* and gastric MALT lymphoma. Other infectious agents have been described in MALT lymphomas involving skin *(Borrelia burgdorferi)*, ocular adnexae *(Chlamydia psittaci)*, and the small intestine *(Campylobacter jejuni);* however, a causal relationship has not yet been demonstrated. Chronic antigenic stimulation is critical to both the development of MALT lymphoma and the maintenance of the neoplastic state. Indeed, in some cases lacking these genetic aberrations, eradication of *H. pylori* by antibiotic therapy has led to the spontaneous remission of gastric MALT lymphoma. MALT lymphomas are positive for B cell–associated antigens but are usually negative for CD5 and CD10. BCL6 and CD10 are helpful markers to identify residual reactive germinal-center cells, especially in cases of follicular colonization. The putative cell of origin of MZL is a post–germinal-center memory B cell.

Nodal Marginal Zone Lymphoma

Nodal marginal zone lymphoma (NMZL), a primary nodal disease, is phenotypically similar to other MZL, extranodal or splenic types. Adult patients often present with bone marrow involvement and tend to have a more aggressive clinical course than those with extranodal MALT. The neoplastic proliferation is mostly composed of small to medium-sized B cells, often with pale cytoplasm and the immunophenotype is $CD20^+$, $CD5^-$, $CD10^-$, with variable IgD expression. Plasmacytoid differentiation may be present, and, in lymph nodes, follicular colonization can be seen.

Splenic Marginal Zone Lymphoma

Splenic marginal zone lymphomas (SMZLs) present in adults and have a slight female gender predilection, usually with splenomegaly but without peripheral lymphadenopathy. Most patients have bone marrow involvement with modest lymphocytosis. Some evidence of plasmacytoid differentiation and the presence of a low-level M ptotein may also be seen. The course is reportedly indolent, and splenectomy may be followed by a prolonged remission.

Histologically, the spleen shows expansion of the white pulp with a characteristic biphasic pattern—central zone of small lymphocytes surrounded by a peripheral zone of larger cells resembling marginal zone cells while an intact mantle is not present. The abundant pale cytoplasm evident in tissue sections may also be seen in peripheral blood smears, and the cytologic features may be mistaken for those of hairy cell leukemia. The phenotype resembles other marginal zone B-cell lymphomas with a more frequent IgD expression.

CLINICAL PEARLS

Aggressive Lymphomas

- Natural history: survival measured in months.
- Successful therapy can be achieved with combination chemotherapy.
- Relapses from chemotherapy-induced remission may be cured with high-dose chemotherapy with hematopoietic support.
- In addition to the International Prognostic Index (IPI), gene expression profiling can be useful in predicting prognosis and survival of individual patients.

Diffuse Large B-Cell Lymphoma, Not Otherwise Specified

DLBCL, not otherwise specified (DLBCL, NOS) is one of the more common subtypes of NHL, representing up to 40% of cases. This diagnosis is used for both primary DLBCL as well as for cases involving transformation from a low-grade lymphoma. It may be nodal or involve extranodal sites, including bone, skin, thyroid gland, GI tract, and lungs.

DLBCL, NOS is composed of large transformed lymphoid cells with nuclei at least twice the size of a small lymphocyte (Fig. 78.3). The nuclei generally have vesicular chromatin, prominent nucleoli, and basophilic cytoplasm—resembling either centroblasts or immunoblasts, albeit with an overall greater cellular pleomorphism.

In terms of morphology and phenotype, DLBCL is one of the most heterogeneous categories in the WHO classification; currently, several morphologic variants as well as specific subtypes are recognized.[2] There has been a great interest in identifying DLBCL features that might be prognostically relevant. To address

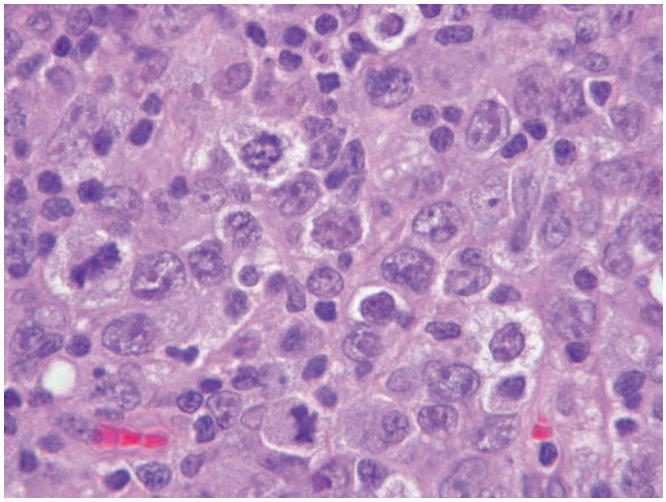


FIG. 78.3 Diffuse Large B-Cell Lymphoma. The neoplastic cells have large round to oval nuclei with vesicular chromatin and multiple eosinophilic nucleoli. Numerous mitoses are also present.

issues traditionally not resolved by morphologic or immunophenotypic features, DLBCL was among the first lymphomas to be analyzed by GEP. Based on the differential expression of a large set of genes by GEP, *germinal center B cell–like* (GCB) group and *activated B cell–like* (ABC) group have been identified,[20] in addition to the previously recognized *primary mediastinal (thymic) large B-cell lymphoma* (PMBCL). GCB DLBCLs express a set of genes that are associated with normal germinal center B cells, whereas ABC DLBCLs show downregulation of these genes and share similarities with post–germinal center B cells.

The t(14;18) translocation involving *BCL2* and Ig heavy-chain gene has been detected in the GCB subtype but not in the other subtypes. Previous studies have shown reduced disease-free survival in DLBCL cases with BCL2 overexpression, irrespective of the translocation. Although no absolute correlation between morphology and GEP has been established, the majority of DLBCLs with centroblastic morphology fall into the GCB subtype, whereas those with immunoblastic morphology usually correlate with the ABC subtype.

DLBCL has an aggressive natural history but generally responds well to chemotherapy. The complete remission rate with modern regimens is 75% to 80%. Currently, with R-CHOP (rituximab–cyclophosphamide, doxorubicin, vincristine, and prednisone) regimen, the 10-year progression-free and overall survival rates for older patients with advanced-stage DLBCL are 36.5% and 43.5%, respectively.[21]

Immunophenotype-based algorithms using CD10/BCL6 positivity for GCB and MUM1/IRF4 expression for the ABC subtype have been proposed as surrogates for GEP; in addition, BCL2 and IPI are also informative to stratify the DLBCL cases. ABC DLBCLs show constitutively activated NF-κB and mutations in the B-cell receptor (BCR) signaling pathway with poorer overall survival when compared with GCB DLBCL using R-CHOP. The addition of BTK inhibitors has improved the survival in those with ABC DLBCL since these drugs target the BCR signaling pathway. Because of the therapeutic impact, it has been recommended that these two subtypes be identified, on the basis of either immunophenotype-based algorithms or GEP.

T-cell/histiocyte–rich large B-cell lymphoma (THRLBCL) is a distinct clinicopathologic entity, rather than a morphologic variant. THRLBCL tends to occur in younger patients compared with other DLBCL, NOS, and often presents with advanced stage and bone marrow involvement with an aggressive clinical behavior.

The WHO classification recognizes that lymphomas arising from certain anatomic sites may have distinctive clinical and biologic features. Among these are *primary DLBCL of the central nervous system* (CNS) and *primary cutaneous DLBCL, leg type*. Primary DLBCL of the CNS with some distinctive GEP features shares some similarities with DLBCL arising in other immune-privileged sites, such as the testis. These lymphomas typically show biallelic *CDKN2A* loss and *MYD88* mutations and have frequent copy number gains in 9p24.1 *(PD-L1* and *PD-L2)* contributing to immune evasion of the tumor cells.[22] Primary cutaneous DLBCL, leg type, has a GEP resembling the ABC subtype of DLBCL, presents most often in older females, and generally has an aggressive clinical course.

A recently described provisional entity usually manifests a follicular growth pattern but may have diffuse areas. This lesion, termed *large B-cell lymphoma with IRF4 rearrangement* is common in children or young adults, typically involves the Waldeyer ring and/or the cervical lymph nodes. These cases show strong IRF4/MUM1 expression, usually with BCL6 and a high proliferative fraction, and are considered more aggressive than other pediatric-type FL; however, when treated, they show good response.[23,24]

Primary Mediastinal Large B-Cell Lymphoma

PMBCL has a distinct constellation of clinical and morphologic features. PMBCL shows marked female predominance in adolescents and young adults. It presents with a rapidly growing anterior mediastinal mass with frequent superior vena cava syndrome and/or airway obstruction. Nodal involvement is uncommon at presentation and at relapse. Frequent extranodal sites of involvement, particularly at relapse, include the liver, kidneys, adrenal glands, ovaries, the GI tract, and the CNS. The treatment approach includes aggressive systemic chemotherapy plus rituximab, along with radiation therapy used in some centers. A relatively abundant pale cytoplasm with distinct cytoplasmic membrane and fine compartmentalizing sclerosis are characteristic. The tumor appears to be derived from medullary B cells within the thymus gland. These cells express CD20 and CD79a but not surface Ig. CD23 is frequently positive, and MUM-1/IRF4 coexpression is also common. A unique signature was identified by GEP in PMBCL that shared similarities with Hodgkin lymphoma (HL) cell lines, including constitutive activation of the NF-κB and recurrent gains and amplification of c-Rel.[25]

The 2008 WHO classification included borderline categories, one of which manifests features intermediate between DLBCL, especially PMBCL and classic Hodgkin lymphoma (CHL) "gray zone lymphomas."[26] A close relationship between PMBCL and CHL has been supported by GEP.[25]

Gray zone lymphomas are more common in males than females, present with bulky mediastinal masses, and appear to have a more aggressive clinical course compared with either PMBCL or CHL. Previously, methylation profiling showed a signature distinct from both CHL and PMBCL.[27] GEP of gray zone lymphomas in comparison with CHL and PMBCL shows that they are distinct from the other entities with downregulation of the B-cell program and higher NF-κB as compared with PMBCL.[28] However, as shown by FISH studies, gray zone lymphomas, PMBCL, and CHL all share a number of common cytogenetic aberrations, including gains at 2p16.1 (*REL/BCL11A*

locus), 9p24.1 *(JAK2/PDL2)*, and rearrangements of 16p13.13 *(CIITA)*. DA-EPOCH R has been effective in both PMBL and DLBCL; however, the therapeutic approach is not clear, although combined modality therapy appears to be beneficial.[29]

Virally Associated B-Cell Lymphoproliferative Disorders

Epstein-Barr virus (EBV)-positive B-cell lymphoproliferations show a range in cytologic features and clinical behavior. In the revised WHO 2016 classification, *EBV-positive DLBCL, NOS*[3] has now replaced a previous provisional entity "*EBV-positive DLBCL of the elderly*" in the 2008 WHO classification because it may occur in younger patients with a broader morphologic spectrum and better survival[30,31] as a consequence of decreased immune surveillance. In this entity, both nodal and extranodal sites may be involved. Large transformed neoplastic cells often show immunoblast and/or Hodgkin/Reed-Sternberg (HRS)-like cell features, in a background consisting of interspersed small lymphocytes, histiocytes, and epithelioid cells. In other cases, a monomorphic pattern is present. (Fig. 78.4)

These should be distinguished from EBV-associated atypical hyperplasia and lesions associated with iatrogenic or age-related immunosuppression, as well as those typified by an isolated and circumscribed cutaneous or mucosal presentation, with an indolent behavior and a self-limiting clinical course as *EBV-positive mucocutaneous ulcer* (new provisional entity).[32] Lesions are most common in the oral mucosa, skin, and GI tract. Ulcerated surfaces have an underlying polymorphic infiltrate with large transformed cells resembling immunoblasts and HRS-like cells (Fig. 78.5). Patients rarely present with lymphadenopathy or systemic symptoms.

Lymphomatoid granulomatosis is an EBV-positive B-cell lymphoproliferative disorder (LPD) associated with an inflammatory background rich in T cells. The lung is nearly always involved; in addition, skin, kidneys, the liver, and the brain are frequently affected.[33] *DLBCL associated with chronic inflammation* are EBV-driven large B-cell proliferations encountered in diverse clinical settings, usually associated with a confined anatomical space and a background of chronic inflammation. These cases appear to have a good prognosis if successfully resected.

Human herpes virus-8/Kaposi sarcoma herpes virus (HHV-8/KSHV)-associated LPDs are also documented. These include *primary effusion lymphoma* (PEL), *multicentric Castleman disease* (MCD), and lymphomas arising in the context of MCD. The cells of PEL are usually coinfected with EBV, and the disease is most often diagnosed in the setting of human immunodeficiency virus (HIV) infection and immunosuppression. Although pleural or peritoneal effusions are most common, extracavitary PEL can present as a tumor mass, usually in extranodal sites. PEL has a phenotype resembling terminally differentiated B cells (*i.e.*, plasmablastic).

Two other lymphomas with a plasmablastic phenotype are *plasmablastic lymphoma* (PBL) and *anaplastic lymphoma kinase (ALK)-positive large B-cell lymphoma*. PBL is usually positive for EBV, most often extranodal, and is associated with immunosuppression from either HIV infection or advanced age. Recent studies have identified a high incidence of *MYC* translocation in PBL.[34] ALK-positive large B-cell lymphomas show overexpression of ALK, usually as a consequence of translocation. They mainly affect older individuals but can occur at any age.

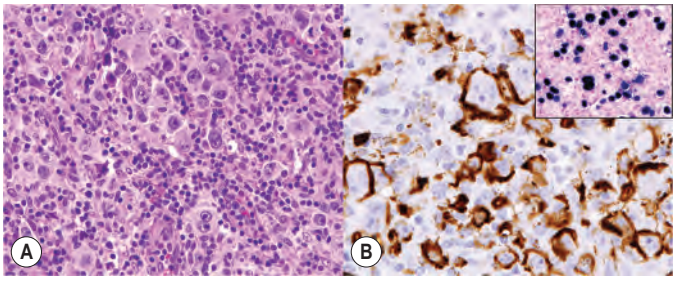


FIG. 78.4 Epstein-Barr virus (EBV)-positive diffuse large B-cell lymphoma, NOS. (A) There is a polymorphous population with scattered large atypical cells. (B) The large cells are positive for CD20 (variable) and EBV by in situ hybridization with EBER (*inset*).

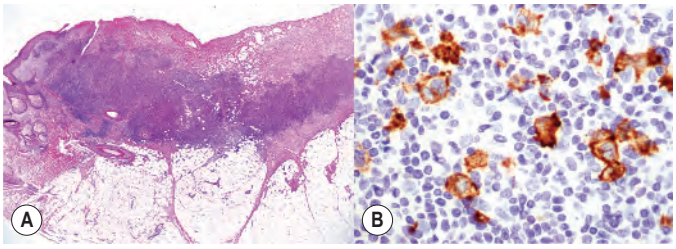


FIG. 78.5 Mucocutaneous Ulcer. (A) Low-power magnification shows an ulcerated surface with a dense lymphoid infiltrate underneath. (B) Higher magnification shows a dense polymorphic infiltrate with histiocytes and large atypical cells resembling immunoblasts and Hodgkin/Reed-Sternberg–like cells that are positive for CD30.

CLINICAL PEARLS

Highly Aggressive Lymphomas

- More common in children
- Natural history similar to acute leukemia
- Successful treatment includes high-dose chemotherapy (induction, consolidation, and maintenance phases) with central nervous system prophylaxis

Burkitt Lymphoma

Burkitt lymphoma (BL) occurs most commonly in children and accounts for up to one-third of all pediatric lymphomas in the United States. Three clinical variants of BL are recognized—endemic, sporadic, and immunodeficiency associated. The endemic cases are seen in equatorial Africa (African BL) in malaria-endemic regions and mostly involve the jaw and other facial structures. In non–malaria-endemic regions, such as the United States, extranodal sites are frequent, including the ileocecal region, ovaries, kidneys, or breasts. Bone marrow involvement is a poor prognostic feature.

EBV has been shown to be a cofactor for the development of BL and shows varying degrees of positivity in the variant subtypes. BL is one of the more common tumors associated with HIV. It can present at any time during the clinical course, including the initial acquired immunodeficiency syndrome (AIDS)-defining illness. GEP data have supported a common pathogenetic mechanism in cases of HIV infection and endemic malaria–related immunosuppression.

Cytologically, BL shows monomorphic medium-sized cells with round nuclei, multiple (2 to 5) basophilic nucleoli, and moderately abundant deeply basophilic cytoplasm. Cytoplasmic lipid vacuoles reflecting a high proliferation rate and apoptosis are frequent. It is the most rapidly growing of all lymphomas, with 100% of the cells in cell cycle at any time. The characteristic

"starry sky" pattern of BL is a manifestation of the numerous benign macrophages that have ingested karyorrhectic or apoptotic tumor cells. BL has a mature B-cell phenotype, expressing CD19, CD20, CD22, CD79a, and monoclonal surface Ig that is nearly always IgM. CD10 is positive in almost all cases, whereas CD5, CD23, and BCL2 are consistently absent.

The pathogenesis of BL is related to the *MYC* oncogene translocations seen in virtually 100% of cases. The *MYC* translocation is considered a primary event and often is the sole karyotypic abnormality detected. This is in contrast to other aggressive lymphomas, in which the *MYC* translocation occurs as a secondary event in a more complex karyotype.[35] Most of the translocations involve the *IGH* gene on chromosome 14 and, less commonly, the light-chain genes on chromosomes 2 and 22. Mutations in the transcription factor *TCF3* or its negative regulator *ID3* are present in approximately 70% of sporadic and immunodeficiency-related BL and 40% of endemic cases.

A subset with chromosome 11q alterations has a new provisional entity "*Burkitt-like lymphoma with 11q aberration*" in the revised classification. The malignant cells resemble BL morphologically, largely phenotypically, and by GEP, but lack *MYC* translocations.[36]

High-Grade B-Cell Lymphoma With *MYC* and *BCL2* and/or *BCL6* Rearrangements

This category attempts to unify aggressive lymphomas carrying rearrangements of *MYC*, as well as *BCL2* or *BCL6*, so-called double- or triple-hit lymphoma.[2] These cases can resemble DLBCL, NOS or have features intermediate between DLBCL and BL (Fig. 78.6). Cases of FL and B lymphoblastic lymphoma that have these rearrangements should be excluded. A less well-defined group is high-grade B-cell lymphoma, NOS. Some of these cases have blastoid morphology, but unifying features are lacking.

T-CELL AND NK-CELL NEOPLASMS

Overview of the Classification of T-Cell Neoplasms

*Peripheral T-cell lymphoma*s (PTCLs) are uncommon, representing fewer than 10% of all NHLs. The classification of PTCL has always been controversial. The genetic landscape of many entities has only recently been defined, and immunophenotypic markers are less specific for apparently distinct disease entities. For these reasons, the WHO classification relied, to a considerable extent, on clinical presentation to subdivide these tumors.[2]

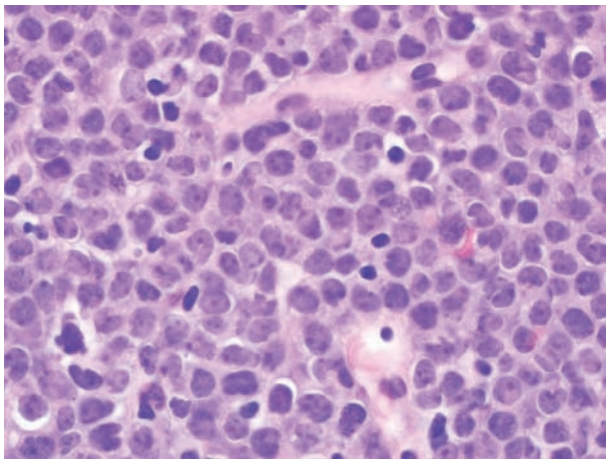


FIG. 78.6 Triple-Hit Lymphoma. A case of high grade B-cell lymphoma with *MYC, BCL2,* and *BCL6* translocations. The tumor cells are monotonous and medium in size with dispersed chromatin and small nucleoli.

Extranodal Natural Killer/T-Cell Lymphoma, Nasal Type

Extranodal NK/T-cell lymphoma, nasal type, is a distinct clinicopathologic entity highly associated with EBV. It affects adults (median age 50 years), and the most common clinical presentation is a destructive nasal or midline facial lesion. Palatal destruction, orbital swelling, and edema can be prominent. NK/T-cell lymphomas have been reported in other extranodal sites, including skin, soft tissue, testes, the upper respiratory tract, and the GI tract. It is much more common in Asians and indigenous populations of the Americas than in those of European background, and this indicates that genetic factors play a role in the pathogenesis of these lymphomas.

Extranodal NK/T-cell lymphoma, nasal type, is characterized by a broad cytologic spectrum. Although the cells express some T cell–associated antigens, most commonly CD2, other T-cell markers, such as surface CD3, are usually absent. Cytoplasmic CD3 is positive, but most cases lack clonal T-cell gene rearrangement. In favor of an NK-cell origin, the cells are nearly always CD56 positive, although CD16 and CD57 are usually negative. EBV is invariably positive as shown by in situ hybridization.

The clinical features and treatment response of non–nasal-type extranodal NK/T-cell lymphoma are different from of those of nasal presentation of this lymphoma in that the addition of radiotherapy in early-stage nasal cases may offer a survival benefit.[37] A hemophagocytic syndrome is a common clinical complication and adversely affects survival. Emerging oncogenic pathways have been identified by GEP.

There are other EBV$^+$ T-cell and NK-cell proliferations seen mainly in Asian children and in indigenous populations from Central and South Americas and Mexico. These show a broad range of clinical manifestations from indolent, localized forms involving skin, such as *hydroa vacciniforme-like lymphoproliferative disorder* (name changed from lymphoma) and *mosquito bite allergy*, the latter usually being derived from NK cells, to *systemic EBV$^+$ T-cell lymphoma* (name changed from lymphoproliferative disease),[3] which is a more systemic form characterized by fever, hepatosplenomegaly and lymphadenopathy with or without cutaneous manifestations[38] and a fulminant clinical course, and hemophagocytic lymphohistiocytosis (HLH). The neoplastic cells are usually positive for CD2, CD3, CD8, TIA-1, and Epstein-Barr virus-encoded small RNA (EBER) and negative for CD56 and most commonly infiltrate the liver, spleen, skin, lung, and lymph nodes (Fig. 78.7).

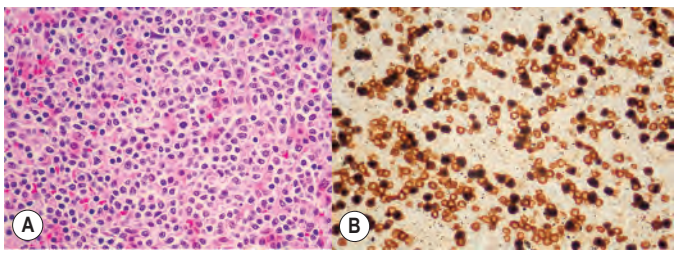


FIG. 78.7 Systemic Epstein-Barr Virus–Positive T-Cell Lymphoma of Childhood. (A) Spleen infiltration by cytologically atypical T cells that are (B) positive for CD3 and EBER-ISH double-staining.

Nodal T-Cell Lymphoma With T_{FH} Phenotype: Angioimmunoblastic T-Cell Lymphoma

Recently angioimmunoblastic T-cell lymphoma (AITL) has been grouped under an umbrella category "nodal T-cell lymphoma with T follicular helper cell (T_{FH}) phenotype" to highlight a spectrum of nodal lymphomas with T_{FH} phenotype, including follicular T-cell lymphoma and other nodal PTCLs with a T_{FH} phenotype, in addition to AITL.[3]

AITL presents in adults with generalized lymphadenopathy and prominent systemic symptoms, including fever, weight loss, and skin rash. Polyclonal hypergammaglobulinemia is usually seen.

Histologically, the nodal architecture is generally effaced, but peripheral sinuses are often open or dilated. Proliferation of high endothelial venules (HEVs) is often prominent. Follicles are regressed, but a proliferation of dendritic cells (DCs) often abutting HEVs is typically seen. The atypical lymphoid cells have clear cytoplasm and are admixed with small lymphocytes, immunoblasts, plasma cells, and histiocytes, with or without eosinophils. A relationship to T_{FH} has been confirmed by GEP. The atypical T cells are usually positive for CD4, CD10, CXCL-13, and PD-1, a phenotype characteristic of T_{FH}. In keeping with an established derivation from T_{FH}, B-cell proliferation, including marked polyclonal plasmacytosis, is often seen. In some cases, the plasma cells may be monoclonal. Background EBV^+ B cells are almost constant, and progression to EBV^+ B-cell lymphoma may occur. The exact role of EBV in AILT is uncertain; however, plausible theories include expansion resulting from decreased immune surveillance. The majority of cases shows clonal rearrangements of T-cell receptor (TCR) genes. Patients may initially respond to steroids or mild cytotoxic chemotherapy, but progression usually occurs. More aggressive combination chemotherapeutic regimens have led to a higher remission rate, but patients are prone to secondary infectious complications. The median survival is usually less than 5 years. Recurrent mutations include epigenetic modifiers *TET2, IDH2, DNMT3A*, and *RHOA*. Moreover, fusion genes encoding a CTLA4-CD28 and *ITK-SYK* have been reported in a significant proportion of PTCL, NOS cases with a T_{FH} phenotype including AITL.

Peripheral T-Cell Lymphoma, Not Otherwise Specified

PTCL, not otherwise specified (PTCL, NOS) is a diagnosis of exclusion with most cases being nodal in origin. PTCLs are characterized by cytologic and phenotypical heterogeneity. Cases have a mature T-cell phenotype and express one of the major subset antigens, with CD4 expression seen more frequently than CD8. These are not clonal markers, and antigen expression can change over time. Loss of one of the pan-T-cell antigens (CD3, CD5, CD2, or CD7) is seen in two-thirds of cases, with loss of CD7 being most frequent.

Recently, GEP has shown a global signature close to one of activated T lymphocytes and has identified at least three subtypes characterized by overexpression of *GATA3, TBX21*, and cytotoxic genes associated with differences in clinical behavior and response to therapy. The clinical course is generally aggressive, especially with a high proliferation signature and a lower response rate than that seen for aggressive B-cell lymphomas. A recent study investigating genomic copy number changes revealed that loss or mutation of genes in the *CDKN2A/B-TP53* and PTEN-PI3K pathways are a feature of *GATA3* expressing PTCL-NOS while co-occurring gains/amplifications of *STAT3* and *MYC* also occur in this subgroup. On the other hand, PTCL-NOS expressing *TBX21* showed mutations of genes regulating DNA methylation.[39,40]

Anaplastic Large-Cell Lymphoma

Anaplastic large-cell lymphoma (ALCL) is most common in children and young adults, with male predominance. Nodal presentations are most common; however, a variety of extranodal sites can be involved.

ALCL is characterized by pleomorphic or monomorphic cells that have a propensity to invade lymphoid sinuses. The cells of classic ALCL have large, often lobulated nuclei with small basophilic nucleoli. In some cases, the nuclei may be round. The cytoplasm is usually abundant and amphophilic and has distinct cytoplasmic borders, with a prominent Golgi region. The expression of the CD30 antigen is a hallmark of this disease (Fig. 78.8). However, CD30 expression is not specific for ALCL and may also be seen in other forms of malignant lymphoma, including CHL. ALCL-ALK^+ is associated with a characteristic chromosomal translocation, t(2;5)(p23;q35), involving the *ALK* and *NPM* genes, respectively. A variety of other *ALK* partners has been identified, and monoclonal antibodies to the ALK protein have been able to identify tumor cells regardless of the underlying translocation. Staining of both the nucleus and cytoplasm is found in cases with the classic *NPM:ALK*; cases with variant translocations show only cytoplasmic staining.

Immunohistochemistry is indispensable in the correct diagnosis of ALCL. The prominent Golgi region usually shows intense staining for CD30 and epithelial membrane antigen (EMA). The cells exhibit an aberrant phenotype with loss of many T cell–associated antigens. Both CD3 and CD45RO, the most widely used pan-T-cell markers, are negative in greater than 50% of cases, while CD2 and CD4 are positive in the majority, whereas CD8 is usually negative. ALCL cells, despite the $CD4^+/CD8^-$ phenotype, often express the cytotoxic-associated antigens TIA-1, granzyme B, and perforin. In addition, clusterin is generally present in ALCL and represents another useful diagnostic marker. Molecular studies in most cases demonstrate TCR rearrangement, confirming a T-cell origin.

Improved criteria now exist for the recognition of ALK-negative ALCL as a separate entity. It occurs in an older age group compared with the ALK-positive cases. Recently, a unique form of ALK-negative ALCL arising in association with breast implants has been identified[41] with a median interval from time of implant to lymphoma of approximately 10 years. Confinement of the neoplastic cells to the seroma fluid without capsule invasion portends a favorable prognosis.

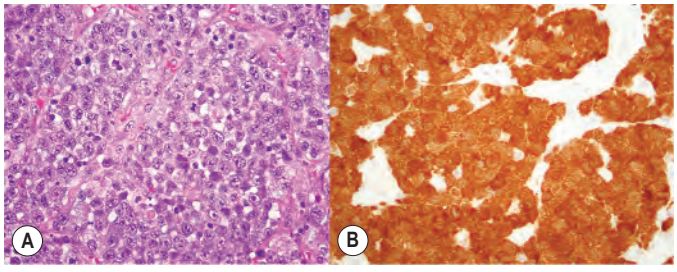


FIG. 78.8 Anaplastic Large-Cell Lymphoma, ALK-Positive. (A) The cells have large, lobulated nuclei with small basophilic nucleoli. The cytoplasm is usually abundant and amphophilic and has distinct cytoplasmic borders, and a prominent Golgi region is generally present. (B) ALK-1 expression is a hallmark of this disease.

GEP studies have shown that ALK-negative ALCLs have a signature close to that of ALK-positive counterparts but distinct from other NK/T-cell lymphomas. Genetic studies have shown convergent mutations and kinase fusions leading to constitutive activation of the JAK/STAT3 pathway. The overall survival and disease-free survival are significantly better in ALK-positive cases than in ALK-negative cases. Clinical or prognostic variations exist in ALK-negative ALCLs—a subset with rearrangements at the *DUSP22* and *IFR4* locus on 6p25 has a superior prognosis, whereas a small subset with *TP63* rearrangements is very aggressive.[42] Loss of *PRDM1 (BLIMP1)* and/or *TP53* has also been associated with poor outcome.

Primary Cutaneous Anaplastic Large-Cell Lymphoma

Primary cutaneous ALCL (pcALCL) is closely related to lymphomatoid papulosis (LyP) and differs at the clinical, immunophenotypic, and molecular levels from the systemic form. Indeed, LyP and cutaneous ALCL represent a histologic or clinical continuum of $CD30^+$ cutaneous lymphoproliferative diseases. Small lesions are likely to regress, whereas patients with large tumor masses may develop disseminated disease with lymph node involvement—a period of observation is usually warranted before the institution of any chemotherapy for isolated lesions. Most patients with primary cutaneous ALCL have multiple skin lesions, but it is a more indolent disease compared with other T-cell lymphomas of the skin. Cutaneous ALCL is $CD30^+$ but is usually ALK and EMA negative and also lacks the t(2;5) translocation. Most cases of pcALCL have clonally rearranged TCR genes and approximately 28% of cases harbor translocations involving *DUSP22*.

LyP is characterized by a chronic course of recurrent, self-healing papulonecrotic or nodular skin lesions. The histologic picture of LyP is variable and may resemble different types of CD30-positive CTCLs. Appreciation of these variants is important as they can mimic aggressive T-cell lymphoma histologically but clinically are similar to other forms of LyP.

Subcutaneous Panniculitis-Like T-Cell Lymphoma

Subcutaneous panniculitis-like T-cell lymphoma (SPTCL) usually presents with subcutaneous nodules primarily affecting the extremities. In some cases, the infiltrate may appear deceptively benign, mimicking lobular panniculitis. The neoplastic cells are CD8 cytotoxic α/β T cells (perforin, granzyme B, and TIA-1 positive), while EBV is negative. Some PTCL of γ/δ T-cell derivation may show similar features but differ from SPTCL in their clinical behavior (more aggressive) and histologic pattern because they are often not confined to the subcutis. Some patients have a history of autoimmune disease, and, in particular, the differential diagnosis of SPTCL with lupus profundus panniculitis can be challenging. Unlike SPTCL, lupus panniculitis usually contains abundant plasma cells, a mixture of CD4 and CD8 cells, and a relative increase of γ/δ T cells and plasmacytoid DCs.

A loss-of-function germline mutation altering TIM-3 was identified in up to 60% of SPTCL, and this TIM-3 deficiency leads to uncontrolled immune activation.[43] Patients present with fever, pancytopenia, hepatosplenomegaly, and HLH. HLH is most readily diagnosed in bone marrow aspirate smears, where histiocytes containing erythrocytes, platelets, and other blood elements may be observed.

Primary Cutaneous γ/δ T-Cell Lymphoma

Primary cutaneous γ/δ T-cell lymphomas are clinically aggressive tumors that can present with involvement of the subcutis and the dermis with or without epidermal infiltration. A hemophagocytic syndrome may be seen and is more common than in SPTCL. Skin is the most common presenting site; however, similar lymphomas of γ/δ T-cell origin can present in other extranodal sites, including the GI tract and lungs. The cells have a cytotoxic phenotype and express cytotoxic molecules and, like normal γ/δ T cells, lack CD5. CD8 may be positive but more often is negative (in which case the T cells are negative for both CD4 and CD8) (Fig. 78.9) while most cases express CD56. Few cutaneous T-cell lymphomas appear to be TCR silent (negative for both TCR-β and TCR-γ) but share the same aggressive clinical behavior.

Mycosis Fungoides and Sézary Syndrome

Mycosis fungoides (MF) and Sézary syndrome (SS) are closely related and often considered together from a clinical and biologic standpoint but are now regarded as separate diseases. Both are primary cutaneous T-cell malignancies derived from mature skin homing CD4 T cells. Epidermotropism is the hallmark of MF; infiltration of the epidermis produces characteristic Pautrier microabscesses. The cutaneous lesions are categorized as patches, plaques, and tumors, based on the extent of the infiltrate, and, clinically, SS is more aggressive than MF. A γ/δ indolent variant of MF has also been described. SS presents with exfoliative erythroderma and circulating cerebriform lymphocytes known as *Sézary cells*.

Intestinal T-cell Lymphoma

Enteropathy-associated T-cell lymphoma (EATL) is highly associated with celiac disease on a worldwide basis. This disease occurs in adults; the majority has either overt or clinically silent gluten-sensitive enteropathy (Chapter 75). Ulcerative jejunitis may precede the development of overt EATL and may share a common clonal T-cell population as with the subsequent lymphoma. The small bowel usually shows ulceration with frequent perforation, which may or may not be accompanied by a mass. EATL shows a cytologic composition of variably sized or polymorphic atypical lymphoid cells. The adjacent small bowel usually shows villous atrophy associated with celiac disease. The neoplastic cells are $CD3^+$, $CD7^+$ T cells that also express the homing receptor CD103. Anaplastic cells strongly positive for CD30 can be present. The cells express cytotoxic molecules, a feature shared by nearly all extranodal T-cell lymphomas. The majority belong to the α/β TCR subset, whereas only a minority expresses the γ/δ TCR. Other PTCLs can present with intestinal disease, including EBV^+ extranodal T/NK-cell lymphomas and γ/δ T-cell lymphomas, and should be distinguished from EATL. The clinical course of EATL is aggressive.

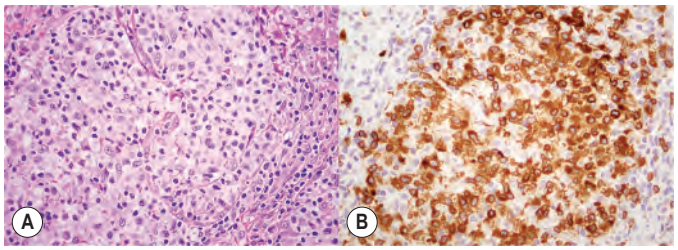


FIG. 78.9 Primary Cutaneous γ/δ T-Cell Lymphoma. (A) The lesions usually involve the dermal and subcutaneous tissue and are composed of atypical cells with irregular nuclear contours. (B) They have a γ/δ phenotype highlighted by the T-cell receptor δ immunostain.

The disease previously known as *EATL type II* has been now been formally designated as *monomorphic epitheliotropic intestinal T-cell lymphoma* (MEITL); it shows no association with celiac disease and therefore is the form most often seen in Asian and Hispanic populations. Unlike classic EATL, MEITL is monomorphic, usually positive for CD8, CD56, and MATK (Fig. 78.10). *STAT5B* and *SETD2* mutations are common in MEITL cases, many of which are of γ/δ T-cell origin.[44]

Hepatosplenic T-Cell Lymphoma

Hepatosplenic T-cell lymphoma (HSTCL) shows a marked male gender predominance, and most patients are young adults. The clinical presentation is that of marked hepatosplenomegaly in the absence of lymphadenopathy.

The liver and the spleen show marked sinusoidal infiltration, with sparing of both portal triads and white pulp, respectively. The homing pattern manifested by the malignant cells is similar to that of normal γ/δ T cells, which also populate the sinusoidal areas of the spleen. The neoplastic cells have a phenotype that resembles normal γ/δ T cells, usually expressing neither CD4 nor CD8 while CD56 is often positive. The cells are positive for the cytotoxic protein TIA-1 but are not activated and generally lack granzyme B and perforin. In situ hybridization for EBV is negative. The recognition of the atypical cells in bone marrow is greatly aided by immunohistochemical stains for CD3 showing a sinusoidal pattern of infiltration.

Isochromosome 7 is a consistent cytogenetic abnormality, usually seen in conjunction with trisomy 8. Rare cases derived from α/β T cells but with similar morphologic and biologic features can be seen. Recurrent mutations of *STAT5B*, and less often, *STAT3* and *PIK3CD*, are seen in HSTCL of γ/δ origin.[45] Clinically, HSTCL is aggressive, and, although patients may respond initially to chemotherapy, relapse occurs frequently. The median survival is less than 3 years. Allogeneic bone marrow transplantation is required for sustained remission.

Adult T-Cell Leukemia/Lymphoma

Adult T-cell leukemia/lymphoma (ATLL) is a distinct form of T-cell lymphoma associated with the retrovirus human T-lymphotropic virus-1 (HTLV-1). The highest number of cases is seen in southwestern Japan and the Caribbean basin. The disease has a long latency, and affected individuals are usually exposed to the virus very early in life. The virus may be transmitted in breast milk and through exposure to blood or blood products. The cumulative incidence of ATLL is estimated to be 2.5% among HTLV-1 carriers. The virus is clonally integrated into tumor DNA. Different clinical variants of ATLL, including the acute, lymphomatous, chronic, and smoldering types, have been recognized. Cutaneous involvement is seen in the majority of patients.

Peripheral blood involvement is very common, often without bone marrow disease. The atypical cells are often markedly polylobated and have been referred to as "flower" cells. The cells have a characteristic phenotype that resembles T regulatory cells (Tregs)—$CD3^+$, $CD4^+$, and $CD25^+$—but FOXP3 is expressed only in a minority of tumor cells. The function of the tumor cells as Tregs may correlate with the associated immunodeficiency.

Somatic gain of function *CCR4* mutations have been implicated in the pathogenesis of ATLL.[46]

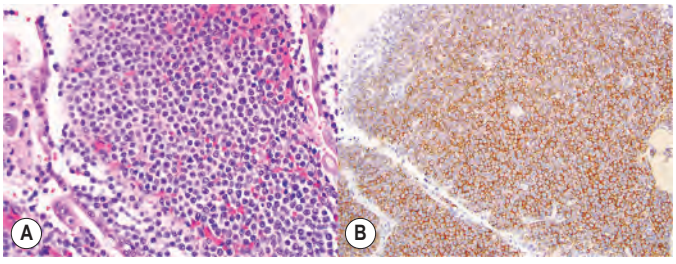


FIG. 78.10 Monomorphic Epitheliotropic Intestinal T-Cell Lymphoma. (A) Monomorphic population of medium-sized cells in the submucosa of the intestine with dispersed chromatin and indistinct nucleoli. (B) The cells are positive with CD56.

CLINICAL PEARLS

Hodgkin Lymphoma

- B-cell lineage established in nearly all cases
- Reed-Sternberg cell, the hallmark of the disease, represents a "crippled" germinal center B cell
- Nodular lymphocyte-predominant Hodgkin lymphoma considered a related but distinct entity
- Eighty percent of patients are curable with current therapy
- Stage of disease guides the choice of therapy; even patients with advanced-stage disease may be cured
- Late complications from treatment include acute leukemia (associated with alkylating agents with extended-field radiation therapy), second solid tumors (radiation therapy), and premature atherosclerotic coronary artery disease (radiation therapy)
- Cause of death in the first 5–10 years is mainly Hodgkin lymphoma; after 10 years, mainly secondary malignant tumors

HODGKIN LYMPHOMAS

HL and NHL have long been regarded as distinct disease entities on the basis of their differences in pathology, phenotype, clinical features, and response to therapy. It is now accepted that the malignant cell of HL is an altered B cell. Thus the current preferred term is *Hodgkin lymphoma*, rather than *Hodgkin disease*, reflecting current knowledge regarding the nature of the cell of origin. Despite the close histogenetic relationship with NHL, these disorders are still treated with different modalities.

Classic Hodgkin Lymphoma

The diagnosis of CHL depends on the identification of HRS cells in an appropriate inflammatory background composed of small T lymphocytes, plasma cells, histiocytes, and granulocytes (often eosinophils). All cases of CHL share certain immunophenotypic and genotypic features. The phenotype is $CD30^+$, $CD15^{+/-}$, $CD45^-$, and EMA^-. Expression of B cell–associated antigens is seen in up to 75% of cases; however, when present, CD20 staining is usually weaker or more variable than that seen in normal B cells, and PAX5 is dimly expressed. Ig and TCR genes are usually germline because of the paucity of tumor cells in the inflammatory background; however, microdissection can enable amplification for clonal rearrangement of the Ig genes by polymerase chain reaction (PCR). In addition, the presence of somatic mutations indicates transit through the germinal center.

Classic Hodgkin Lymphoma, Nodular Sclerosis

This variant of HL, classic Hodgkin lymphoma, nodular sclerosis (CHLNS) is most commonly seen in adolescents and young adults but can occur at any age. It is more common in females

than males, the mediastinum is commonly involved, and stage and bulk of disease have prognostic importance.

The tumor has at least a partially nodular pattern, with fibrous bands separating the nodules in most cases. The characteristic cell is the "lacunar-type" Reed-Sternberg (RS) cell, which may be very numerous. Classic RS cells are usually also present. The background contains lymphocytes, histiocytes, plasma cells, eosinophils, and neutrophils. Grading (I and II) is based on the proportion of the tumor cells and the presence of necrosis but is considered optional. The immunophenotype and genotype are characteristic of CHL. However, EBV is infrequently positive, seen in less than 15% of cases.

CHLNS is often curable; however, in long-term survivors, the risk of secondary malignancies is increased, especially in those receiving both chemotherapy and radiation. CHLNS of the mediastinum is thought to be closely related to PMBCL, and both types of tumors can be seen in the same patient, either as composite malignancy or sequentially.

Classic Hodgkin Lymphoma, Mixed Cellularity

Classic Hodgkin lymphoma, mixed cellularity (CHLMC) is predominantly seen in male adults. A bimodal age distribution, with peaks in young children and later in older adults can be present. Both CHLMC and the lymphocyte-depleted form (see later) can be associated with underlying HIV infection. The infiltrate is diffuse without band-forming sclerosis, although fine interstitial fibrosis may be present (Fig. 78.11). HRS cells are of the classic type. It is often EBV positive, seen in up to 75% of cases. The stage is often advanced at diagnosis. The clinical course is moderately aggressive but is often curable.

Classic Hodgkin Lymphoma, Lymphocyte Depletion

Classic Hodgkin lymphoma, lymphocyte depletion (CHLLD) is the least common variant of CHL and is most common in older people, in HIV-positive individuals, and in populations of nonindustrialized countries. It frequently presents with abdominal lymphadenopathy, and spleen, liver, and bone marrow involvement but without peripheral adenopathy. The infiltrate is diffuse and often appears hypocellular because of the presence of diffuse fibrosis and necrosis. The immunophenotype is characteristic of CHL. Because the histologic differential diagnosis often includes B- or T-large-cell lymphoma or ALCL, immunohistochemistry should be performed in most cases. EBV is positive in the majority of cases, and the stage is usually advanced at diagnosis.

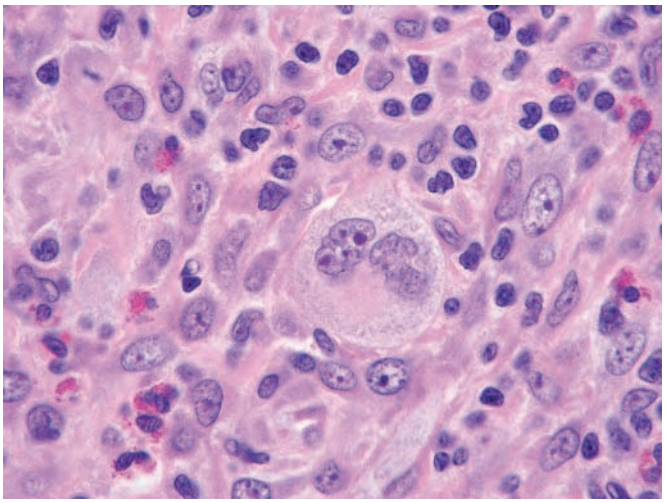


FIG. 78.11 Classic Hodgkin Lymphoma, Mixed Cellularity Subtype. A classic Reed-Sternberg cell is shown in a mixed inflammatory background with eosinophils, plasma cells, histiocytes, and small lymphocytes.

Classic Hodgkin Lymphoma, Lymphocyte-Rich

Classic Hodgkin lymphoma, lymphocyte-rich (CHLLR) may be nodular or diffuse and contains relatively infrequent classic HRS cells. Eosinophils and plasma cells are rare. In the nodular form, the HRS cells are seen at the periphery of B cell–rich nodules, mainly in the marginal zone. CHLLR has some features that are intermediate between other CHL and nodular lymphocyte-predominant Hodgkin lymphoma (NLPHL).[47] Immunophenotypically, the neoplastic cells resemble classic HRS cells, but morphologic distinction from LP cells may be difficult in some cases. Thus, in the past, many cases were misdiagnosed as NLPHL. However, patients usually present with localized disease and tend to be older than patients with NLPHL. The genetic features are similar to other variants of CHL.

Nodular Lymphocyte-Predominant Hodgkin Lymphoma

NLPHL is considered a distinct entity. Although it resembles other types of HL in having a minority of putative neoplastic cells in a background of benign inflammatory cells, it differs morphologically, immunophenotypically, and clinically from CHL.

NLPHL occurs at all ages but is more common in adult males. It usually involves peripheral lymph nodes with sparing of the mediastinum and is localized at diagnosis, although rarely it may be disseminated. NLPHL usually has a nodular growth pattern, with or without diffuse areas. The number of infiltrating reactive T cells is variable, and various patterns have been described on the basis of the cellular composition and growth pattern.[48] The atypical cells have vesicular, polylobated nuclei and small nucleoli. These had been called *lymphocytic and/or histiocytic (L&H) cells*, or "popcorn" cells, but the term *lymphocyte-predominant (LP) cell* is currently preferred. LP cells differ from classic HRS cells. The background is composed predominantly of lymphocytes with or without clusters of epithelioid histiocytes. Plasma cells are infrequent; eosinophils and neutrophils are also rare. The atypical cells are $CD45^+$, positive for B cell–associated antigens (CD19, CD20, CD22, CD79a), $CDw75^+$, $EMA^{+/-}$ $CD15^-$, $CD30^{-/+}$, and usually sIg^-. Small lymphocytes in the nodules are predominantly B cells with a mantle zone phenotype. However, numerous T cells are present, with $CD279^+$ T cells "rosetting" the LP cells. The proportion of T cells tends to increase over time in sequential biopsies. A prominent follicular DC meshwork is present within the nodules. LP cells, when microdissected, have shown to have clonally rearranged Ig genes with evidence of somatic hypermutation.

Survival for localized cases with or without treatment is long. However, cases with variant histology are more likely to have advanced-stage disease with a higher relapse rate. Moreover, patients with advanced-stage disease respond poorly to HL regimens, such as Adriamycin (doxorubicin), bleomycin, vinblastine, dacarbazine (ABVD), and benefit from treatment for aggressive B-cell lymphoma.[49]

ON THE HORIZON

- In recent years, there has been a greater appreciation of early events in lymphoid neoplasia.
- These early lesions in some ways can be considered equivalent to benign neoplasms in the epithelial system.
- These are clonal proliferations of B or T cells that carry genetic aberrations associated with specific forms of lymphoid neoplasia: chronic lymphocytic leukemia, multiple myeloma, follicular lymphoma, mantle cell lymphoma.
- Examples include monoclonal gammopathy of undetermined significance, monoclonal B lymphocytosis, in situ mantle cell neoplasia, in situ follicular neoplasia, lymphomatoid papulosis, patch stage of mycosis fungoides.
- Early lesions appear to lack the secondary and tertiary "hits" seen in lymphoid neoplasms that are clinically significant, and most patients have a very low risk of clinical progression.
- Current focus is to define the precise genetic features that distinguish early lesions from lymphoma, assess the risk of clinical progression, and determine their clinical management.

REFERENCES

1. Nogai H, Dorken B, Lenz G. Pathogenesis of non-Hodgkin's lymphoma. *J Clin Oncol*. 2011;29(14):1803–11. Epub 2011/04/13.
2. Swerdlow S, Campo E, Harris NL, et al. *WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues*. Lyon, France: International Agency for Research on Cancer; 2017.
3. Swerdlow SH, Campo E, Pileri SA, et al. The 2016 revision of the World Health Organization classification of lymphoid neoplasms. *Blood*. 2016;127(20):2375–90.
4. Campo E, Swerdlow SH, Harris NL, et al. The 2008 WHO classification of lymphoid neoplasms and beyond: evolving concepts and practical applications. *Blood*. 2011;117(19):5019–5032. Epub 2011/02/09.
5. Treon SP, Xu L, Yang G, et al. MYD88 L265P somatic mutation in Waldenstrom's macroglobulinemia. *N Engl J Med*. 2012;367(9):826–33.
6. Fu K, Weisenburger DD, Greiner TC, et al. Cyclin D1-negative mantle cell lymphoma: a clinicopathologic study based on gene expression profiling. *Blood*. 2005;106(13):4315–21.
7. Salaverria I, Royo C, Carvajal-Cuenca A, et al. CCND2 rearrangements are the most frequent genetic events in cyclin D1(-) mantle cell lymphoma. *Blood*. 2013;121(8):1394–1402. Epub 2012/12/21.
8. Mozos A, Royo C, Hartmann E, et al. SOX11 expression is highly specific for mantle cell lymphoma and identifies the cyclin D1-negative subtype. *Haematologica*. 2009;94(11):1555–62. Epub 2009/11/03.
9. Fernandez V, Salamero O, Espinet B, et al. Genomic and gene expression profiling defines indolent forms of mantle cell lymphoma. *Cancer Res*. 2010;70(4):1408–18. Epub 2010/02/04.
10. Scott DW, Abrisqueta P, Wright GW, et al. New molecular assay for the proliferation signature in mantle cell lymphoma applicable to formalin-fixed paraffin-embedded biopsies. *J Clin Oncol*. 2017;35(15):1668–1677. Epub 2017/03/14.
11. Bouska A, Zhang W, Gong Q, et al. Combined copy number and mutation analysis identifies oncogenic pathways associated with transformation of follicular lymphoma. *Leukemia*. 2017;31(1):83–91.
12. Loeffler M, Kreuz M, Haake A, et al. Genomic and epigenomic co-evolution in follicular lymphomas. *Leukemia*. 2015;29(2):456–463. Epub 2014/07/17.
13. Liu Q, Salaverria I, Pittaluga S, et al. Follicular lymphomas in children and young adults: a comparison of the pediatric variant with usual follicular lymphoma. *Am J Surg Pathol*. 2013;37(3):333–343. Epub 2012/10/31.
14. Louissaint A, Jr. Ackerman AM, Dias-Santagata D, et al. Pediatric-type nodal follicular lymphoma: an indolent clonal proliferation in children and adults with high proliferation index and no BCL2 rearrangement. *Blood*. 2012;120(12):2395–2404. Epub 2012/08/03.
15. Schmidt J, Gong S, Marafioti T, et al. Genome-wide analysis of pediatric-type follicular lymphoma reveals low genetic complexity and recurrent alterations of TNFRSF14 gene. *Blood*. 2016;128(8):1101–1111.
16. Jegalian AG, Eberle FC, Pack SD, et al. Follicular lymphoma in situ: clinical implications and comparisons with partial involvement by follicular lymphoma. *Blood*. 2011;118(11):2976–2984. Epub 2011/07/20.
17. Mamessier E, Song JY, Eberle FC, et al. Early lesions of follicular lymphoma: a genetic perspective. *Haematologica*. 99:481–8, 2014. Epub 2013/10/29.
18. Swerdlow SH. Pediatric follicular lymphomas, marginal zone lymphomas, and marginal zone hyperplasia. *Am J Clin Pathol*. 2004;122(suppl): S98–S109.
19. Rinaldi A, Mian M, Chigrinova E, et al. Genome-wide DNA profiling of marginal zone lymphomas identifies subtype-specific lesions with an impact on the clinical outcome. *Blood*. 2011;117(5):1595–1604.
20. Rosenwald A, Wright G, Chan WC, et al. The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma. *N Engl J Med*. 2002;346(25):1937–1947. Epub 2002/06/21.
21. Coiffier B, Thieblemont C, Van Den Neste E, et al. Long-term outcome of patients in the LNH-98.5 trial, the first randomized study comparing rituximab-CHOP to standard CHOP chemotherapy in DLBCL patients: a study by the Groupe d'Etudes des Lymphomes de l'Adulte. *Blood*. 2010;116(12):2040–2045.
22. Chapuy B, Roemer MGM, Stewart C, et al. Targetable genetic features of primary testicular and primary central nervous system lymphomas. *Blood*. 2016;127(7):869–881.
23. Salaverria I, Philipp C, Oschlies I, et al. Translocations activating IRF4 identify a subtype of germinal center-derived B-cell lymphoma affecting predominantly children and young adults. *Blood*. 2011;118(1):139–147.
24. Ramis-Zaldivar JE, Gonzalez-Farre B, Balague O, et al. Distinct molecular profile of IRF4-rearranged large B-cell lymphoma. *Blood*. 2020;135(4):274–286.
25. Rosenwald A, Wright G, Leroy K, et al. Molecular diagnosis of primary mediastinal B cell lymphoma identifies a clinically favorable subgroup of diffuse large B cell lymphoma related to Hodgkin lymphoma. *J Exp Med*. 2003;198(6):851–862. Epub 2003/09/17.
26. Traverse-Glehen A, Pittaluga S, Gaulard P, et al. Mediastinal gray zone lymphoma: the missing link between classic Hodgkin's lymphoma and mediastinal large B-cell lymphoma. *Am J Surg Pathol*. 2005;29(11):1411–1421. Epub 2005/10/15.
27. Eberle FC, Rodriguez-Canales J, Wei L, et al. Methylation profiling of mediastinal gray zone lymphoma reveals a distinctive signature with elements shared by classical Hodgkin's lymphoma and primary mediastinal large B-cell lymphoma. *Haematologica*. 2011;96(4):558–566. Epub 2011/04/02.
28. Pittaluga S, Nicolae A, Wright GW, et al. Gene expression profiling of mediastinal gray zone lymphoma and its relationship to primary mediastinal B-cell lymphoma and classical Hodgkin lymphoma. *Blood Cancer Discov*. 2020;1(2):155–161.
29. Dunleavy K, Pittaluga S, Maeda LS, et al. Dose-adjusted EPOCH-rituximab therapy in primary mediastinal B-cell lymphoma. *N Engl J Med*. 2013;368(15):1408–1416. Epub 2013/04/12.
30. Dojcinov SD, Venkataraman G, Pittaluga S, et al. Age-related EBV-associated lymphoproliferative disorders in the Western population: a spectrum of reactive lymphoid hyperplasia and lymphoma. *Blood*. 2011;117(18):4726–4735. Epub 2011/03/10.
31. Nicolae A, Pittaluga S, Abdullah S, et al. EBV-positive large B-cell lymphomas in young patients: a nodal lymphoma with evidence for a tolerogenic immune environment. *Blood*. 2015;126(7):863–872.
32. Dojcinov SD, Venkataraman G, Raffeld M, et al. EBV positive mucocutaneous ulcer—a study of 26 cases associated with various sources of immunosuppression. *Am J Surg Pathol*. 2010;34(3):405–417. Epub 2010/02/16.
33. Song JY, Pittaluga S, Dunleavy K, et al. Lymphomatoid granulomatosis—a single institute experience: pathologic findings and clinical correlations. *Am J Surg Pathol*. 2015;39(2):141–156.
34. Taddesse-Heath L, Meloni-Ehrig A, Scheerle J, et al. Plasmablastic lymphoma with MYC translocation: evidence for a common pathway in the generation of plasmablastic features. *Mod Pathol*. 2010;23(7):991–999. Epub 2010/03/30.
35. Dave SS, Fu K, Wright GW, et al. Molecular diagnosis of Burkitt's lymphoma. *N Engl J Med*. 2006;354(23):2431–2442.

36. Salaverria I, Martin-Guerrero I, Wagener R, et al. A recurrent 11q aberration pattern characterizes a subset of MYC-negative high-grade B-cell lymphomas resembling Burkitt lymphoma. *Blood.* 2014;123(8):1187–1198.
37. Au WY, Weisenburger DD, Intragumtornchai T, et al. Clinical differences between nasal and extranasal natural killer/T-cell lymphoma: a study of 136 cases from the International Peripheral T-Cell Lymphoma Project. *Blood.* 2009;113(17):3931–3937. Epub 2008/11/26.
38. Quintanilla-Martinez L, Ridaura C, Nagl F, et al. Hydroa vacciniforme-like lymphoma: a chronic EBV$^+$ lymphoproliferative disorder with risk to develop a systemic lymphoma. *Blood.* 2013;122(18):3101–3110. Epub 2013/08/29.
39. Heavican TB, Bouska A, Yu J, et al. Genetic drivers of oncogenic pathways in molecular subgroups of peripheral T-cell lymphoma. *Blood.* 2019;133(15):1664–1676. Epub 2019/02/21.
40. Watatani Y, Sato Y, Miyoshi H, et al. Molecular heterogeneity in peripheral T-cell lymphoma, not otherwise specified revealed by comprehensive genetic profiling. *Leukemia.* 2019;33(12):2867–2883.
41. Thompson PA, Lade S, Webster H, et al. Effusion-associated anaplastic large cell lymphoma of the breast: time for it to be defined as a distinct clinico-pathological entity. *Haematologica.* 2010;95(11):1977–1979.
42. Parrilla Castellar ER, Jaffe ES, Said JW, et al. ALK-negative anaplastic large cell lymphoma is a genetically heterogeneous disease with widely disparate clinical outcomes. *Blood.* 2014;124(9):1473–1480.
43. Gayden T, Sepulveda FE, Khuong-Quang DA, et al. Germline HAVCR2 mutations altering TIM-3 characterize subcutaneous panniculitis-like T cell lymphomas with hemophagocytic lymphohistiocytic syndrome. *Nat Genet.* 2018;50(12):1650–1657. Epub 2018/10/31.
44. Roberti A, Dobay MP, Bisig B, et al. Type II enteropathy-associated T-cell lymphoma features a unique genomic profile with highly recurrent SETD2 alterations. *Nat Commun.* 2016;7:12602. Epub 2016/09/08.
45. Nicolae A, Xi L, Pittaluga S, et al. Frequent STAT5B mutations in gamma-madelta hepatosplenic T-cell lymphomas. *Leukemia.* 2014;28(11):2244–2248. Epub 2014/06/21.
46. Nakagawa M, Schmitz R, Xiao W, et al. Gain-of-function CCR4 mutations in adult T cell leukemia/lymphoma. *J Exp Med.* 2014;211(13):2497–2505. Epub 2014/12/10.
47. Nam-Cha SH, Montes-Moreno S, Salcedo MT, et al. Lymphocyte-rich classical Hodgkin's lymphoma: distinctive tumor and microenvironment markers. *Mod Pathol.* 2009;22(8):1006–1015. Epub 2009/05/26.
48. Fan Z, Natkunam Y, Bair E, et al. Characterization of variant patterns of nodular lymphocyte predominant Hodgkin lymphoma with immunohistologic and clinical correlation. *Am J Surg Pathol.* 2003;27(10):1346–1356.
49. Xing KH, Connors JM, Lai A, et al. Advanced-stage nodular lymphocyte predominant Hodgkin lymphoma compared with classical Hodgkin lymphoma: a matched pair outcome analysis. *Blood.* 2014;123(23):3567–3573. Epub 2014/04/10.

79

Monoclonal Gammopathies

Angela Dispenzieri

The monoclonal gammopathies include a spectrum of disorders involving an over-proliferation of plasma cells or B cells that have the ability to secrete monoclonal proteins into the serum or the urine. Some of these disorders are overt malignancies, characterized by significant overgrowth of malignant cells in the bone marrow, and in the instance of Waldenström macroglobulinemia (WM) in lymph nodes; others are premalignant conditions that are asymptomatic; and yet others create pathology through deposition and or humoral-mediated mechanisms.

DIAGNOSING A MONOCLONAL GAMMOPATHY

As will be discussed throughout this chapter, there are different pathways by which each of the clinical diagnoses are made, but basic common elements to the diagnosis of monoclonal gammopathies rely on the detection of a monoclonal protein and bone marrow evaluation. Monoclonal proteins are defined as proteins that have restricted migration on an electrophoretic gel (Fig. 79.1, A). The clonality has historically been established by immunoelectrophoresis (see Fig. 79.1, B) or immunosubstraction and most recently by mass spectrometry (see Fig. 79.1, C). An expansive repertoire of plasma cells will make a diverse population of immunoglobulins, but clonal expansion of secretory plasma cells in the bone marrow space will result in overproduction of a single (monoclonal) immunoglobulin (Ig), which in turn may be seen on protein electrophoresis of the blood (serum protein electrophoresis [SPEP]) and/or urine as a restricted band, which has been called an M component, an M protein, or an M spike. "M" in this context refers to "monoclonal." The isotype of this Ig is then clarified by immunologic techniques, demonstrating whether the heavy chain is present or not, and if present whether it is IgG, IgA, IgM, IgD, or IgE. The same methodology provides information about whether the light chain is kappa or lambda. The level (quantity) of the M protein typically has been determined by the area under the curve on the electrophoretic pattern.

Nephelometry, which measures light scatter, contributes to the assessment of the quantity of serum immunoglobulins. However, there is imperfect correlation between the M protein calculated by SPEP and nephelometry, due to technical issues. First, quantitative immunoglobulins by nephelometry measure both monoclonal and polyclonal immunoglobulins, thereby overestimating the size of small M proteins in the context of polyclonal activation of plasma cells (e.g., infection, autoimmune disease, liver disease). Second, the SPEP can underestimate the size of M proteins when they migrate in the middle of the beta-region. Third, the SPEP can underestimate the size of very large M proteins due to dye saturation.

The nephelometric test measuring serum immunoglobulin free light chain (FLC) has proven to be most useful, particularly in the context of immunoglobulin light-chain (AL) amyloidosis and also in myeloma when much of the secreted immunoglobulin is light chain without an accompanying heavy chain. Measuring ***total*** immunoglobulin light chains by nephelometry is typically not a constructive exercise because even in diseases in which large amounts of immunoglobulin free (or unbound) light chains are made, these quantities are orders of magnitude lower than that of the overall total circulating immunoglobulin levels. Therefore the measurement of total light chains merely provides the amount of intact immunoglobulins with that particular light chain, and the unbound or FLCs are a rounding error. In contrast, the reagents specifically designed to detect FLCs do not detect light chains bound to heavy chains. These tests revolutionized physicians' ability to follow patients with AL amyloidosis and have been useful for the diagnosis, risk assessment, and monitoring of many other monoclonal gammopathies.

The newest technology for identifying and monitoring monoclonal proteins is mass spectrometry, a technique penned as miRAMM (monoclonal immunoglobulin rapid accurate mass measurement).[1] This methodology can be carried out on a liquid chromatography–coupled electrospray ionization time of flight (TOF)-based mass spectrometer or on a simpler, less-expensive tabletop matrix-assisted laser desorption/ionization (MALDI)-TOF mass spectrometer, the latter of which is penned MASS-FIX. miRAMM identifies the M protein from the accurate molecular mass of the light chain. The advantages to this methodology are multiple. First, it provides greater sensitivity and specificity than does immunoelectrophoresis. Second, it can distinguish therapeutic monoclonal antibodies from a patient's own monoclonal protein by differences in mass. Third, the methodology is quicker and requires less technical time to generate the result.

The other test that is essential for evaluating most of the monoclonal gammopathies is a bone marrow aspiration for morphologic assessment, flow cytometry, and genetic analysis (most often fluorescence in situ hybridization [FISH]). In addition to the bone marrow aspiration, a biopsy is typically done for a better morphologic appraisal.

KEY CONCEPTS

1. Low-abundance monoclonal proteins are found in monoclonal gammopathy of unknown significance (MGUS) and merely warrant patient observation
2. A monoclonal gammopathy with clinical findings often signals a malignant condition including multiple myeloma, Waldenström macroglobulinemia, or AL amyloidosis that requires specific therapy
3. Monoclonal proteins are also observed in patients with smoldering multiple myeloma or smoldering Waldenström macroglobulinemia
4. Low-level monoclonal proteins also can be associated with a number of serious conditions, known as monoclonal gammopathy of clinical significance (MGCS), that often warrant plasma cell–directed therapy

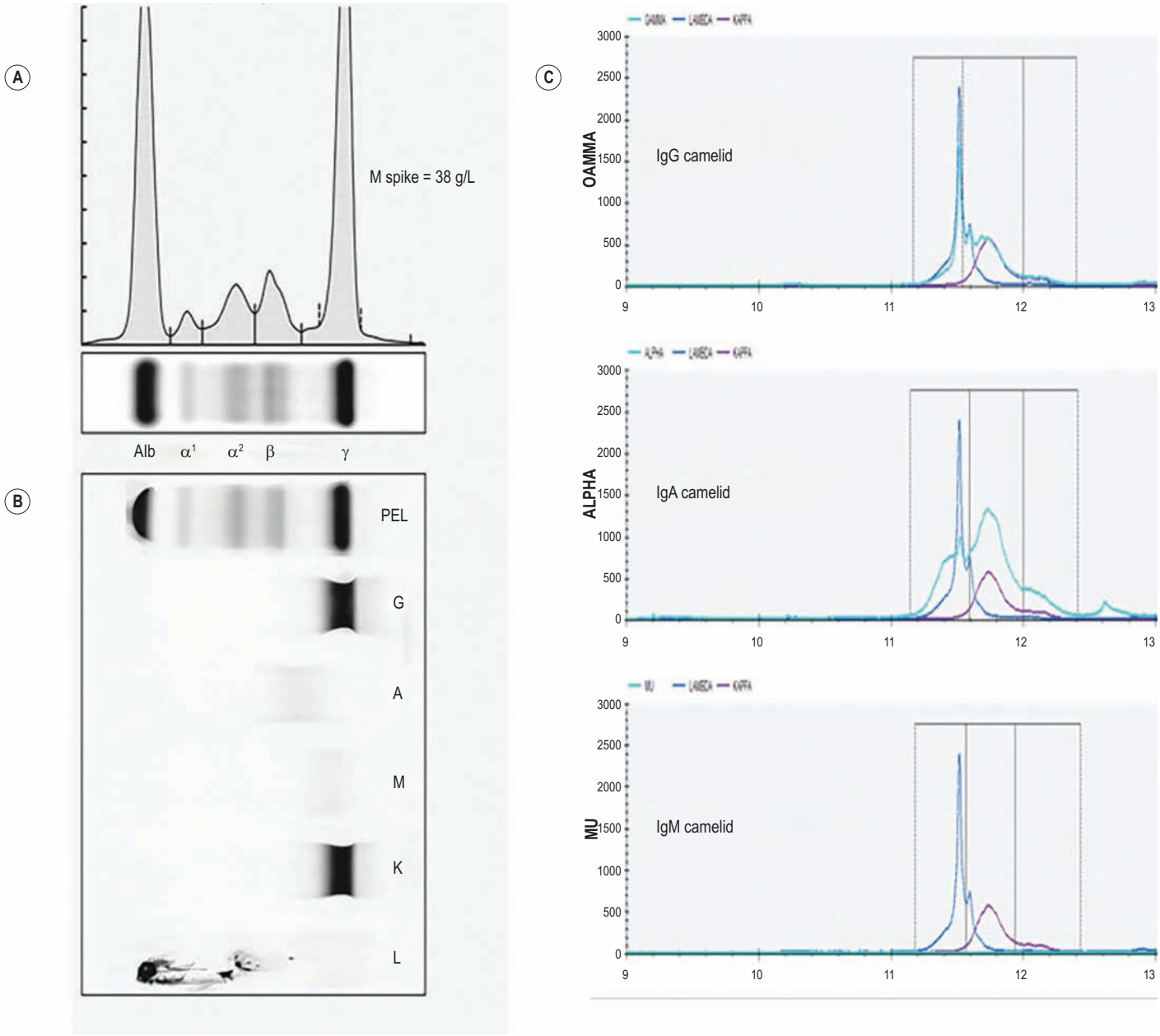


FIG. 79.1 Selected Techniques for Identifying Monoclonal Proteins. (A) Protein electrophoresis in a patient with multiple myeloma. (B) Immunoelectrophoresis in an immunoglobulin G (IgG) lambda patient. (C) MASS-FIX in a different immunoglobulin G (IgG) lambda patient. (C) Blue, purple, and turquoise represent lambda, kappa, and specified heavy chain camelids, respectively.

CLINICAL RELEVANCE

Accurate diagnosis is essential for each monoclonal gammopathy to determine whether observation or directed therapy is the appropriate management strategy.

MONOCLONAL GAMMOPATHY OF UNDETERMINED SIGNIFICANCE

Monoclonal gammopathy of undetermined significance (MGUS) is a term coined by Kyle in 1978 to account for the observation that patients found with small monoclonal proteins on electrophoresis without multiple myeloma (MM) or WM had a condition that could progress to these malignancies over time. These small proteins that persisted as such for years without end-organ damage attributable to the plasma cell disorder progressed at a rate of approximately 1% per year to myeloma and related disorders.

Epidemiology

MGUS is the most common plasma cell proliferative disorder, with a prevalence of approximately 4% among adults 50 years or older.[2] Rates are higher in men than women, increase with increasing age, are more than twofold more common in African Americans than Whites, and twofold to threefold more common in first-degree relatives of individuals with MGUS or MM. The annual incidence of MGUS in males is estimated to be 120/100,000 at age 50 and rises to 530/100,000 at age 90 years. The rates for women are 60/100,000

TABLE 79.1 Definition of Plasma Cell Dyscrasia

	M Protein	BM Clonal Cells	Other
MGUS (all required)			
Non-IgM (IgA, IgG, IgD, or IgE)	Serum <3 g/dL	PC <10%[a]	No CRAB attributable
IgM	Serum <3 g/dL	LP <10%	No CRAB attributable
Light chain	Abn FLC ratio due to elevation of one of the FLCs. No heavy chain Urine M protein <0.5/24 h	PC <10%	No anemia, hyperviscosity, adenopathy, organomegaly attributable
SMM (either M protein or BM criteria met)	Serum (IgA, IgG, IgD, or IgE) ≥3 g/dL or urine M protein ≥ 0.5/24 h	PC 10%–60%	No CRAB attributable
SWM (all criteria required)	Serum (IgM) ≥3 g/dL or urine M protein ≥ 0.5/24 h	LP ≥10%[b]	No anemia, hyperviscosity, adenopathy, organomegaly attributable
Multiple myeloma[c] (any criteria met)	Serum (IgA, IgG, IgD, or IgE) ≥3 g/dL Urine M protein ≥ 0.5/24 h	PC ≥10%[d]	CRAB attributable, >1 focal lesion on MRI, FLC ratio ≥100 with involved FLC also ≥100, or BMPC ≥60%
Waldenström macroglobulinemia (all required)	Serum IgM protein	LP ≥10%[b]	Anemia, hyperviscosity, adenopathy, or organomegaly attributable
Solitary plasmacytoma	Any	No clonal BM PC	Biopsy proven solitary lesion of bone or soft tissue with clonal PC
Solitary plasmacytoma with minimal marrow involvement	Any	PC <10%	Normal skeletal survey and MRI (or CT) of spine and pelvis (except for the primary solitary lesion) No CRAB attributable

[a]Bone marrow can be deferred in patients with low-risk MGUS.
[b]Phenotype (e.g., surface IgM$^+$, CD5$^{+/-}$, CD10$^-$, CD19$^+$, CD20$^+$, CD23$^-$) that satisfactorily excludes other lymphoproliferative disorders including chronic lymphocytic leukemia and mantle cell lymphoma.
[c]Plasma cell leukemia diagnosis is made if more than 20% plasma cells in periphery and a circulating absolute plasma cell count of 2×10^9/L.
[d]Few plasma cells may be present if biopsy proven bony or extramedullary plasmacytoma.
BM, Bone marrow; *CRAB*, hypercalcemia, renal dysfunction, anemia, bone lesions; *CT*, computed tomography; *FLC*, free light chain; *Ig*, immunoglobulin; *LP*, lymphoplasmacytic infiltrate; *MGUS*, monoclonal gammopathy of undetermined significance; *MRI*, magnetic resonance imaging; *PC*, plasma cells; *SMM*, smoldering multiple myeloma; *SWM*, smoldering Waldenström macroglobulinemia.

at age 50, and 370/100,000 at age 90. The fact that the increased prevalence of MGUS with increasing age is not just related to accumulation of new cases but due to an actual increase in incidence suggests that an age-related cumulative damage model is at play in the pathogenesis of MGUS.

Clinical Presentation and Diagnosis

By definition (Table 79.1), individuals with MGUS are asymptomatic in terms of their plasma cell disorder, and it typically represents an incidental finding. A monoclonal protein may be found when an individual is found to have a high total serum protein or high erythrocyte sedimentation rate on routine blood work or when the patient has some indeterminant symptomatology, and a SPEP is performed as part of the evaluation. In this latter scenario, a patient may have symptoms and signs, but when the evaluation is complete, none of these are attributed to the monoclonal gammopathy. To exclude other diagnoses, an extensive review of symptoms looking for symptoms referable to monoclonal gammopathy of clinical significance (MGCS), MM, AL amyloidosis, etc., is required. In terms of testing, in addition to the SPEP and immunofixation which had already been done to diagnose the monoclonal gammopathy, the individual should have a complete blood count, a serum calcium, serum creatinine, alkaline phosphatase, spot urine looking for excess urinary protein, and serum FLC assay. The FLC assay is essential for MGUS prognostication. If there are cardiac symptoms, an N-Terminal proB-type natriuretic peptide (NT-proBNP) blood test should also be done as an initial screen for AL amyloidosis cardiomyopathy. A bone marrow is not necessary if the patient has low-risk MGUS and is potentially elective if the patient has low-intermediate–risk MGUS (see later).

Differential Diagnosis

The differential diagnosis for MGUS is as listed earlier. The monoclonal protein could be part of a malignant disease, a deposition disease, or a humoral-mediated disease (see later).

Management and Prognosis

Once the diagnosis is of MGUS is established, follow-up in 3 to 6 months is important to exclude the possibility that the patient was on his or her initial trajectory toward a malignant disorder. Once again, an extensive review of systems and the same blood work done at diagnosis minus the immunoelectrophoresis and the bone marrow should be performed. If the results are stable, follow-up can be based on the MGUS risk assessment.[3] Patients are assigned a point for each of the following factors: immunoglobulin M (IgM) or IgA isotype; abnormal FLC ratio; and M protein greater than 1.5 g/dL. A score of zero is deemed low-risk MGUS, and these patients have a 0.5%/year risk of progression to MM or related disorders. A score of 1 is low-intermediate risk, and these patients have a 1% per year risk for progression; 2 is intermediate risk with a 1.5% per year risk for progression; and a score of 3 is high-risk MGUS with a risk of 2% per year for progression. Overall risk is considerably lower if one accounts for competing risk of death in this generally elderly population. Other risk factors for progression include suppression of uninvolved immunoglobulins, higher bone marrow plasmacytosis, and increased percent aberrant phenotype plasma cells in the marrow.

After stability has been demonstrated, it is quite reasonable to do no additional follow-up in the low-risk MGUS population and annual follow-up for the remaining groups and/or additional special follow-up if there are symptoms worrisome for progression.

SMOLDERING MULTIPLE MYELOMA

As shown in Table 79.1, smoldering multiple myeloma (SMM) is another asymptomatic condition. Patients who clinically look like MGUS, but who have either an M protein or plasmacytosis beyond what is allowable for MGUS, are designated SMM. These patients have a significantly higher rate of progression to MM and related disorders. SMM is not a true biologic entity but rather a gray zone between MGUS and MM. Genetic studies of bone marrow plasma cells from SMM patients and MM patients are indistinguishable.

Epidemiology

SMM accounts for approximately 15% of all cases of newly diagnosed myeloma. There are no epidemiologic studies defining SMM prevalence.

Clinical Presentation and Diagnosis

Like MGUS, SMM is an incidental finding. Individuals with SMM by definition should not have anemia, renal dysfunction, hypercalcemia, bone disease, or plasmacytosis more than 60%. Evaluation for these patients is similar to that of MGUS patients, but bone radiographs—preferably a computed tomography (CT) skeletal survey, or magnetic resonance imaging (MRI)—should also be performed to exclude bone involvement.

Differential Diagnosis

The differential diagnosis is active MM, AL amyloidosis, and the other MGCS.

Management and Prognosis

Progression of SMM, based on a study of 276 patients with SMM, was 10% per year for the first 5 years, 5% per year for the next 3 years, and then 1% to 2% per year thereafter.[4] This pattern of progression in which there is a plateau after 10 years is consistent with the heterogeneous nature of SMM; in the first 10 years, the subset of patients with early MM declare themselves with symptomatic disease, while after 10 years, the remaining cohort of patients is identical to MGUS in biology and clinical behavior. A subset of patients can remain free of progression for several years.[5]

Risk factors for progression include level of the M protein, extent of abnormal FLC ratio, uninvolved immunoglobulin suppression, circulating plasma cells evaluated by flow cytometry, extent of aberrant bone marrow plasmacytosis and overall bone marrow plasmacytosis, proliferative rate of bone marrow plasma cells, and cytogenetic changes. Patients with t(4;14) and deletion 17p have the highest risk of progression, and patients with no FISH abnormalities are at lowest risk.

Patients with SMM are generally observed with serial follow-up blood tests and radiographs, the former of which are done quarterly and the latter typically every year.

The exception to the aforementioned observation strategy should be followed for those patients who are classified as having high-risk SMM defined by having two or three of the following risk factors: bone marrow plasma cells greater than 20%; M protein greater than 2 g/dL; and FLC ratio greater than 20.[6] Patients with high-risk SMM have a median time to MM progression of 29 months as compared with patients with none or one risk factors, who have a progression to MM of 110 and 68 months, respectively. Two randomized trials have demonstrated improved progression-free survival (PFS) for the high-risk population when treated with lenalidomide or lenalidomide and dexamethasone therapy. In addition, therapy trials showed improvement in overall survival.[7,8] There is controversy as to whether these patients be recategorized as active myeloma and treated as such, with ongoing trials underway to address this question.

MULTIPLE MYELOMA

MM is a neoplastic plasma cell dyscrasia (PCD) characterized by CRAB: (1) hypercalcemia; (2) renal insufficiency; (3) anemia; and (4) bone lesions and/or bone pain (Fig. 79.2). According to the World Health Organization (WHO) classification system, there is only one category for MM.[9]

Epidemiology

With the exception of MGUS, MM is the most common PCD, with an incidence of approximately 1 per 10,000 per year in the United States.[10] The 2018 annual estimate in the United States for new cases of MM is 30,770 and for deaths is 12,770. The age-adjusted rates in the United States show an incidence among African Americans that is approximately twofold higher than Whites and the incidence in men is approximately 1.6-fold that of women. The median age at diagnosis is 69 years. Since the turn of the century, 5-year survival rates in ethnic groups noted have equalized. Epidemiologic risk factors for MM include radiation, chronic antigenic stimulation, obesity, and low socioeconomic status.

Clinical Presentation and Diagnosis

The diagnosis is made according to criteria shown in Table 79.1, and one of the most important distinctions is whether the disorder is SMM or active MM requiring treatment. Anemia, bone disease, hypercalcemia, and renal dysfunction are the hallmarks of MM resulting in a typical clinical presentation of fatigue and/or bone pain. Due to suppression of uninvolved immunoglobulin production, patients with MM are at higher risk for infection. In addition, hyperviscosity occurs rarely and may be a cause for increased bleeding.

Essential testing at the time of diagnosis includes: SPEP and urine protein electrophoresis with isotype monoclonal protein using one of the methods discussed previously; complete blood count; calcium; creatinine; serum albumin (found on SPEP); lactate dehydrogenase. In addition, a bone marrow biopsy with FISH, looking for IgH translocations (t(11;14), t(4;14), t(14;16), t(14;20)), evidence of hyperdiploidy (focusing on odd chromosomes), copy number of 1q+, and deletions

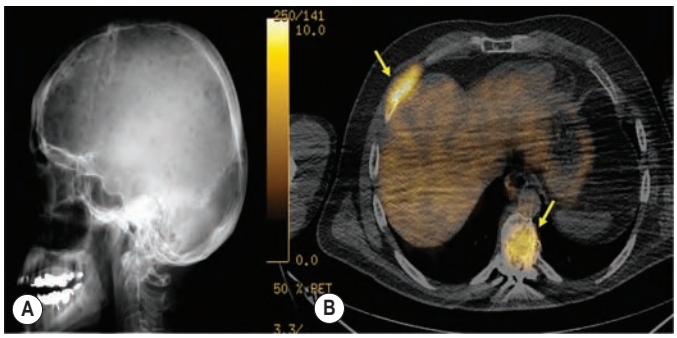


FIG. 79.2 Lytic Lesions in Multiple Myeloma. (A) Lytic lesions in skull. (B) ^{18}Fluorodeoxyglucose positron emission tomography/computed tomography demonstrating two avid lytic lesions in right rib and vertebra. *Yellow arrows* refer to lesions.

(chromosome 13, 17p, 1p).[11] Additional testing such as gene expression profiling and flow cytometry looking for plasma cell phenotype and proliferative rate may also be done.

Two-thirds of myeloma patients have a monoclonal kappa clone. Approximately 1% of patients with MM have nonsecretory disease and have no detectable circulating or urinary monoclonal proteins. The Ig isotype has been reported to be IgG, IgA, IgD, and light chain only in 52%, 20%, 2%, and 16% of cases, respectively. No individual bone marrow finding is pathognomonic for a malignant plasma cell process other than plasmacytosis of greater than 60%; the bone marrow diagnosis of MM relies on percentage of clonal bone marrow plasma cells, with 10% accepted as a cutoff to move a patient from MGUS to SMM or MM. However, a clinical diagnosis of MM can be made with fewer bone marrow plasma cells if CRAB or a myeloma-defining event is present.

Differential Diagnosis

The differential diagnosis is SMM, MGUS, AL amyloidosis, MGCS, solitary plasmacytoma, POEMS (polyradiculoneuropathy, organomegaly, endocrinopathy, monoclonal plasma cell disorder, and skin changes) syndrome, plasma cell leukemia (PCL), and diseases that can cause reactive plasmacytosis with resultant polyclonal hypergammaglobulinemia (e.g., chronic infection and autoimmune diseases). Because a very small minority of MM cases (≤1%) can have an IgM monoclonal protein, it is unusual to confuse MM with WM.

Management and Prognosis

The management involves aggressive supportive care and treatments directed at killing the myeloma cells. The complications of hypercalcemia and/or renal failure at presentation require aggressive treatment using hydration, occasionally with plasmapheresis, calcium-lowering agents, and urgent chemotherapy. There is an expanding repertoire of therapies for patients with myeloma. The two most important considerations are fitness (which often correlates with age) and genetic risk. Many will consider a pathway that includes autologous stem cell transplant (ASCT) in fit patients younger than age 75; alternate, effective strategies exist for less fit patients. In general, allogeneic stem cell transplant is not considered a standard therapy.

The mainstays of chemotherapy include the following different classes of drugs. Proteasome inhibitors include bortezomib, carfilzomib, and ixazomib. Immune modulators include thalidomide, lenalidomide, and pomalidomide. Alkylators include melphalan and cyclophosphamide. Corticosteroids include dexamethasone, prednisone, and methylprednisolone. Monoclonal antibodies include daratumumab, elotuzumab, and isatuximab. The most commonly used induction (first-line therapy) involves a combination of an immune modulator, a proteasome inhibitor, and a corticosteroid (e.g., bortezomib, lenalidomide, and dexamethasone). This highly effective regimen can be continued for approximately 4 months in those patients headed for ASCT as part of their first-line treatment and for approximately 6 months in those patients who are not ASCT candidates. In this latter group, maintenance with the single agent lenalidomide (or sometimes with dexamethasone) is continued for a minimum of 2 years but often indefinitely. For those patients intended for early transplant, their stem cells are collected before receiving myeloablative melphalan, followed by infusion of their stem cells. Maintenance with lenalidomide is then started around day 100 and continued for a minimum of 2 years to indefinitely.[12] With these strategies, the vast majority—upwards of 75%—of patients achieve a very good partial response or better. Durations of response vary depending on risk, but with these strategies, median PFS is 3 years in the non-ASCT group and more than 4 years in the ASCT group. Alternate induction regimens gaining popularity include combinations of an immune modulator drug, daratumumab, and a corticosteroid.

Relapsing patients have a number of therapeutic options including combinations of the drugs already mentioned, as well as innovative clinical trial strategies including check-point inhibitors, antibody-drug conjugates, chimeric antigen receptor T-cell (CAR-T) therapies and bi-specific T-cell engagers (BITE). Most commonly targeted antigens include B-cell maturation antigen (BCMA) and CD38.[13] The drug venetoclax, which blocks the antiapoptotic B-cell lymphoma-2 (BCL-2) protein, has shown efficacy in myeloma patients with the t(11;14) as a single agent but also in combination with bortezomib. This drug is US Food and Drug Administration (FDA) approved for other hematologic malignancies. The check-point inhibitor strategy had initially looked promising in conjunction with immune modulator drugs at the level of response, but randomized trials showed higher mortality using check-point inhibitors, stalling these initiatives. GSK2857916 (belantamab mafodotin), an antibody-drug conjugate against BCMA studied in multiply refractory myeloma, has yielded 60% response rates with a PFS of nearly 8 months. The most common adverse events were thrombocytopenia and corneal events. There are multiple CAR-T therapies in clinical trials. The products with the most experience in the United States include the bb2121 CAR-T (idecaptagene cicleucel), which targets BCMA.

There are many prognostic factors recognized in myeloma, but the most universally used are based on the International Staging System (ISS) and the Revised ISS (R-ISS).[11] The ISS is a simple system that incorporates the baseline serum albumin and beta-2 microglobulin. In 2015 FISH (t(4;14), t(14;16), or deletion 17p) and serum lactate dehydrogenase (LDH) was added to the ISS to further stage patients.

PLASMA CELL LEUKEMIA

PCL is a rare form of PCD. By definition, there are more than 20% plasma cells in the peripheral blood with an absolute plasma cell count of more than 2×10^9/L. Some authors accept the diagnosis with only one of these criteria. Less rigorous cutoffs have been considered to establish the diagnosis of PCL. The presentation may be primary, de novo, or secondary, evolving from an existing case of myeloma as part of the terminal phase of the disease. The morphology and immunophenotype of the plasma cells in PCL are indistinguishable from that of myeloma, but CD56 is more frequently found to be expressed on myeloma cells and CD20 on PCL cells.

Epidemiology

Between 2% and 4% of malignant PCD cases are PCL, and 60% to 70% of these cases are primary. The incidence of PCL in a Danish registry was 1.2 cases per million persons.[14]

Clinical Presentation and Diagnosis

As compared with patients with myeloma, primary PCL patients tend to be younger and sicker and have higher rates of light chain–only disease, high LDH, 17p deletion, and extramedullary disease.

The diagnostic work-up is similar to that of myeloma, but a positron emission tomography (PET)/CT may be warranted to enable better assessment for extramedullary disease.

Differential Diagnosis

Myeloma or severe infection is in the differential diagnosis. Because high levels of circulating polyclonal plasma cells can be seen in severe infections and serum sickness, establishing clonality of the plasma cells is important.

Management and Prognosis

The prognosis of primary PCL is worse than that of MM. In contrast, the survival of patients with secondary PCL is universally dismal. The most optimistic median survival rates are approximately 3 years, but those are in patients who lived long enough to make it to stem cell transplantation. Most other studies/reports estimate median survival of 1.5 to 2 years for patients with primary PCL. There is no clear guidance on what the most effective therapies are, but immune modulator drug (IMiD)-proteasome inhibitor (PI) combinations have yielded good response rates, as have complex multidrug regimens including conventional chemotherapy. Consolidation with ASCT should be done if feasible, and all patients should have maintenance therapy. So far, there are no data to support allogeneic stem cell transplant over ASCT in PCL. There is an anecdotal report of a patient with relapsed primary t(11;14)-PCL who was treated with venetoclax monotherapy and who enjoyed remission for more than 9 months.[15]

PLASMACYTOMA

Solitary plasmacytomas come in two varieties: (1) bone (SPB) and (2) extramedullary spaces (SEP).[16] In the old literature there was an overcall of SPB because immunohistochemistry of the bone marrow was not always done and imaging was primitive by current standards. Many of these older cases diagnosed as SPB were early myeloma with either low levels of MM in the marrow or multiple other bone lesions that were below the level of detection using imaging of the period. This yielded a falsely high rate of progression for true plasmacytoma of bone. Definitions were also not standardized in that some reports allowed for a patient to be diagnosed with SPB even if they had documented clonal plasma cells found on a random bone marrow biopsy. The International Myeloma Working Group developed more restricted diagnostic criteria and allowed for a new category called plasmacytoma of bone with minimal marrow involvement (see Table 79.1). In contrast, cases of SEP have increased over the decades, with better imaging and biopsy techniques, but this diagnosis still remains a rare entity.

Epidemiology

Good epidemiologic studies are lacking, but SPB and SEP account for approximately 2% to 5% of malignant PCD treated at large referral centers.

Clinical Presentation and Diagnosis

Criteria for diagnoses are shown in Table 79.1. For SPB, there is a male predominance and the median age of patients is 55 years. The axial skeleton is more often affected, and pain is the usual presentation. SPB can sometimes be associated with POEMS syndrome (see later). The presenting symptoms of SPE align with the location of the mass, with approximately 80% involving the oronasopharynx and paranasal sinuses, while other sites of involvement include the gastrointestinal tract, lung, liver, lymph nodes, skin, and central nervous system. Careful staging should be done akin to the evaluation performed for patients with MM, but PET/CT or whole-body MRI is required to exclude additional disease. Careful evaluation of the random bone marrow biopsy for clonal disease is also required.

Differential Diagnosis

The SPB differential diagnosis includes MM and POEMS syndrome. For SEP, reactive plasmacytosis, plasma cell granuloma, and immunoblastic lymphoma should be excluded. Some cases of SEP may represent marginal zone B-cell lymphoma that has undergone plasmacytic differentiation.

Management and Prognosis

These are potentially curable diseases. For SPB the most commonly used therapy is radiation, 40 to 50 Gy. Surgery should be reserved for patients with mechanical instability. For SEP, treatment options include surgery or radiation, whereas adjuvant chemotherapy is not recommended.[16] Many risk factors for progression have been identified, including size of the lesion, a nonvertebral presentation, abnormal FLC ratio, nonsecretory disease, and depression of immunoglobulins at presentation. Local progression is rare (<10%), and the risk for progression to myeloma appears to be less common for patients with SEP than SPB, with 10% progressing to myeloma within 3 years.

IMMUNOGLOBULIN LIGHT-CHAIN AMYLOIDOSIS

AL amyloidosis (formerly called primary amyloidosis) is a complex plasma cell disorder that impacts patients through immunoglobulin deposition in vital organs in the form of amyloid fibrils. These 8- to 10-nm fibrils have a beta pleated sheet structure and are highly insoluble.[17] Some immunoglobulin light chains are more prone to form amyloid, including the light chains from IGVL-6 gene; light-chain glycosylation may also be a risk factor for amyloid formation. Currently it is unknown why AL amyloid is formed.

There are two types of AL amyloidosis: a localized form and a systemic form. The former refers to the condition with little to no circulating clonal immunoglobulin and the amyloid deposits occur at the site of the cells producing the clonal protein. In systemic AL, there is circulating clonal immunoglobulin—most often in the form of an FLC—and the amyloid deposition is distant from the immunoglobulin-producing cells that reside in the bone marrow.

Epidemiology

There are limited data on the epidemiology of AL amyloidosis. The rates of the disease rise with increasing age, with the reported median in the 70 s. There is a male predominance ranging from 54% to 70%, and there are no identified risk factors for AL amyloidosis other than a preexisting monoclonal gammopathy. The incidence of AL has been estimated to be approximately 10.8 to 15.2 cases per million.

Clinical Presentation and Diagnosis

Systemic AL amyloidosis is an insidious disease that often goes undiagnosed for a year or longer. Approximately 70% have cardiac and/or renal involvement, and 15% have liver and/or 15% nerve involvement. Patients will have symptoms referable to these organ systems, including dyspnea, fluid retention, light-headedness, dysesthesias, and fatigue. Other symptoms may include periorbital purpura and jaw claudication (due to

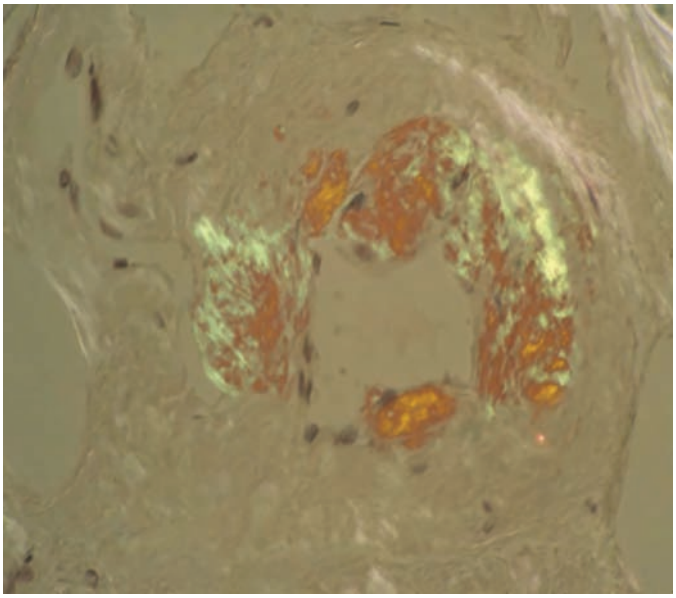


FIG. 79.3 Congo Red Stain of a Bone Marrow. Positive Congo red stain (1000×) under polarized light demonstrating apple green birefringence.

microvascular disease), macroglossia, vocal hoarseness, diarrhea, constipation, early satiety, hematochezia, weight loss, and carpal tunnel syndrome.

In contrast, localized AL amyloidosis occurs in one of the following systems: the respiratory tract, the genitourinary tract, the skin, the gastrointestinal tract, or the lymphatics. Symptoms are associated with the involved organ systems.

The diagnosis of AL amyloidosis is a tissue diagnosis. Biopsy of the affected organ in localized AL amyloidosis is required to make that diagnosis, but for systemic AL amyloidosis, surrogate tissues can provide a diagnosis in the right clinical situation. The fat aspirate is positive for AL amyloidosis in approximately 85% of cases demonstrating Congo red positivity and apple green birefringence on polarized light (Fig. 79.3). The application of mass spectrometry typing showing either kappa or lambda light chains is sufficient to make the diagnosis of systemic AL amyloidosis. In such a patient, the presence of an enlarged, stiff heart is presumed to be associated with AL amyloid deposition in the tissue without the need for an endomyocardial biopsy.

Differential Diagnosis

The differential diagnosis for AL amyloidosis is MGUS, SMM, MM, light-chain deposition disease, and other types of amyloidosis. Not all amyloid is AL amyloidosis, with more than 30 proteins having been found to cause disease in humans. Once amyloidosis is identified on a biopsy, the causative protein must be typed to distinguish the type of amyloid present. A tissue mass spectrometry approach is preferable, but in some instances other techniques can be used. The importance of typing cannot be overemphasized, because the therapeutic strategy is different for patients with AL amyloidosis compared with other acquired or inherited forms of amyloidosis.

Management and Prognosis

Patients are staged using simple blood tests: troponin, NT-proBNP, and immunoglobulin FLCs. The Mayo 2012 staging system assigns those patients with none of these markers elevated Stage I, with one elevated Stage II, with two elevated Stage III, and all three elevated Stage IV. Patients with Stage IV disease have very poor overall survival, whereas patients with Stage I disease have median survival of more than 6 years.

The mainstay of therapy for patients with AL amyloidosis is plasma cell–directed chemotherapy. The same drugs used to treat myeloma are used to treat AL amyloidosis, although often at slightly different doses and schedules. Patients with AL amyloidosis have more side effects due to their compromised organs. The flip side is that, because approximately half of patients with AL amyloidosis have such a low plasma cell burden, they often respond faster and more completely than myeloma patients and also have a longer time to progression. Daratumumab, a therapeutic monoclonal antibody directed against CD38, is especially active in patients with AL amyloidosis,[18] and anecdotal evidence shows great promise for venetoclax,[19] a BCl-2 inhibitor, in the 50% of AL patients whose plasma cells harbor the t(11:14) translocation. However, for patients with AL amyloidosis, therapy is a race against time because these patients can die if their disease is too advanced at diagnosis, due to organ dysfunction (most commonly cardiac) before (and even after) they achieve an effective hematologic response.

To date, there are no proven therapies that are effective in increasing removal of amyloid from tissues. Neod-001, a therapeutic antibody, was thought to be promising, but its development was terminated when a phase III trial did not demonstrate efficacy. There is one antiamyloid antibody, Cael-01, that is currently being tested in clinical trials. The antibiotic doxycycline disrupts amyloid fibrils in vitro, and there are some clinical data that suggest it may reduce mortality.

WALDENSTRÖM MACROGLOBULINEMIA AND SMOLDERING WALDENSTRÖM MACROGLOBULINEMIA

WM is a rare hematologic malignancy characterized by accumulation of malignant IgM-secreting lymphoplasmacytic lymphoma cells in the bone marrow and lymph nodes.[20] *MYD88* L256P is the somatic point mutation that is found in more than 90% of patients with WM. Those patients without the mutation appear to be at higher risk of transformation to diffuse large B-cell lymphoma. Approximately 30% to 40% of patients have mutations in *CXCR4*, and these patients typically have higher serum IgM levels and lower rates of lymphadenopathy and hepatosplenomegaly. The clonal lymphoplasmacytic cells express CD19, CD20, CD22, and CD79a (lymphoid population) and CD38 on the plasmacytic population.

Epidemiology

The age-adjusted incidence for males and females is approximate 0.9 and 0.3 per 100,000 person-years, respectively.[21] The median age at presentation is 69 years. It is more common in Whites than African Americans, and there is a familial predisposition.

Clinical Presentation and Diagnosis

Most symptomatic patients present with anemia and fatigue and less often with symptoms and signs of hyperviscosity. A large proportion of patients are diagnosed when they are asymptomatic, and these patients are referred to as smoldering WM.

Differential Diagnosis

The differential diagnosis is smoldering WM, IgM MGUS, AL amyloidosis, cryoglobulinemia, Schnitzler syndrome, other lymphomas, and IgM myeloma.

Management and Prognosis

The International Prognostic Staging System for WM includes age plus serum albumin, lactate dehydrogenase, and beta-2 microglobulin levels. The median diseases specific survival is approximately 10 years. The size of the M protein spike and the degree of anemia are also prognostic.

For patients who are symptomatic (e.g., bulky disease or the presence of cytopenias), either bendamustine and rituximab or ibrutinib with or without rituximab are the mainstays of first-line therapy. If the patient has symptomatic hyperviscosity, plasmapheresis is indicated followed by the therapy noted earlier. For the elderly or less symptomatic patients, single-agent rituximab with or without an alkylator can be tried, but more often these patients are given bendamustine and rituximab as well. In contrast, patients with smoldering WM are observed.

MONOCLONAL GAMMOPATHY OF CLINICAL SIGNIFICANCE

MGCS was coined after the expression monoclonal gammopathy of renal significance (MGRS, see later).[22,23] It became increasingly apparent that a diagnosis was required for a patient with a small B-cell clone and low-level monoclonal protein who was presenting with serious and even life-threatening disease. Under these circumstances, diagnosing these patients with MGUS was inappropriate. In addition, some of these patients satisfied the definition for SMM or smoldering WM but did not require treatment directed at their clone because they did not have the classic end-organ damage that defines active MM or WM. Although nearly half of AL amyloidosis cases could be called MGCS, the expert community decided that AL amyloidosis is discrete and complex enough to stand on its own as a separate diagnosis. The MGCS can be broken down into different systems that are affected, the most common of which are kidney, nerve, and skin (Table 79.2). There is overlap because some of these rare disorders can have a more systemic, multiorgan presentation and/or course.

Monoclonal Gammopathy of Renal Significance

This was the first consensus of monoclonal gammopathies of clinical significance. There have been multiple iterations since the first publication by Leung et al. in 2012.[24] Fig. 79.4 demonstrates the spectrum of diagnoses based on location and type of deposits. All of these diagnoses are made by the renal pathologist.

With the exception of C3 glomerulopathy with monoclonal gammopathy and thrombotic microangiopathy, all other MGRS are broken into either "nonorganized" or "organized" monoclonal immunoglobulin deposits. The nonorganized includes two disease entities: monoclonal immunoglobulin deposition disease (see later), and proliferative glomerulonephritis with monoclonal immunoglobulin deposits (PGNMID). In contrast, the organized deposits can be further broken down into fibrillar deposits, which include AL amyloidosis (previously discussed) and monoclonal fibrillary glomerulonephritis (GN); and microtubular, which includes immunotactoid GN and cryoglobulinemia GN. Finally, there is the subcategory of inclusions or crystalline deposits, which includes light-chain proximal tubulopathy, crystal storing histiocytosis, and cryocrystalglobulin GN.

This classification allows for a uniform vocabulary among nephrologists, renal pathologists, and hematologists such that diagnostic algorithms can be constructed, and natural history and therapeutic interventions can be analyzed. As an example, careful study has clarified that patients with MGCS-associated fibrillary GN, immunotactoid GN, and PGNMID, who receive renal allografts, commonly have disease recurrence in the allograft. This is less commonly seen in patients with AL amyloidosis. Which diseases best respond to which plasma cell–directed therapies remains a work in progress. The most data exist on light-chain deposition disease, so that topic is described next. Cryoglobulinemia is discussed under dermatologic MGCS. An extensive review of MGRS was published by Leung and colleagues.[24]

Light-Chain Deposition Disease

Light-chain deposition disease shares similarities with AL amyloidosis in that both are immunoglobulin deposition diseases. Both can rarely involve immunoglobulin heavy chains as well. Light-chain deposition disease is less common than AL amyloidosis, is more often due to a kappa-restricted light chain (IGKV4), and presents with impaired creatinine clearance. This disease primarily

TABLE 79.2 Summary of Disease Characteristics of Selected Monoclonal Gammopathies of Clinical Significance

	Peripheral Nerves	Kidney[a]	Skin	Heart	Liver	Lymph Nodes	Gastrointestinal	Eyes	Lungs
Systemic AL amyloidosis[b]	+	++	++	++	+	+	+	+[b]	+
Light-chain deposition disease	0–+	+++	0–+	+	+	0	0–+	0	0
POEMS syndrome	+++	+	++	0–+	++	++	+	++	++
DADS-M PN	+++	0	0	0	0	0	0	0	0
Cryoglobulinemia	+–++	+	+++	0	++	+	0	+	0
Scleromyxedema	+	0	+++	+	–	–	++	–	+
Necrobiotica xanthogranuloma	0	– to +	+++	0–+	–	–	0–+	+, ++[c]	0–+
Schnitzler syndrome	0	0	+++	0	+	++	0	0	0
TEMPI syndrome	0	+++	+++	0	0	0	0	0	+++
Clarkson disease	0	++	0	+	0	0	0	0	++

[a]MGRS diagnoses are not included in this table, since all by definition are renal diseases.
[b]Not officially deemed a MGCS.
[c]Periorbital tissues only.
0, ~0%; +, 1%–39%; ++, 40%–99%; +++, ~100%.
AL, Immunoglobulin light chain; *DADS-M PN*, distal, acquired demyelinating, symmetric, neuropathy with M protein; *MGRS*, monoclonal gammopathy of renal significance; *MGCS*, monoclonal gammopathy of clinical significance; *POEMS*, polyradiculoneuropathy, organomegaly, endocrinopathy, monoclonal plasma cell disorder, and skin changes; *TEMPI*, telangiectasias; elevated erythropoietin and erythrocytosis; monoclonal gammopathy; perinephric fluid collections; and intrapulmonary shunting.

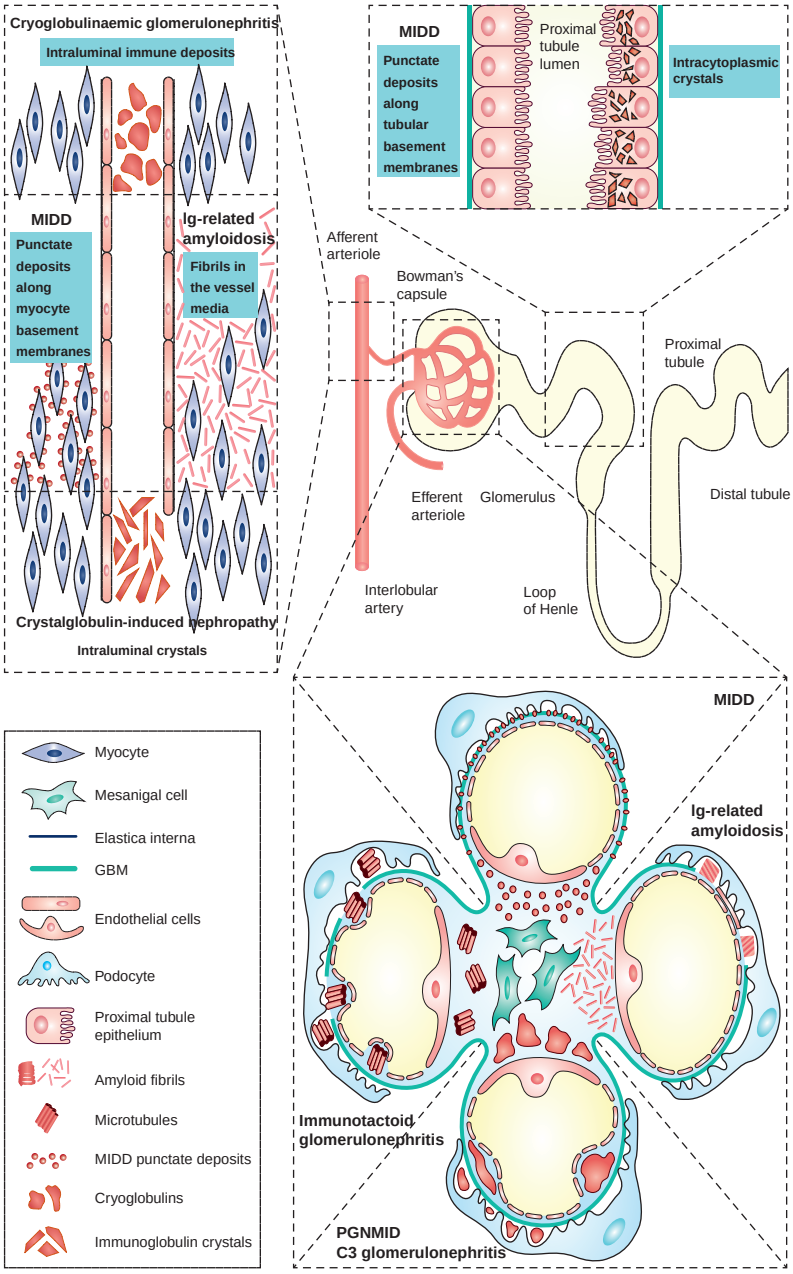


FIG. 79.4 Monoclonal Gammopathy of Renal Significance. Monoclonal gammopathy of renal significance (MGRS)-associated lesions can involve one or more renal compartments. In immunotactoid glomerulonephritis, C3 glomerulopathy, and proliferative glomerulonephritis with monoclonal immunoglobulin deposits (*PGNMIDs*), MGRS-associated lesions involve only the glomeruli, whereas in light-chain proximal tubulopathy (LCPT), MGRS-associated lesions involve only the proximal tubules. MGRS-associated lesions in cryoglobulinemic glomerulonephritis mainly involve the glomeruli but can occasionally affect blood vessels in the form of intravascular cryoglobulin thrombi or endovasculitis. Immunoglobulin *(Ig)*-related amyloidosis and monoclonal immunoglobulin deposition disease (*MIDD*) usually affect all renal compartments, including glomeruli, vessels, and the tubulointerstitium. MGRS-associated lesions can involve one or more renal compartments. In immunotactoid glomerulonephritis, C3 glomerulopathy, and PGNMIDs, MGRS-associated lesions involve only the glomeruli, whereas in LCPT, MGRS-associated lesions involve only the proximal tubules. MGRS-associated lesions in cryoglobulinemic glomerulonephritis mainly involve the glomeruli but can occasionally affect blood vessels in the form of intravascular cryoglobulin thrombi or endovasculitis. Immunoglobulin-related amyloidosis and MIDD usually affect all renal compartments, including glomeruli, vessels, and the tubulointerstitium. *GBM,* Glomerular basement membrane. (Redrawn with permission from Leung, N., Bridoux F, Batuman V, et al. The evaluation of monoclonal gammopathy of renal significance: a consensus report of the International Kidney and Monoclonal Gammopathy Research Group. *Nat Rev Nephrol.* 2019;15[1]:45–59.)

affects the kidneys but can occasionally affect other organs. As in the case of AL amyloidosis, glycosylation may be a risk factor for the disease.[25] Also like AL amyloidosis, the median plasmacytosis in the bone marrow is approximately 10%, making half with myeloma and the other half with a bone marrow that would appear to be MGUS (now classified as MGRS).

On immunofluorescence of the kidney biopsy, linear deposits of the involved monoclonal immunoglobulin are seen along

tubular and glomerular basement membranes. On electron microscopy, these deposits have a granular appearance.

Treatment is similar to that of patients with AL amyloidosis, and prognosis tends to be better because it is extraordinarily unusual for patients with light-chain deposition disease to have cardiac involvement, which is the major determinant of poor outcomes in patients with AL amyloidosis.

Monoclonal Gammopathy of Neurologic Significance

The three most common monoclonal gammopathies that have neurologic significance are POEMS syndrome, cryoglobulinemia, and AL amyloidosis (see Table 79.2). The neuropathy associated with cryoglobulinemia and AL amyloidosis is not discussed here because these diseases are covered in other sections, but suffice it to say that the former is more often an axonal neuropathy due to vasculitis and the latter a small-fiber neuropathy, which can eventually become axonal. Distal, acquired demyelinating, symmetric neuropathy with M protein (DADS-M), previously known as MGUS-associated neuropathy, is an entity that probably has significance only among patients with IgM monoclonal proteins. Scleromyxedema can sometimes cause neuropathy. Finally, there is a condition called sporadic late-onset nemaline myopathy, which is not a neuropathy but causes severe motor issues as will be discussed.

POEMS Syndrome

POEMS syndrome is the acronym for polyradiculoneuropathy, organomegaly, endocrinopathy, monoclonal plasma cell disorder, and skin changes.[26] It is a rare condition with a prevalence of 3 per million people in Japan. Median age at presentation is in the 50s, and there are slightly more men than woman affected. Symptoms and signs refer to the acronym as well as features covered by the other acronym associated with this syndrome, PEST (papilledema, extravascular volume overload, sclerotic bone lesions, thrombocytosis, and erythrocytosis). Other findings not covered by either of the acronyms are elevated vascular endothelial growth factor, pulmonary hypertension, reduced diffusion capacity of lungs for carbon monoxide (DLCO), and arterial and venous thromboembolism. The dominant symptom in this disease is a progressive length-dependent ascending sensorimotor peripheral neuropathy. Diagnostic criteria are shown in Table 79.3.

The most significant risk factors in this disease are age, pleural effusion, reduced estimated glomerular filtration rate (eGFR), and pulmonary hypertension. Coexisting Castleman disease also is an adverse risk factor as is lack of complete hematologic response. ASCT is a favored therapy, but lenalidomide and dexamethasone are also active. Data are emerging with proteasome inhibitors and daratumumab. Overall survival in patients with POEMS syndrome is excellent with plasma cell–directed therapy, with estimated 10-year survivorship at 79%.

Distal, Acquired Demyelinating, Symmetric Neuropathy With M Protein

The IgM monoclonal gammopathy accounts for approximately 60% of neuropathies associated with monoclonal gammopathy.[27] Pathologic studies in WM and IgM DADS-M have identified demyelination and widened myelin lamellae with IgM deposits detected in the widened lamellae of myelin fibers and myelin debris contained in Schwann cells and macrophages. These M proteins may bind to myelin-associated glycoprotein (MAG) or other gangliosides. However, anti-MAG antibodies are not specific for peripheral neuropathy, and reduction in anti-MAG antibody titers with rituximab or other anti-CD20 antibodies has not correlated with clinical improvement.

Patients are more often male and in their 50s to 80s. They present with a distal, demyelinating symmetric neuropathy, with sensory ataxia the most common finding. The diagnosis is one of exclusion. Even in the presence of a monoclonal gammopathy, other explanations including inherited neuropathies, diabetes, alcoholism, and drugs should be ruled out, as should POEMS syndrome and AL amyloidosis. Anti-MAG antibodies can be ordered, but as mentioned, these are not specific. Treatments include intravenous immunoglobulin (IVIG) and rituximab.

TABLE 79.3 Criteria for the Diagnosis of POEMS Syndrome[a]

Mandatory major criteria	1. Polyneuropathy (typically demyelinating) 2. Monoclonal plasma cell-proliferative disorder (almost always lambda)
Other major criteria (one required)	3. Castleman disease[a] 4. Sclerotic bone lesions 5. Vascular endothelial growth factor elevation
Minor criteria	6. Organomegaly (splenomegaly, hepatomegaly, or lymphadenopathy) 7. Extravascular volume overload (edema, pleural effusion, or ascites) 8. Endocrinopathy (adrenal, thyroid,[b] pituitary, gonadal, parathyroid, pancreatic[b]) 9. Skin changes (hyperpigmentation, hypertrichosis, glomeruloid hemangiomata, plethora, acrocyanosis, flushing, white nails) 10. Papilledema 11. Thrombocytosis/polycythemia[c]
Other symptoms and signs	Clubbing, weight loss, hyperhidrosis, pulmonary hypertension/restrictive lung disease, thrombotic diatheses, diarrhea, low vitamin B_{12} values

[a]There is a Castleman disease variant of POEMS syndrome that occurs *without* evidence of a clonal plasma cell disorder that is not accounted for in this table. This entity should be considered separately.
[b]Because of the high prevalence of diabetes mellitus and thyroid abnormalities, this diagnosis alone is not sufficient to meet this minor criterion.
[c]Approximately 50% of patients will have bone marrow changes that distinguish it from a typical monoclonal gammopathy of undetermined significance (MGUS) or myeloma bone marrow.[36] Anemia and/or thrombocytopenia are distinctively unusual in this syndrome unless Castleman disease is present.
The diagnosis of POEMS syndrome is confirmed when both of the mandatory major criteria, one of the three other major criteria, and one of the six minor criteria are present.
POEMS, Polyneuropathy, organomegaly, endocrinopathy, monoclonal plasma cell disorder, and skin changes.

Sporadic Late-Onset Nemaline Myopathy

Sporadic late-onset nemaline myopathy (SLONM) is a rare muscle disease that can be associated with a monoclonal protein or human immunodeficiency virus (HIV) infection. On biopsy, muscle fibers accumulate nemaline rods, with no associated inflammation. Patients present with predominantly proximal or axial muscle weakness, including respiratory muscle weakness.[28] Treatment strategies include IVIG and plasma cell–directed therapies, including ASCT.

Monoclonal Gammopathy of Dermatologic Significance

These conditions include Schnitzler syndrome, scleromyxedema, necrobiotic xanthogranuloma (NXG), TEMPI (telangiectasias; elevated erythropoietin and erythrocytosis; monoclonal gammopathy; perinephric fluid collections; and intrapulmonary shunting) syndrome, cryoglobulinemia, capillary leak syndrome, and POEMS syndrome (Table 79.4). This last entity is described

earlier, but these patients have skin findings including hyperpigmentation, hypertrichosis, thickening, rubor, and white nails.

Schnitzler Syndrome

Schnitzler syndrome is a condition that is characterized primarily by chronic urticaria and the presence of an IgM monoclonal gammopathy. It was described in 1972 by the French dermatologist Liliane Schnitzler and is currently classified as an autoinflammatory disorder. Interleukin (IL)-1β plays a critical role in in the disease, with aberrant NLRP3 inflammasome signaling and cytokine pathway dysregulation having been identified. Schnitzler syndrome can rarely be associated with an IgG monoclonal gammopathy. Other features and diagnostic criteria are shown in Table 79.4.[29] Although the presence of a dermal neutrophilic infiltrate on skin biopsy became a minor criterion with the Strasbourg revision, such a biopsy finding is nonspecific, and its relative importance in the diagnostic criteria has been questioned. Therapy with the anti–IL-1 monoclonal antibody anikinra and newer anti–IL-1 monoclonal antibodies (rilonacept and canakinumab) are effective.

Scleromyxedema

Scleromyxedema affects the skin primarily but can also affect other systems. It is characterized by generalized papular and sclerodermoid cutaneous eruptions and is typically associated with an IgG monoclonal gammopathy. Aside from cutaneous involvement, extracutaneous involvement can include the nervous system, joints, gastrointestinal system, and heart. The infiltrates are composed of mucin. The mechanism associated with a PCD and its monoclonal protein–inducing fibroblast proliferation remains unknown but is gradually emerging through skin transcriptome analyses and the study of peripheral blood immune cells. Transforming growth factor β (TGF-β) is overexpressed as well as other proteins, including collagen 1a and several interferon-inducible proteins.[30,31] Diagnostic criteria are shown in Table 79.5.

The two mainstays of treatment in this disease are IVIG and plasma cell–directed therapy. In one study, baseline levels of peripheral blood Tc17 cells ($CD8^+CCR\gamma^+CXCR3^+CCR4^-$) correlated with extent of skin involvement and decreased after IVIG therapy. RNA analysis of skin tissue before and after treatment revealed a decrease in gene expression of TGF-β–induced cytokines and several interferon-inducible proteins.[30] The combination of IVIG, dexamethasone, and either lenalidomide or bortezomib seems to be effective therapy in the majority of patients.[31] IVIG is considered first-line therapy, with plasma cell–directed therapy added if no response or more severe disease.

Necrobiotic Xanthogranuloma

NXG is a non-Langerhans cell histiocytosis typically associated with monoclonal proteins attributable to PCD or lymphoproliferative disorders (LPDs).[32] It was first described by Kossard and Winkelmann in 1980. The classic presentation is yellow to orange papules, plaques, and/or nodules involving the eyelids. Cutaneous lesions may also be found on other places of the face, the trunk, and the extremities. NXG plaques can occasionally be pruritic and also painful, especially if they ulcerate. Extracutaneous involvement includes the eye, heart, gastrointestinal tract, liver, and lung but is relatively rare.

On biopsy, palisading granulomas are found with nonclonal lymphoplasmacytic infiltrates and zones of necrobiosis. Cholesterol clefts and large bizarre foreign body giant cells are also classic. The pathogenesis of the disease is unknown, but it has been speculated that there is a monoclonal protein–lipoprotein interaction. Diagnostic criteria have been proposed (Table 79.6).

TABLE 79.4 Strasbourg Diagnostic Criteria of Schnitzler Syndrome

Obligate Criteria
1. Chronic urticarial rash and
2. Monoclonal immunoglobulin M (IgM) or IgG

Minor Criteria
1. Recurrent fever[a]
2. Objective findings of abnormal bone remodeling with or without bone pain[b]
3. A neutrophilic dermal infiltrate on skin biopsy[c]
4. Leukocytosis and/or elevated C-reactive protein (CRP)[d]

Definite Diagnosis
If IgM, both obligate criteria AND at least 2 minor criteria
If IgG both obligate criteria AND 3 minor criteria

Probable Diagnosis
If IgM, both obligate criteria AND 1 minor criteria
If IgG, both obligate criteria AND 2 minor criteria

[a]Must be greater than 38°C, and otherwise unexplained. Occurs usually—but not obligatory—together with the skin rash.
[b]As assessed by bone scintigraphy, magnetic resonance imaging (MRI), or elevation of bone alkaline phosphatase.
[c]Corresponds usually to the entity described as "neutrophilic urticarial dermatosis"; absence of fibrinoid necrosis; and significant dermal edema.
[d]Neutrophils greater than 10,000/mm^3 and/or C-reactive protein *(CRP)* greater than 30 mg/L.
From Simon, A., Asli, B., Braun-Falco, M., et al. Schnitzler's syndrome: diagnosis, treatment, and follow-up. *Allergy.* 2013;68:562–568.

TABLE 79.5 Rongioletti Diagnostic Criteria for Scleromyxedema

1. Generalized papular and sclerodermoid eruption
2. Evidence of monoclonal gammopathy
3. Microscopic triad associating dermal mucin deposition, thickened collagen, and fibroblast proliferation or an interstitial granuloma annulare–like pattern
4. Absence of thyroid disease

From Rongioletti, F., Merlo, G., Carli, C., et al. Histopathologic characteristics of scleromyxedema: A study of a series of 34 cases. *J Am Acad Dermatol.* 2016;74:1194–1199.

TABLE 79.6 Proposed Diagnostic Criteria for Necrobiotic Xanthogranuloma[a]

Major Criteria
1. Cutaneous papules, plaques, and/or nodules, most often yellow or orange in color
2. Histopathologic features demonstrating palisading granulomas with lymphoplasmacytic infiltrate and zones of necrobiosis. Characteristic features that are variably present include cholesterol clefts and/or giant cells (Touton or foreign body)

Minor Criteria
1. Paraproteinemia, most often IgG-κ, plasma-cell dyscrasia, and/or other associated lymphoproliferative disorder
2. Periorbital distribution of cutaneous lesions

[a]Both major criteria and at least 1 minor criterion are required for diagnosis, applicable only in the absence of foreign body, infection, or other identifiable cause.
From Nelson, C.A., Zhong, C.S., Hashemi, D.A., et al. A Multicenter Cross-Sectional Study and Systematic Review of Necrobiotic Xanthogranuloma With Proposed Diagnostic Criteria. *JAMA Dermatol.* 2020;156(3):270–279.

The leading differential diagnosis is necrobiosis lipoidica, which is a necrotizing skin condition that most often occurs in patients with diabetes mellitus but can also occur in patients with rheumatoid arthritis.

Treatment with IVIG is one of the most promising therapies, but limited success has also been reported with plasma cell–directed therapies, intralesional triamcinolone, and antimalarials.

Cryoglobulinemia

Cryoglobulinemia is a multisystem disease that can affect almost any organ system. It is discussed under dermatologic conditions because cutaneous manifestations are almost always present (see Table 79.2). A cryoglobulin is an immunoglobulin that precipitates at a low temperature but that can dissolve upon warming. The Brouet classification divides cryoglobulins into three types: type I, a monoclonal protein; type II, polyclonal immunoglobulins that form immune complexes with one or more monoclonal immunoglobulins; and type III polyclonal immunoglobulins. Type I cryoglobulins are most often IgM followed by IgG. Type II cryoglobulins are typically IgM with rheumatoid factor activity that binds to the Fc of the IgGs, which are themselves bound to antigen—most often hepatitis C viral antigens. Type I cryoglobulins arise from PCD or LPDs. In contrast, type II and type III cryoglobulins may be related to PCDs or LPDs but are more often due to infections, particularly hepatitis C, or connective tissue disorders.

Manifestations are quite variable, as is the severity of disease. This is in part due to the immunoglobulin level and biophysical characteristics (see Table 79.2). Type I cryoglobulins more often cause occlusive symptoms due to occlusion of capillary lumina, whereas vasculitis is uncommon.[33] Low complement levels and rheumatoid factor are rare. Patients report cold-induced skin symptoms, including purpura, livedo, and cold urticaria. Ulceration can occur. Fewer than one-third of patients will have renal involvement, but as many as 50% may have peripheral neuropathy. In contrast, in type II/III cryoglobulinemia or mixed cryoglobulinemia, small vessel vasculitis is the major mechanism driving morbidity. Skin symptoms including purpura occur in the vast majority of patients; arthralgia is also very common, as is peripheral neuropathy followed by renal involvement.

Treatment for cryoglobulinemia is based on the underlying cause. For those driven by PCD or LPD, clone-directed therapy is appropriate. For patients with HCV, which comprises approximately 70% to 90% of these cases, treating the underlying hepatitis is most appropriate. Sustained virologic responses can be achieved in greater than 50% of these patients. For patients not responding to antiviral therapy, rituximab and other immunosuppressants can play an important role in treating the vasculitis. Plasmapheresis can be used in patients with severe end-organ damage and/or refractory disease. Other disease modifiers including corticosteroids and cyclophosphamide can also play a role in therapy. For disease driven by autoimmunity, rituximab and corticosteroids are the best first-line option.

TEMPI Syndrome

TEMPI syndrome is a rare (22 reported cases as of December 2019) acquired disorder characterized by the features that comprise the acronym: telangiectasias; elevated erythropoietin and erythrocytosis; monoclonal gammopathy; perinephric fluid collections; and intrapulmonary shunting.[34] The underlying pathophysiology is not understood, but it is clear that plasma cell–directed therapy reverses the clinical manifestations.

TABLE 79.7 Proposed Diagnostic Criteria for TEMPI Syndrome

Major
1. Telangiectasias
2. Elevated erythropoietin and erythrocytosis
3. Monoclonal gammopathy

Minor
1. Perinephric fluid
2. Intrapulmonary shunting

Other
1. Venous thrombosis

Patients most often present with telangiectasias involving the face and upper body and erythrocytosis (see Table 79.2). Unlike the erythrocytosis of polycythemia rubra vera and of POEMS syndrome, patients with TEMPI syndrome have a high erythropoietin. Patients develop progressive hypoxia, but the underlying pulmonary shunting is not evident on high-resolution chest CT and is best demonstrated by ^{99m}Tc macro-aggregated albumin scintigraphy. The perinephric fluid collections have the same electrolyte composition as serum. Proposed diagnostic criteria are shown in Table 79.7. Unlike POEMS syndrome, there is no bias in clonality for lambda-restricted clones and there are no features of a myeloproliferative neoplasm.

Plasma cell–directed therapy appears to be useful, specifically bortezomib, daratumumab, lenalidomide, and high-dose melphalan. All of the features can improve upon achievement of a complete hematologic response.

Idiopathic Systemic Capillary Leak Syndrome (Clarkson Disease)

This devastating disease was described by Dr. Bayard Clarkson in 1960. Systemic capillary leak syndrome (SCLS) is characterized by capillary leak resulting in sudden-onset shock and anasarca caused by plasma extravasation (up to 70% of total plasma volume). The diagnostic triad is composed of the "3 Hs," hypotension, hemoconcentration, and hypoalbuminemia, which occur in the absence of secondary causes. Sixty-eight percent of adult cases of SCLS have monoclonal proteins, most commonly IgG kappa. There is no clear pathologic role for the monoclonal protein in this disorder, other than its presence in the majority of cases. A recent review provides the current understanding of the disease mechanism(s) associated with the vascular endothelial hyperpermeability in SCLS.[35]

The differential diagnosis for an acute attack includes sepsis, anaphylaxis, and hereditary angioedema. Treatment at the time

ON THE HORIZON

- Innovative diagnostics will emerge that more precisely differentiate the various forms of monoclonal gammopathies
- Novel immunotherapies are in clinical trials to treat multiple myeloma
- New approaches to therapy should eventually prove to be helpful for managing AL amyloidosis and the spectrum of the monoclonal gammopathy of clinical significance (MGCS)
- The MGCS represent a group of disorders whose pathogenic mechanisms will be unraveled in the future

of an acute attack is supportive with fluid resuscitation until the flare subsides, which typically occurs over the course of days. Empiric prophylaxis with IVIG is recommended based on the demonstration of fewer attacks with this therapy.

REFERENCES

1. Mills JR, Kohlhagen MC, Dasari S, et al. Comprehensive assessment of M-Proteins using nanobody enrichment coupled to MALDI-TOF mass spectrometry. *Clin Chem*. 2016;62(10):1334–1344. https://doi.org/10.1373/clinchem.2015.253740. PubMed PMID: 27540026.
2. Kyle RA, Larson DR, Therneau TM, et al. Long-term follow-up of monoclonal gammopathy of undetermined significance. *N Engl J Med*. 2018;378(3):241–249. https://doi.org/10.1056/NEJMoa1709974. Epub 2018/01/18. PubMed PMID: 29342381. PMCID: PMC5852672.
3. Rajkumar SV, Kyle RA, Therneau TM, et al. Serum free light chain ratio is an independent risk factor for progression in monoclonal gammopathy of undetermined significance (MGUS). *Blood*. 2005;106:812–817. PMCID: PMC1895159.
4. Kyle RA, Remstein ED, Therneau TM, et al. Clinical course and prognosis of smoldering (asymptomatic) multiple myeloma. *N Engl J Med*. 2007;356(25):2582–2590. https://doi.org/10.1056/NEJMoa070389. Epub 2007/06/22. PubMed PMID: 17582068.
5. Kyle RA, Greipp PR. Smoldering multiple myeloma. *N Engl J Med*. 1980;302(24):1347–1349.
6. Lakshman A, Rajkumar SV, Buadi FK, et al. Risk stratification of smoldering multiple myeloma incorporating revised IMWG diagnostic criteria. *Blood Cancer J*. 2018;8(6):59. https://doi.org/10.1038/s41408-018-0077-4. Epub 2018/06/14. PubMed PMID: 29895887; PMCID: PMC5997745.
7. Mateos MV, Gonzalez-Calle V. Smoldering multiple myeloma: who and when to treat. *Clin Lymphoma Myeloma Leuk*. 2017;17(11):716–722. https://doi.org/10.1016/j.clml.2017.06.022. Epub 2017/07/16. PubMed PMID: 28709797.
8. Lonial S, Jacobus S, Fonseca R, et al. Randomized trial of lenalidomide versus observation in smoldering multiple myeloma. *J Clin Oncol*. 2019:JCO1901740. https://doi.org/10.1200/JCO.19.01740. Epub 2019/10/28. PubMed PMID: 31652094.
9. Harris NL, Jaffe ES, Stein H, et al. A revised European-American classification of lymphoid neoplasms: a proposal from the International Lymphoma Study Group. *Blood*. 1994;84(5):1361–1392. PubMed PMID: 8068936.
10. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. *CA Cancer J Clin*. 2018;68(1):7–30. https://doi.org/10.3322/caac.21442. PubMed PMID: 29313949.
11. Palumbo A, Avet-Loiseau H, Oliva S, et al. Revised International Staging System for multiple myeloma: A report from International Myeloma Working Group. *J Clin Oncol*. 2015;33(26):2863–2869. https://doi.org/10.1200/JCO.2015.61.2267. Epub 2015/08/05. PubMed PMID: 26240224.
12. Attal M, Lauwers-Cances V, Hulin C, et al. lenalidomide, bortezomib, and dexamethasone with transplantation for myeloma. *N Engl J Med*. 2017;376(14):1311–1320. https://doi.org/10.1056/NEJMoa1611750. PubMed PMID: 28379796.
13. Soekojo CY, Ooi M, de Mel S, Chng WJ. Immunotherapy in multiple myeloma. *Cells*. 2020;9(3):601 https://doi.org/10.3390/cells9030601. PubMed PMID: 32138182; PMCID: PMC7140529.
14. Gundesen MT, Lund T, Moeller HEH, Abildgaard N. Plasma cell leukemia: definition, presentation, and treatment. *Curr Oncol Rep*. 2019;21(1):8https://doi.org/10.1007/s11912-019-0754-x. PubMed PMID: 30689121; PMCID: PMC6349791.
15. Nalghranyan S, Singh AP, Schinke C. The combination of venetoclax, daratumumab and dexamethasone for the treatment of refractory primary plasma cell leukemia. *Am J Hematol*. 2020;95(2):E34–E35. https://doi.org/10.1002/ajh.25676. PubMed PMID: 31709578.
16. Caers J, Paiva B, Zamagni E, et al. Diagnosis, treatment, and response assessment in solitary plasmacytoma: updated recommendations from a European Expert Panel. *J Hematol Oncol*. 2018;11(1):10. https://doi.org/10.1186/s13045-017-0549-1. PubMed PMID: 29338789; PMCID: PMC5771205.
17. Merlini G, Dispenzieri A, Sanchorawala V, et al. Systemic immunoglobulin light chain amyloidosis. *Nat Rev Dis Primers*. 2018;4(1):38https://doi.org/10.1038/s41572-018-0034-3. Epub 2018/10/27. PubMed PMID: 30361521.
18. Dispenzieri A. AL patients don't dare go without dara. *Blood*. 2020;135(18):1509–1510. https://doi.org/10.1182/blood.2020005436. Epub 2020/05/01. PubMed PMID: 32353125.
19. Leung N, Thome SD, Dispenzieri A. Venetoclax induced a complete response in a patient with immunoglobulin light chain amyloidosis plateaued on cyclophosphamide, bortezomib and dexamethasone. *Haematologica*. 2018;103(3):e135–e137. https://doi.org/10.3324/haematol.2017.183749. Epub 2018/01/21. PubMed PMID: 29351984.
20. Castillo JJ, Treon SP. What is new in the treatment of Waldenström macroglobulinemia? *Leukemia*. 2019;33(11):2555–2562. https://doi.org/10.1038/s41375-019-0592-8. Epub 2019/10/09. PubMed PMID: 31591468.
21. Kyle RA, Larson DR, McPhail ED, et al. Fifty-year incidence of Waldenström macroglobulinemia in Olmsted County, Minnesota, from 1961 Through 2010: a population-based study with complete case capture and hematopathologic review. *Mayo Clin Proc*. 2018;93(6):739–746. https://doi.org/10.1016/j.mayocp.2018.02.011. Epub 2018/04/17. PubMed PMID: 29656787; PMCID: PMC5988946.
22. Leung N, Bridoux F, Hutchison CA, et al. Monoclonal gammopathy of renal significance: when MGUS is no longer undetermined or insignificant. *Blood*. 2012;120(22):4292–4295. https://doi.org/10.1182/blood-2012-07-445304. PubMed PMID: 23047823.
23. Fermand JP, Bridoux F, Dispenzieri A, et al. Monoclonal gammopathy of clinical significance: a novel concept with therapeutic implications. *Blood*. 2018;132(14):1478–1485. https://doi.org/10.1182/blood-2018-04-839480. Epub 2018/07/18. PubMed PMID: 30012636.
24. Leung N, Bridoux F, Batuman V, et al. The evaluation of monoclonal gammopathy of renal significance: a consensus report of the International Kidney and Monoclonal Gammopathy Research Group. *Nat Rev Nephrol*. 2019;15(1):45–59. https://doi.org/10.1038/s41581-018-0077-4. Epub 2018/12/05. PubMed PMID: 30510265; PMCID: PMC7136169.
25. Joly F, Cohen C, Javaugue V, et al. Randall-type monoclonal immunoglobulin deposition disease: novel insights from a nationwide cohort study. *Blood*. 2019;133(6):576–587. https://doi.org/10.1182/blood-2018-09-872028. Epub 2018/12/24. PubMed PMID: 30578255.
26. Dispenzieri A. POEMS syndrome: 2019 update on diagnosis, risk-stratification, and management. *Am J Hematol*. 2019https://doi.org/10.1002/ajh.25495. Epub 2019/04/24. PubMed PMID: 31012139.
27. Chaudhry HM, Mauermann ML, Rajkumar SV. Monoclonal gammopathy-associated peripheral neuropathy: diagnosis and management. *Mayo Clin Proc*. 2017;92(5):838850https://doi.org/10.1016/j.mayocp.2017.02.003. Epub 2017/05/06. PubMed PMID: 28473042; PMCID: PMC5573223.
28. Naddaf E, Milone M, Kansagra A, et al. Sporadic late-onset nemaline myopathy: clinical spectrum, survival, and treatment outcomes. *Neurology*. 2019;93(3):e298–e305. https://doi.org/10.1212/WNL.0000000000007777. Epub 2019/06/07. PubMed PMID: 31167932.
29. Gusdorf L, Lipsker D. Schnitzler syndrome: a review. *Curr Rheumatol Rep*. 2017;19(8):46https://doi.org/10.1007/s11926-017-0673-5. Epub 2017/07/19. PubMed PMID: 28718061.
30. Mecoli CA, Talbot Jr. CC, Fava A, et al. Clinical and molecular phenotyping in scleromyxedema pre- and post-treatment with intravenous immunoglobulin. *Arthritis Care Res (Hoboken)*. 2019;72(6):761–767. https://doi.org/10.1002/acr.23908. Epub 2019/04/23. PubMed PMID: 31008568; PMCID: PMC6810715.
31. Mahevas T, Arnulf B, Bouaziz JD, et al. Plasma cell-directed therapies in monoclonal gammopathy-associated scleromyxedema. *Blood*. 2020;135(14):1101–1110. https://doi.org/10.1182/blood.2019002300. Epub 2020/02/07. PubMed PMID: 32027747.
32. Dellatorre G, Miqueloto JK. Necrobiotic xanthogranuloma. *JAMA Dermatol*. 2020https://doi.org/10.1001/jamadermatol.2020.0897. Epub 2020/05/07. PubMed PMID: 32374351.
33. Desbois AC, Cacoub P, Saadoun D. Cryoglobulinemia: an update in 2019. *Joint Bone Spine*. 2019;86(6):707–713. https://doi.org/10.1016/j.jbspin.2019.01.016. Epub 2019/02/08. PubMed PMID: 30731128.

34. Sykes DB, O'Connell C, Schroyens W. The TEMPI syndrome. *Blood*. 2020;135(15):11991203https://doi.org/10.1182/blood.2019004216. Epub 2020/02/29. PubMed PMID: 32108223.
35. Druey KM, Parikh SM. Idiopathic systemic capillary leak syndrome (Clarkson disease). *J Allergy Clin Immunol*. 2017;140(3):663–670. https://doi.org/10.1016/j.jaci.2016.10.042. Epub 2016/12/26. PubMed PMID: 28012935; PMCID: PMC5481509.
36. Dao LN, Hanson CA, Dispenzieri A, et al. Bone marrow histopathology in POEMS syndrome: a distinctive combination of plasma cell, lymphoid and myeloid findings in 87 patients. *Blood*. 2011;117(24):6438–6444. https://doi.org/10.1182/blood-2010-11-316935. Epub 2011/03/10. PubMed PMID: 21385854.

Immune Responses to Solid Tumors and Immune Checkpoint Therapy

Satya Das, Justin M. Balko, and Douglas B. Johnson

Harnessing the immune response to target cancer has been a long-standing and elusive goal of cancer researchers. Several early advances suggested that such an approach was possible, including Coley toxins (heat-killed Serratia directly injected into tumors), cytokine therapy (specifically high-dose interleukin-2 in melanoma and renal cell carcinoma), and hematologic stem cell transplants (for certain hematologic malignancies). Despite these proof-of-concept applications, most cancer patients lack benefit to any type of immune therapy.

Several discordant threads characterized cancer immunology research over past decades. First, the concept of cancer immune surveillance suggested that the immune system was able to suppress and eliminate nascent cancers. However, even "true believers" in this concept readily admitted that established tumors seemed unimpeded by this immune surveillance. Second, a variety of immune suppressive mechanisms were identified in tumor cells, ranging from loss of antigen-presentation machinery (e.g., loss of major-histocompatibility complex [MHC] class I) to immunosuppressive subpopulations (e.g., regulatory T cells and myeloid derived suppressor cells) to promotion of immunosuppressive factors in the tumor microenvironment (e.g., hypoxia, indolamine dioxygenase). Overcoming this gauntlet of negative regulators was thought to be a daunting challenge. Third, repeated disappointing and thoroughly negative results of numerous tumor vaccine and cytokine studies further dampened enthusiasm for immunotherapy approaches.

Still, glimmers of hope remained and convinced a dedicated group of researchers to continue searching for methods to harness the immune system to elicit effective antitumor responses. It remained clear that many, if not all, tumors harbored immunologically distinct molecular features from host tissue, including somatic mutations, cancer-testis antigens, and differentiation antigens. Furthermore, many tumors were obviously infiltrated by immune cells, suggesting that these antigens might be driving an immune response. In addition, while most vaccine studies were negative, correlative data suggested that patients were mounting an immune response, and occasionally clinical evidence of antitumor and immune activation occurred, including tumor shrinkage and development of vitiligo. Other clinical hints remained, including the durable complete responses generated by high-dose interleukin-2 (IL-2) in 5% to 8% of metastatic melanoma and renal cell carcinoma patients. Other labor intensive, cellular therapy approaches also showed promise, with phase II studies of tumor-infiltrating lymphocytes showing response rates as high as 50%. Still, the ability to routinely unlock effective antitumor immune responses remains limited.

IMMUNE CHECKPOINTS

T lymphocytes constitute a major portion of adaptive immunity. Expression of a unique T-cell receptor (TCR) is a hallmark of T cells and represents the mechanism by which they can identify pathogens and cancerous tissue. During development, each T cell obtains a unique TCR—composed of an alpha (α) and beta (β) chain arising from two independent genes—through a process called TCR V(D)J gene rearrangement (see Chapter 4). Positive and negative selection of immature T cells occurs in the thymus, ensuring that each uniquely generated TCR combination can recognize an antigen with low affinity, while not completely activating against self-antigens. Breaks or failures in this process are one of the mechanisms at play in the generation of autoimmunity.

Each unique TCR, expressed in a single T-cell clone, can bind to a cognate antigen bound to a MHC on the surface of a cell. The interaction of a TCR with its MHC-antigen complex (present on the antigen-presenting cell; signal 1) together with the binding of T-cell CD28 receptor with a co-stimulatory ligand (also present on the antigen presenting cell; signal 2) can prime T cells and prepare them for functional activation. Following this initial activation, the primed T cell can activate and eliminate other MHC-antigen-presenting cells in the absence of signal 2 through a variety of cytotoxic mechanisms, or in the case of CD4 T helper cells, secrete critical supportive cytokines required for effective adaptive immune function (Fig. 80.1).

Potent and sustained inflammation has the potential to hyperactivate T cells, expanding T-cell clones with weak affinity for self-antigens. As such, development of self-reactivity in the form of autoimmunity can be a negative downstream consequence. Thus, several safety mechanisms are programmed into the activation of adaptive immunity that aid in limiting or thwarting chronic inflammation and sustained T-cell activity, called immune checkpoints. Immune checkpoints are often induced by signaling cascades parallel or even identical to those potentiating inflammation and cytotoxic activity or are present on regulatory cells that are recruited at later stages to a site of chronic inflammation. The expression of immune checkpoints and their binding to their cognate ligands negatively regulate T-cell function.

CTLA-4

CTLA-4 was identified in 1987 by the laboratory of Pierre Golstein and was later found by the same group as having highly similar structure to CD28 (responsible for "signal 2").

However, its function as an immune checkpoint was not described until 5 years later by the laboratory of Jeffrey Bluestone, who found that a soluble CTLA-4 molecule could inhibit T-cell function. Eventually it was understood that CTLA-4 functioned by competing with CD28 for co-stimulation, resulting in suppression of priming of new T-cell clones against antigen (Fig. 80.2). James Allison later applied this principle to cancer biology, demonstrating that inhibition of the CTLA-4 signal through anti-CTLA-4 antibodies could potentiate antitumor immunity, paving the way for an entire class of therapeutic molecules in cancer treatment.[1]

While there was initial skepticism about this approach clinically, eventually a clinical trial of anti-CTLA-4 was initiated. Among 14 patients treated prior to 2003 with the agent that became ipilimumab (MDX-010), three patients experienced objective responses. This agent also was associated with a novel set of autoimmune toxicities, including colitis, hypophysitis, hepatitis, dermatitis, and others, which are discussed in depth subsequently. Several phase II studies were conducted as well, demonstrating an approximately 10% to 20% response rate in advanced melanoma and suggesting that responses were durable in nature.

The responses from these studies set the stage for two randomized phase III studies in patients with advanced or metastatic melanoma, a cancer-type with no approved therapies that had demonstrated improved survival in the metastatic setting. The first was a randomized phase III study comparing ipilim-

KEY CONCEPTS

Box A: CTLA-4 Inhibition in Cancer

- Anti-CTLA-4 antibodies are associated with relatively modest activity in most cancers, although it was the first class of agents to improve overall survival in melanoma.
- Ipilimumab demonstrated key proof of principle features of immune-checkpoint inhibitor response, including atypical responses (including growth before shrinkage), durability, and novel toxicity profiles.
- The role of ipilimumab moving forward is largely as combination therapy.

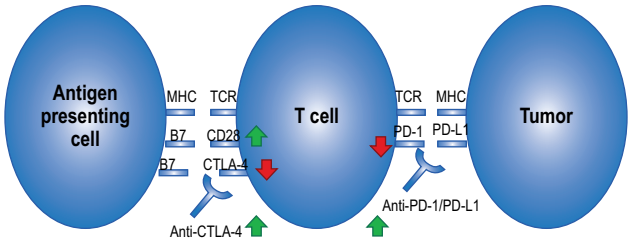


FIG. 80.2 Simplified schematic of mechanism of action of immune-checkpoint inhibitors; anti-CTLA-4 is highlighted on the left, anti-PD-1/PD-L1 on the right. *MHC*, Major histocompatibility complex; *TCR*, T-cell receptor.

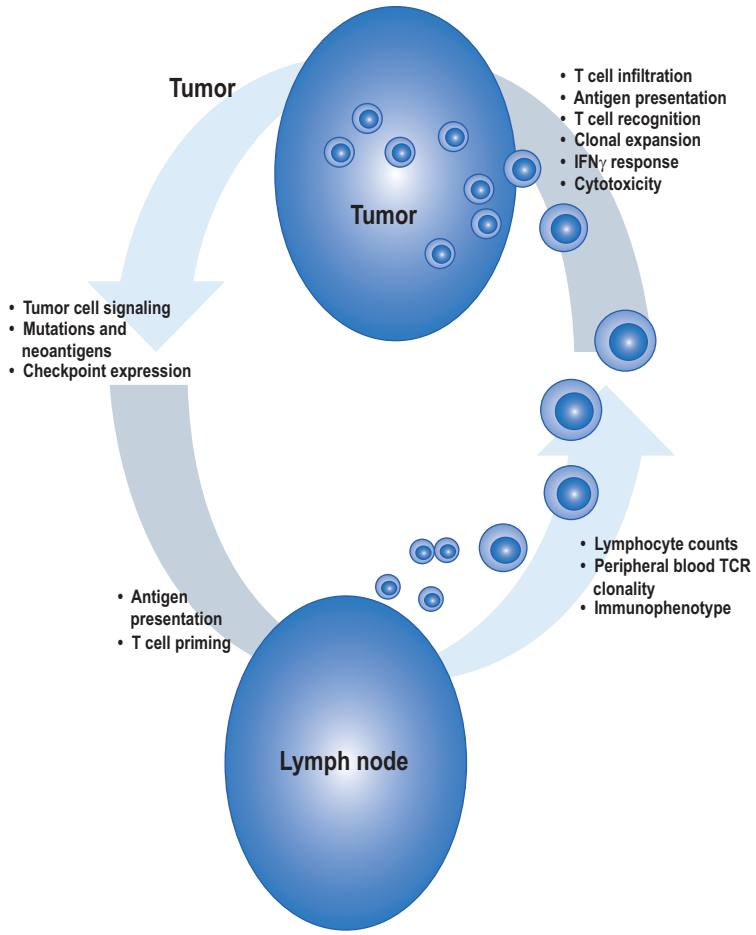


FIG. 80.1 Necessary components for generating an antitumor response against cancer. *IFN γ*, Interferon-γ; *TCR*, T-cell receptor.

umab with a gp100 vaccine and the combination in 676 patients who had already received other treatments (chemotherapy or high-dose IL-2). Both ipilimumab containing arms improved overall survival (OS) compared with chemotherapy (median survivals of approximately 10 months vs. 6 months). The second was a randomized phase III study comparing ipilimumab (given at a higher dose: 10 mg/kg) and dacarbazine (850 mg/m2) versus dacarbazine alone in patients without prior treatment. The combination arm was also associated with improved survival (median 11.2 months vs. 9.1 months), although the combination was also associated with substantially high rates of liver toxicity. These studies ultimately led to the approval of ipilimumab (at the dose of 3 mg/kg for a maximum of four doses) for metastatic melanoma.[2] A subsequent study showed that ipilimumab 10 mg/kg improved OS compared with 3 mg/kg, albeit at the cost of increased toxicities. These data show that both the efficacy and toxicity of ipilimumab appears to be dose dependent. Currently, 3 mg/kg is the approved and widely used dose of ipilimumab for metastatic melanoma.

Subsequent studies have shown that ipilimumab responses are quite distinct from chemotherapy in many cases. Similar to the experience with IL-2, many responding patients experience long-lasting benefit, with responses now lasting more than 10 years.[3] Distinct from IL-2, often partial responses or even stable disease were quite persistent. For patients who respond but later relapse, retreatment with an additional course of ipilimumab may be effective. The phenomenon of "pseudoprogression" was also identified with ipilimumab. At times, the first scan obtained after starting treatment would demonstrate increased tumor size, and patients would be considered as having progressed (Fig. 80.3). However, subsequent scans showed tumor shrinkage, and biopsies of these "progressing" lesions showed substantial immune cell infiltration. Thus, patients occasionally (5% to 10% of the time) experience tumor growth prior to shrinkage when assessed early. Accordingly, obtaining follow-up imaging for patients on ipilimumab is deferred until about 12 weeks on therapy to decrease (but not eliminate) the chances of observing this phenomenon. Novel methods of quantitating responses were generated to allow the possibility of pseudoprogression (immune-related response criteria; IRRC and immune RECIST).[4] Of note, even among patients without pseudoprogression, responses were usually relatively slow in onset and rarely occurred in patients with rapidly progressing disease.

Ipilimumab was also tested in the adjuvant setting in high-risk stage III melanoma patients who had disease that was resected. Compared with placebo, ipilimumab improved OS (hazard ratio 0.75). However, this setting and dose (10 mg/kg) was associated with high rates of severe toxicity (greater than 50% events requiring steroids and greater than 1% treatment-related deaths). Although this regimen was approved and briefly considered a standard of care, it has been supplanted by anti-PD-1 antibody therapy.

Despite the efficacy of ipilimumab in some melanoma patients, most (more than 80%) failed to respond. Moreover, relatively limited activity was observed in other cancers, including negative studies in lung and prostate cancer, suggesting that other immune targets would be needed to more fully generate an effective immune response. Thus, anti-CTLA-4 as a single agent remains primarily useful in demonstrating the potential of immune-checkpoint inhibitors to produce durable antitumor responses.

Anti-Programmed Death-1/Programmed Death-L1

The second immune-checkpoint pathway to be discovered was that of the programmed death-1 (PD-1)/programmed-death-ligand-1 interaction.

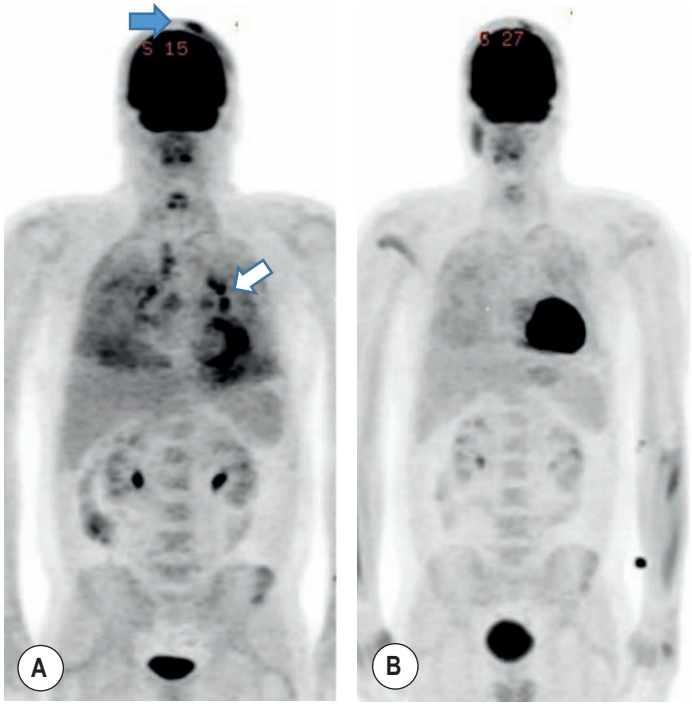


FIG. 80.3 Schematic showing pseudoprogression. Baseline scans (not shown) revealed no metastatic disease outside subcutaneous metastases in the scalp. 3-month scans *(left)*, showed increased scalp lesion *(blue arrow)* and enlarged hilar and mediastinal lymph nodes (white arrow). Biopsy of both lesions showed inflammation and no tumor cells. These lesions resolved on next imaging *(right)*.

KEY CONCEPTS

Box B: Anti-PD-1/PD-L1 Monotherapy in Cancer

- Anti-PD-1/PD-L1 antibodies are now approved in 17 different cancers, producing often durable responses in ~10% to 80% of treated patients, depending on the cancer.
- The greatest clinical activity is seen in tumors with high rates of somatic mutations, T-cell infiltration, and PD-L1 expression, which include Hodgkin lymphoma, skin cancers, and renal cell carcinoma.
- Many other solid tumors, including urothelial, lung, head and neck, and hepatocellular carcinomas have response rates in the range of 15%–20%
- Within most individual tumor types, PD-L1 expression on tumor or infiltrating immune cells correlates with higher response rates and superior clinical outcomes.

The PD-1 gene and gene-product were initially identified in the laboratory of Tasuku Honjo. The expression of PD-1 was found to be induced on T lymphocytes following activation. Upon binding of its ligand PD-L1, also identified by the Honjo and Gordon Freeman laboratories, negative signals are transmitted to the T cell, forcing it into a state of temporary, and eventually terminal, exhaustion (see Fig. 80.2). In *in vitro* and *in vivo* preclinical models, blockade of this axis using either PD-1 or PD-L1 blocking antibodies demonstrated therapeutic potential for modulating T-cell responses.[5]

Importantly, although PD-L1 may be expressed by numerous tumor types in the setting of inflammation and interferon production, as well as in mediated tolerogenic states (e.g., by the placenta during pregnancy), it is particularly highly expressed in the tumor microenvironment. Thus, one might predict that responses generated by blocking the PD-1/PD-L1 axis might be more tumor-specific than blocking the CTLA-4 axis.

Indeed, early preclinical work using murine-tumor models overexpressing PD-L1, or using knockout mice lacking the PD-L1 or PD-1 gene, demonstrated suppressive effects on tumor growth in a variety of settings. Later, more elaborate studies showed that these effects were driven by a renewal of activated T cells in the microenvironment that could elicit tumor control or even elimination. In contrast to the CTLA-4 axis which is largely thought to hold greatest activity in lymphoid tissue (e.g., tumor-draining lymph nodes and spleen), the PD-1/L1 pathway was primarily functioning as a peripheral tolerance mechanism at the site of inflammation, in this case the tumor microenvironment. These differing mechanisms of action suggested not only therapeutic potential of PD-1/L1 blockade in cancer, but also potential for combinatorial activity between the two pathways.

Anti-Programmed Death-1 in Melanoma

In view of the success of IL-2, tumor-infiltrating lymphocytes, and ipilimumab in melanoma, melanoma was a logical starting point for anti-PD-1/PD-L1 antibody trials. Early phase I data suggested response rates of greater than 30% in previously treated patients, exceeding that observed with ipilimumab. Importantly, toxicities were also less frequent than with CTLA-4 blockade, with high-grade events in 10% to 20% and requiring discontinuation in less than 5% of cases. Responses also occurred more rapidly and with lower incidence of pseudoprogression (Table 80.1 for list of FDA-approved indications for anti-PD-1 agents in cancer).

These data set the stage for several phase III trials of both nivolumab and pembrolizumab, which are briefly summarized. Patients treated with anti-PD-1 antibodies experienced statistically significant improvements in OS (median of about 24 months in initial trials), progression-free survival (PFS) (median 6 to 8 months), and objective-response rates (35% to 45%) compared with either ipilimumab or chemotherapy (depending on several different trials). Subsequent analyses have shown a 5-year survival of 44% for patients without prior treatment, with a 29% 5-year PFS.[6] Largely equivalent efficacy and toxicity profiles were observed with both pembrolizumab and nivolumab at various doses and schedules. Both agents are thought to have largely dose-independent effects and are now approved at a variety of doses and schedules.

Anti-PD-1 agents were also tested as adjuvant therapy in two phase III studies of patients with high risk, resected stage III-IV melanoma. Pembrolizumab, compared with placebo, and nivolumab, compared with ipilimumab, were both associated with superior relapse-free survival in these patients (hazard ratios 0.57 and 0.65, respectively). These agents are now standards of care for patients with high-risk, resected stage III-IV disease, and were the first anti-PD-1 agents successfully tested as adjuvant therapy.[7]

Anti-Programmed Death-1/Programmed Death-L1 in Lung Cancer

As the leading cause of cancer-related mortality, developing improved treatments in non-small-cell lung cancer (NSCLC) as well as SCLC have been major priorities. Intriguing responses were observed in initial phase I studies of nivolumab (approximately 20%), setting the stage for future studies. As second-line treatment, nivolumab was associated with improved OS compared with docetaxel (median 12.2 vs. 9.4 months in non-squamous NSCLC) in patients who previously received platinum-containing chemotherapy. Subsequently, in the first-line, pembrolizumab improved OS compared with chemotherapy in patients with NSCLC with PD-L1 expression greater than 1% (hazard ratio 0.81).[8] Nearly all the benefit, however, was in the patients with PD-L1 expression greater than 50% (median 20 months vs. 12 months). Of note, patients harboring epidermal growth factor receptor (EGFR) mutations and anaplastic lymphoma kinase gene (ALK) translocations were largely excluded from these trials, as these molecular subsets appear to have much lower rates of benefit from anti-PD-1 therapy. Thus, appropriate, molecularly-selected patients with NSCLC may receive chemotherapy-free front-line regimens. Along with chemo-immunotherapy combinations, (see Chemotherapy-containing regimens section), pembrolizumab (front-line, PD-L1$^+$) and nivolumab (second-line, irrespective of PD-L1 expression) are standard care regimens. Importantly, response rates and clinical outcomes appear largely similar in squamous vs. non-squamous NSCLC.

Although stage IV NSCLC is the largest contributor to lung cancer mortality, patients with unresectable stage III NSCLC also have very poor outcomes, with long-term cure rates of less than 15%. One phase III study assessed durvalumab (anti-PD-L1) compared with placebo following completion of chemotherapy and radiation and found prolonged survival (median 28.3 months vs. 16.2 months, hazard ratio 0.53). Importantly, pneumonitis (which can occur with both radiation and anti-PD-1/PD-L1) occurred in 4.8% of patients (compared with 2.6% of placebo-treated patients).[9] Adjuvant and neoadjuvant (e.g., treatment initiated prior to surgery) studies are ongoing for earlier stage (I and II) lung cancer.

TABLE 80.1 FDA-Approved Indications for Anti-PD-1 Antibody Monotherapy in Metastatic/Unresectable Malignancies

Drug	Disease Indication	Phase of Study [NCT Number]	Response Rate [95% CI]	Median Overall Survival (mon)
Nivolumab	First-line melanoma	Phase III [NCT01844505]	45% [39.1%–50.3%]	36.9
Pembrolizumab	First-line melanoma	Phase III [NCT01866319]	42% [38.1%–46.5%]	32.7
Nivolumab	Second-line NSCLC	Phase III (squamous) [NCT01642004] Phase III (non-squamous) [NCT01673867]	20% [14%–28%] (squamous) and 19% [15%–24%] (non-squamous)	9.2 (squamous) and 12.2 (non-squamous)
Pembrolizumab	First-line NSCLC (PD-L1 ≥ 50%)	Phase III [NCT02142738]	44.8% [36.8%–53%]	30
Nivolumab	Third-line SCLC	Phase I/II [NCT01928394]	12% [6.5%–19.5%]	4.4
Nivolumab	Second-line RCC	Phase III [NCT01668784]	25% [not mentioned]	25
Nivolumab	Second-line UCC	Phase II [NCT02387996]	19.6% [15.0%–24.9%]	8.74
Pembrolizumab	Second-line UCC	Phase III [NCT02256436]	21.2% [16.4%–26.5%]	10.3
Pembrolizumab	First-line Merkel cell carcinoma	Phase II [NCT02267603]	56% [41.3%–70%]	NR
Pembrolizumab	Any refractory MSI-H tumor	Phase II [NCT01876511]	53% [42%–64%]	NR
Cemiplimab	Cutaneous SCC not amenable to local resection	Phase I [NCT02383212] and [NCT02760498]	47% [34%–61%]	NR
Nivolumab	Later-line MSI-H CRC	Phase II [NCT02060188]	31.1% [20.8%–42.9%]	NR
Pembrolizumab	Later-line MSI-H CRC	Phase II [NCT02460198]	33% [21%–46%]	31.4
Pembrolizumab	Third-line (PD-L1 ≥ 1%) gastric and gej adenocarcinoma	Phase II [NCT02335411]	22.7% [13.8%–33.8%]	NA
Nivolumab	HCC	Phase I/II [NCT01658878]	20% [15%–26%]	NR
Nivolumab	Second-line head and neck SCC	Phase III [NCT02105636]	13.3% [9.3%–18.3%]	7.5
Pembrolizumab	Second-line head and neck SCC	Phase III [NCT02252042]	14.6% [10.4%–19.6%]	8.4
Pembrolizumab	First-line head and neck SCC (PD-L1 ≥ 1%)	Phase III [NCT02358031]	19% [not mentioned]	12.3
Nivolumab	Relapsed/refractory cHL	Phase II [NCT02181738]	69% [63%–75%]	NR
Pembrolizumab	Relapsed/refractory cHL	Phase Ib [NCT01953692]	65% [48%–79%]	NR
Pembrolizumab	Relapsed/refractory PMBCL	Phase II [NCT02576990]	45% [32%–60%]	NR
Pembrolizumab	Second-line (PD-L1 ≥ 1%) cervical carcinoma	Phase II [NCT02628067]	14.6% [7.8%–24.1%]	11

cHL, Classical Hodgkin lymphoma; *CRC*, colorectal cancer; *GEJ*, gastroesophageal junction; *mon*, months; *MSI-H*, microsatellite instability-high; *NA*, not available; *NR*, not reached; *NSCLC*, non-small cell lung cancer; *PMBCL*, primary mediastinal B-cell lymphoma; *RCC*, renal cell carcinoma; *SCC*, squamous cell carcinoma; *SCC*, squamous cell carcinoma; *SCLC*, small cell lung cancer; *UCC*, urothelial carcinoma.

SCLC remains one of the more aggressive cancers and has very poor outlook despite high initial response rates to chemotherapy. In previously treated patients, nivolumab demonstrated a response rate of 10% to 12% in initial studies, leading to emergency FDA approval of nivolumab as third-line therapy for SCLC. However, subsequent confirmatory studies failed to confirm these early promising findings, leading to withdrawal of FDA approval of nivolumab for SCLC. Mesothelioma is another treatment-refractory malignancy with poor long-term survival, particularly after failure of front-line platinum-based chemotherapy. Several small studies with pembrolizumab or nivolumab demonstrated responses in the range of 20% to 40% in patients progressing on chemotherapy.

Anti-Programmed Death-1/Programmed Death-L1 in Renal Cell Carcinoma

Along with melanoma, renal-cell carcinoma (RCC) has been classically considered among the most responsive tumors to immune therapy, including to IL-2 and interferon. Distinct from both melanoma and NSCLC, which both have a high tumor-mutation burden (a metric correlated with response to treatment, see Biomarkers section), RCC has a low/intermediate mutation burden. On the other hand, it is one of the more T-cell inflamed tumors (perhaps due to endogenous retroviral sequences which generate an immune response), which may partially explain the more immune responsive nature of this disease. Although responses to ipilimumab were infrequent, the phase I study of nivolumab demonstrated responses in 29% of previously treated RCC patients. This led to a phase III trial of nivolumab vs. everolimus in patients previously treated with anti-angiogenic therapy, which resulted in improved OS (median 25 months vs. 19.6 months, hazard ratio 0.73), with a response rate of 25%.[10] Although monotherapy remains an option, the major impact of immune therapy in RCC has been in combination regimens, both with anti-angiogenic agents and with combination ipilimumab and nivolumab (see Combination Regimens section).

Anti-Programmed Death-1/Programmed Death-L1 in Urothelial Carcinoma

Historically, urothelial carcinoma has been associated with dismal outcomes in patients who fail or are ineligible for platinum-doublet chemotherapy. Multiple randomized phase III studies, however, have now demonstrated improved OS in patients progressing on chemotherapy, showing approximately 3-month improvement in median OS and response rates in the range of 15% to 20%.[11] Five agents (nivolumab, pembrolizumab, atezolizumab, avelumab, and durvalumab) are all approved in this setting. Two phase III studies were also conducted in patients who were ineligible for first-line cisplatin-based chemotherapy. Notably, atezolizumab was not associated with improved OS compared with investigator's choice chemotherapy (although improvements were observed in PD-L1$^+$ tumors) (Table 80.2 for list of FDA approved anti-PD-L1 agents). Pembrolizumab was associated with improved OS in the front-line setting, however.

Anti-Programmed Death-1/Programmed Death-L1 in Skin Cancers

Although melanoma is the most frequently metastatic skin cancer requiring systemic therapy, Merkel-cell carcinoma (MCC), cutaneous-squamous-cell carcinoma (cuSCC), and basal-cell carcinoma may also present with locally advanced or metastatic disease. All of these tumors, with the exception of the MCCs, which are polyoma-virus associated, have extensive ultraviolet DNA damage signatures and high mutation burdens.

MCC is a neuroendocrine tumor with poor outcomes, particularly after failure of platinum-based chemotherapy. Pembrolizumab was tested as front-line therapy and was associated with a 56% response and nearly 50% 2-year PFS. Outcomes were similar in viral-positive vs. negative tumors. Avelumab (anti-PD-L1) was tested in previously treated patients and was associated with a 33% response rate; subsequent data showed a 62% response rate in untreated patients. Thus, these agents are now FDA approved and standard of care for advanced or unresectable MCC.

cuSCC is usually resectable but may occasionally present as metastatic or locally advanced disease. Cemiplimab (anti-PD-1) was associated with an approximately 50% response rate and is now an approved standard-of-care agent for this disease. Similarly, basal-cell carcinoma may require systemic therapy occasionally. Several case reports and small studies have reported responses in this disease, and phase II studies are ongoing.

Anti-PD-1/PD-L1 in GI Malignancies

Anti-PD-1 and anti-PD-L1 monotherapy has been trialed across GI malignancies, with relatively few successes.[12] The several instances where monotherapy has been effective include microsatellite-instability-high (MSI-H) colorectal adenocarcinoma, MSI-H refractory non-colorectal gastrointestinal malignancies, hepatocellular carcinoma (HCC), and gastric/gastroesophageal junction (GEJ) adenocarcinoma. In the nivolumab monotherapy cohort of the phase II Checkmate-142 trial, 74 later-line MSI-H colorectal cancer patients (54.1% with 3 or more prior lines of therapy) treated with the anti-PD-1 antibody experienced a 31.1% (20.8% to 42.9%) response rate. Median duration of response was not yet reached at the time of reporting, at which time all responding patients were alive. Median PFS and OS have not yet been reached for the cohort. In the phase II Keynote-164 trial, 63 MSI-H CRC patients (median two prior lines of treatment) were treated with pembrolizumab. The response rate was 32% with a median PFS of 4.1 months and a median OS that was not reached. Median duration of response was not reached with a duration of response of 6 months or more in 75% of responding patients.

In the landmark study[13] that led to pembrolizumab's tissue agnostic approval for refractory solid tumors, 53.4% of patients (46) had tumors of non-colorectal cancer origin. Of these patients, 58.6% (27) had GI origin primaries. Response rate in the non-colorectal cancer patients was 54% (39% to 69%). Median OS and PFS have not been reached in the study patients, with a 2-year OS rate of 64%.

Pembrolizumab garnered accelerated approval in third-line gastric/GEJ adenocarcinoma patients with tumor PD-L1 expression (CPS of 1 or greater) based on findings from the Keynote-059 study.[14] In the pembrolizumab monotherapy cohort of the phase II trial, 259 patients were treated with the anti-PD-1 antibody. Response rate in the entire cohort was 11.6% with a median duration of response of 8.4 months. Response rate in patients with PD-L1 positive tumors (CPS of 1 or greater) was 15.5% with a median response duration of 16.3 months. Specifically, in third-line PD-L1 positive patients, response rate was 22.7%.

The phase III Keynote-062 study explored pembrolizumab monotherapy versus platinum-doublet chemotherapy in treatment naive metastatic gastric/GEJ adenocarcinoma patients with PD-L1 positive (CPS of 1 or greater) tumors.[15] Patients treated with pembrolizumab demonstrated non-inferiority for OS when compared to patients treated with chemotherapy. In the MSI-H subgroup (33 patients), patients treated with pembrolizumab demonstrated markedly improved overall survival compared to patients treated with chemotherapy (12-month overall survival of 79% compared to 49%).

Nivolumab was tested in HCC patients (both with and without viral hepatitis) who had progressed on sorafenib or were intolerant of the agent in the Checkmate-040 study. In this

TABLE 80.2 FDA-Approved Indications for Anti-PD-L1 Antibody Monotherapy

Drug	Disease Indication	Phase of Study [NCT Number]	Response Rate [95% CI]	Median OS (mon)
Durvalumab	Unresectable stage III NSCLC, post-consolidation CCR	Phase III [NCT02125461]	30.0% [25.8%–34.5%]	28.3
Avelumab	Platinum-refractory UCC	Phase Ib [NCT01772004]	17% [11%–24%]	6.5
Durvalumab	Platinum-refractory UCC	Phase I/II [NCT01693562]	31% [17.6%–47.1%]	NR
Avelumab	Stage IV Merkel cell carcinoma	Phase II [NCT02155647]	62.1% [42.3%–79.3%]	NR

CCR, Concurrent chemoradiation; *mon*, months; *NSCLC*, non-small cell lung cancer; *UCC*, urothelial carcinoma.

multi-cohort phase I/II trial,[16] 262 patients with Child-Pugh A or Child-Pugh B7 liver function were treated with the anti-PD-1 drug. The response rate in treated patients was 20% (15% to 26%) and median duration of response was 9.9 months. The 9-month OS rate was 74% (67% to 79%). The results from this study granted nivolumab accelerated approval in the post-sorafenib setting in HCC patients.

Pembrolizumab was also tested in HCC patients in the single-arm phase II study Keynote-224. In this study, 104 patients with Child-Pugh A liver function received the anti-PD-1 agent. The response rate in treated patients was 17% and median OS was 12.9 months (9.7 to 15.5 months). The drug gained accelerated approval in second-line HCC patients based on findings from this study. In the confirmatory phase III study Keynote-240 of pembrolizumab versus best supportive care in HCC patients whose disease had progressed on sorafenib, pembrolizumab improved OS (HR 0.78; one-sided $P = 0.024$) and PFS (HR 0.78; one-sided $P = 0.021$) although neither of these endpoints met prespecified significance.[17]

Anti-Programmed Death-1/Programmed Death-L1 in Head and Neck Tumors

Nivolumab was initially tested in the second-line setting post-progression on platinum chemotherapy in recurrent squamous cell carcinoma of the head and neck (HNSCC).[18] Patients were randomized to nivolumab or physicians choice of chemotherapy (methotrexate, cetuximab, docetaxel). OS was longer in patients receiving the anti-PD-1 agent (HR 0.7, 97.7% CI 0.51 to 0.73; 1-year survival 36% vs. 16.6%). Nivolumab was FDA approved based on findings from the study. Pembrolizumab also obtained accelerated approval in recurrent head and neck SCC based on findings from the Keynote-012 study. In this phase 1b expansion cohort, 192 patients were treated with the anti-PD-1 agent. Of these patients 24% (47) received the checkpoint inhibitor in the second-line setting. In the entire cohort of treated patients, response rate was 18%. Similar response rates (17% and 15%) were reported between patients whose disease had progressed on prior platinum chemotherapy or patients whose disease had progressed on platinum chemotherapy plus cetuximab. Median duration of response was not reached and median OS was 8 months. The findings from this study were confirmed in the subsequent confirmatory phase III study Keynote-040. In this 495-patient study, first-line platinum progressive patients with HNSCC were randomized to pembrolizumab or standard-of-care chemotherapy. A statistically significant difference in median OS was observed in the immunotherapy treated patients (HR 0.8, 95% CI 0.65 to 0.98; 1-year survival 37% vs. 26.5%).

A portion of the Keynote-048 study tested pembrolizumab monotherapy versus cetuximab, platinum and fluorouracil triplet chemotherapy in the first-line setting for metastatic HNSCC. In patients with tumors with PD-L1 expression (CPS of 1 or greater) or PD-L1 high expression (CPS of 20 or greater), pembrolizumab was associated with a statistically significant improvement in OS (HR 0.78, 95% CI 0.64 to 0.96 [CPS $\geq$ 1] and HR 0.61, 95% CI, 0.45 to 0.83 [CPS $\geq$ 20]). These findings led to the FDA approval for pembrolizumab in first-line metastatic HNSCC patients with tumors expressing PD-L1.

Anti-Programmed Death-1/Programmed Death-L1 in Lymphoma

Nivolumab was tested in relapsed/refractory classical Hodgkin lymphoma (cHL) patients in the multi-cohort Checkmate-205 study.[19] Most patients had disease that had progressed after autologous stem cell transplant and brentuximab vedotin and had received a median of five prior lines of therapy. The response rate with nivolumab was 65% with a median duration of response of 8.7 months. Nivolumab was granted accelerated approval in patients with disease that had relapsed post-autologous transplant and post-transplant brentuximab vedotin. In the longer term follow up from the study, response rate was 65% to 73% in each cohort, median duration of response was 16.6 months, and median PFS was 14.7 months.

Pembrolizumab was trialed in relapsed/refractory cHL patients in a cohort of the phase 1b Keynote-013 trial. In this cohort, 31 cHL patients, the majority of whom had received prior autologous stem cell transplant (71%) and failed brentuximab vedotin, were treated with the anti-PD-1 agent. Response rate in treated patients was 65% with 16% of patients achieving complete responses. PFS rates at 1-year were 46%. Findings from this cohort led to the phase II Keynote-087 trial. This was a 210-patient study where most enrolled patients had progressed after autologous stem cell transplant (ASCT; 61.4%), brentuximab vedotin (71.4%), and received a median of four prior treatments. Response rate was 69%, median duration of response was not reached, and median OS was not reached in all treated patients. The 9-month OS in study patients was 96%. Pembrolizumab received accelerated approval in cHL patients who had disease that was refractory to three or more prior therapies.

Another cohort of the Keynote-13 trial explored pembrolizumab in non-Hodgkin lymphoma patients (primary mediastinal B-cell lymphoma [PMBCL] specifically). In this cohort, 18 patients received the anti-PD-1 agent. Of treated patients, 61% had received three or more lines of prior treatment and 33% had been pre-treated with ASCT. Response rate in treated patients was 41% and median duration of response was not reached. The subsequent Keynote-170 trial was a single arm phase II trial of relapsed refractory PMBCL patients. Among 53 patients whose data were reported, response rate was 45% (11% complete response) and median duration of response was not reached. Based on these data, the FDA granted accelerated approval to pembrolizumab in PMBCL patients with refractory disease or those with disease that had relapsed after two or more prior lines of therapy.

Anti-Programmed Death-1/Programmed Death-L1 in Cervical Carcinoma

Pembrolizumab was tested in patients with recurrent or metastatic cervical cancer in a cohort of the phase II basket Keynote-158 trial.[20] In this cohort, 98 patients were treated with the anti-PD-1 therapy; of these patients, 83.6% had PD-L1 positive tumors (CPS of 1 or greater) and 65.3% were treated with two or more lines of prior therapy. Across all patients, the response rate was 12.2% (6.5% to 20.4%). However, no responses were seen in patients without PD-L1 expression. Median duration of response was not reached in treated patients. Median OS was 9.4 months (7.7 to 13.1 months) and 11 months (9.1 to 14.1 months) in the entire population and patients with PD-L1 positive tumors, respectively. Based on data from this study, the FDA granted accelerated approval to pembrolizumab in patients with PD-L1-positive cervical cancer that had progressed on chemotherapy. Although nivolumab has also demonstrated anti-tumor efficacy against progressive cervical cancer in the Checkmate-358 trial, the FDA has not yet provided a decision with regard to approval for the agent in this disease space.

KEY CONCEPTS

Box C: Combination Checkpoint Blockade

- Combined PD-1/CTLA-4 blockade has demonstrated additive or synergistic activity in several cancer types, although at a cost of increased toxicity
- Anti-PD-1 plus chemotherapy has demonstrated improved survival in several cancers, including NSCLC, HNSCC, and TNBC, although the durability of the responses is unclear at this time.
- Kinase inhibitors, such as VEGF or BRAF/MEK inhibitors, also appear to provide improved efficacy in combination with anti-PD-1 antibodies.
- Numerous clinical trials testing a wide variety of combinations are underway in this rapidly evolving field.

Immunotherapy Combinations

Despite the success of anti-CTLA-4 and anti-PD-1/PD-L1 monotherapies, most cancer patients still fail to benefit from treatment. One potential method to extend the benefit of immune checkpoint inhibitors (ICIs) is to combine them with other immune, targeted chemotherapies.

The development of combination regimens that include ICI has rapidly expanded, and a comprehensive characterization of each regimen is beyond the scope of this chapter. Thus, we will primarily focus on approved regimens and highlight a few regimens that may be cornerstones of therapy in the future (Table 80.3 for list of approved combination therapies).

Immune-Immune Combinations

The most obvious combination regimen was the first approved: combined PD-1 and CTLA-4 blockade. Pre-clinical and mechanistic data suggested that the immune activation induced by each agent was non-overlapping and potentially synergistic. Further, patients who failed one class of therapy often responded to the other class following progression. Thus, after initial promising responses were observed in metastatic melanoma in phase I/II studies, a phase III study comparing nivolumab, ipilimumab, and the combination was conducted in metastatic melanoma patients. Study metrics were superior in both nivolumab containing arms, particularly in the combination for response rate (58%, 45%, and 19% for nivolumab/ipilimumab, nivolumab, and ipilimumab, respectively), PFS (median 11.5, 6.9, and 2.9 months, respectively), and OS (median not reached at 60 months, 36.9 months, and 19.9 months). The difference in OS between the combination and nivolumab was not statistically significant, although of potential clinical significance (HR 0.83, 95% CI 0.67 to 1.03; 5-year survival 52% vs. 44%), and the trial was not powered to find a difference in these arms. The increased efficacy was accompanied by increased toxicity (59% vs. 23% for high-grade toxicities).[6]

Ipilimumab and nivolumab have also improved demonstrated efficacy in other cancers. Among patients with intermediate or poor risk metastatic RCC, ipilimumab and nivolumab was superior to sunitinib in terms of OS (median not reached vs. 26.6 months, HR 0.66) and response rate (42% vs. 29%). Thus, ipilimumab and nivolumab is now a standard of care for frontline, metastatic RCC.[21]

Similarly, in NSCLC, ipilimumab and nivolumab improved outcomes for patients compared with cytotoxic chemotherapy. In untreated patients, ipilimumab and nivolumab was superior to platinum-based chemotherapy in both PD-L1$^+$ patients (median OS 17.1 months vs. 14.9 months) with a more pronounced improvement in PD-L1$^-$ patients (median OS 17.2 months vs. 12.2 months).[22] Thus, one emerging paradigm in chemotherapy-free regimens is to treat patients with PD-L1$^+$ tumors with pembrolizumab monotherapy, and patients with PD-L1$^-$ with ipilimumab and nivolumab.

TABLE 80.3 FDA-Approved Indications for Immune Checkpoint Inhibitor Combinations in Metastatic/Unresectable Malignancies

ICI Drug	Partner Drug(s)	Disease Indication	Phase of Study [NCT Number]	Response Rate [95% CI]	Median OS (Mon)
Nivolumab	Ipilimumab	First-Line Melanoma	III [NCT01844505]	57.6% [52.0%–63.2%]	NR
Nivolumab	Ipilimumab	Later-Line MSI-H CRC	II [NCT02060188]	58% [49%–67%]	NR
Nivolumab	Ipilimumab	First-Line Poor to Intermediate-Risk RCC	III [NCT02231749]	42% [37%–47%]	NR
Atezolizumab	Carboplatin and Etoposide	First-Line Extensive-Stage SCLC	III [NCT02763579]	60.2% [53.1%–67%]	12.3
Pembrolizumab	Axitinib	First-Line RCC	III [NCT02853331]	59.3% [54.5%–63.9%]	NR
Avelumab	Axitinib	First-Line RCC	III [NCT02684006]	51.4% [46.6%–56.1%]	NR
Pembrolizumab	Carboplatin and Pemetrexed	First-Line NSCLC Adenocarcinoma	III [NCT02578680]	47.6% [42.6%–52.5%]	NR
Pembrolizumab	Carboplatin and Nab-Paclitaxel/Paclitaxel	First-Line NSCLC Squamous	III [NCT02775435]	57.9% [51.9%–63.8%]	15.9
Atezolizumab	Carboplatin and Nab-Paclitaxel	First-Line NSCLC Adenocarcinoma	III [NCT02367781]	49·2% [44.5%–54·0%]	18.6
Atezolizumab	Carboplatin, Paclitaxel and Bevacizumab	First-Line NSCLC Adenocarcinoma	III [NCT02366143]	63.5% [58.2%–68.5%]	19.2
Atezolizumab	Nab-Paclitaxel	First-Line PD-L1 (≥ 1%) TNBC	III [NCT02425891]	NA	25
Pembrolizumab	Platinum and 5-FU	First-Line Head and Neck SCC	III [NCT02358031]	36% [NA]	13
Pembrolizumab	Lenvatinib	Second-Line Endometrial Cancer	II [NCT02501096]	39.6% [26.5%–54.0%]	NA

CRC, Colorectal cancer; *ICI*, immune-checkpoint inhibitor; *mon*, months; *MSI-H*, microsatellite instability-high; *NA*, not available; *NR*, not reached; *NSCLC*, non-small cell lung cancer; *RCC*, renal cell carcinoma; *SCLC*, small cell lung cancer; *TNBC*, triple-negative breast cancer.

Other cancers have also demonstrated benefit from this combination, including MSI-H colorectal cancer (response rate 55%; 24-month OS rate 74%), SCLC (response rate 22%, compared with 11.6% with nivolumab), and urothelial bladder cancer (26% to 38%). Of note, the dose of ipilimumab appears important in determining the rate of high-grade toxicities. In renal, colorectal, and lung cancer studies, ipilimumab is usually given at 1 mg/kg, compared with the 3 mg/kg dose in melanoma studies. The lower dose seems to result in 10% to 20% fewer patients experiencing high-grade toxicities (approximately 30% vs. 40% to 50%), although with both regimens approximately 1% of patients experience fatal toxicities.[23]

Other immune/immune combinations are being developed. A host of other alternative immune checkpoints, such as LAG-3, TIM-3, VISTA, B7-H3, and others have been discovered to have broadly similar roles as PD-1 or PD-L1 in suppressing T-cell function.[24] Early data for PD-1/LAG-3 blockade have shown promise in metastatic melanoma, including in anti-PD-1 progressors. T-cell co-stimulators also are thought to have great promise, as "the other side of the coin" from immune checkpoints. Inhibitors of IDO, which has many common properties to the immune checkpoints, have also been tested in combination with PD-1 blockade. Despite the anticipated promise, a phase III study of epicadostat plus pembrolizumab versus pembrolizumab alone in metastatic melanoma was negative for all prespecified endpoints. Agonist antibodies against OX40, 4-1BB activate T cells in preclinical data (both *in vitro* and *in vivo*) and produce antitumor responses. However, these agents seem to lack efficacy in early clinical trials.

Immune Checkpoint/Cytokine Combinations

The relative efficacy of IL-2 has spurred attempts to produce less toxic reformulations. Bempegaldesleukin is a pegylated recombinant IL-2 that has shown preliminary signs of efficacy with manageable toxicity (minimal hypotension and cytokine release) in combination with nivolumab in melanoma, RCC, NSCLC, and urothelial cancer. Larger studies are currently ongoing.[25] Pegilodecakin (pegylated IL-10) promoted proliferation, expansion, and reversal of exhaustion of CD8 T cells in preclinical models. Early data have suggested that pegilodecakin, in combination with pembrolizumab or nivolumab, has activity in NSCLC and RCC (greater than 40% response rates in approximately 30 patients each). This agent is associated with cytopenias rather than classical immune toxicities.[26]

Immune Checkpoint/Intratumoral Injection Combinations

Intratumoral injections have a long history in immune-oncology, dating back to Coley toxins.[27] Melanoma, with frequent cutaneous, subcutaneous, and palpable lymph node metastases, has been a particularly fertile ground for testing these agents and led to the first approved oncolytic virus. Talimogene laherparepvec (T-VEC) is a modified herpes simplex virus type one with neurovirulence components removed and the GM-CSF promoter inserted, which leads to durable responses in 15% to 25% of melanoma patients. Early phase studies have showed excellent response rates in combination with either pembrolizumab (approximately 60%) or ipilimumab (approximately 40%); a phase III study of pembrolizumab plus T-VEC versus pembrolizumab alone has completed accrual. Other oncolytic viruses are currently being developed with several early ongoing studies involving combinations of these agents with anti-PD-1/PD-L1 antibodies.

Toll-like receptor agonists also may synergize with immune checkpoint inhibitors by activating innate immunity, specifically through stimulating type 1 interferons and promoting dendritic cell maturity and activation. Early data combining TLR7/8 agonists with pembrolizumab have demonstrated response rates as high as 60% in phase II studies. Early data combining ipilimumab with a TLR9 agonist (tilsotolimod) have shown 30% to 40% response rates in patients refractory to anti-PD-1 antibodies.

Immune Checkpoint/Chemotherapy Combinations

On the surface, the idea of combining chemotherapy (and its frequent attendant steroid premedications) with ICI appears counterproductive. Specifically, one would suppose that myelosuppressive and lympholytic agents would counteract T-cell activation/reinvigoration mediated by checkpoint inhibitors. Nonetheless, there was some preclinical rationale for this approach. First, cell-death mediated by chemotherapy could augment an antitumor immune response by releasing additional immunogenic peptides or neoantigens, and by inducing immunogenic cell death (with resulting T cells and additional immune cell recruitment). Second, chemotherapy depletes protumor immune-cell populations, including myeloid-derived suppressor cells and regulatory T cells, thus potentially tipping the balance toward antitumor immunity.

Initial phase II studies in NSCLC demonstrated superior response rates for pembrolizumab and carboplatin plus pemetrexed compared with chemotherapy alone (55% vs. 29%) with only modestly increased toxicities.[28] This led to two phase III studies; pembrolizumab plus chemotherapy improved outcomes in both squamous and non-squamous NSCLC irrespective of PD-L1 status compared with chemotherapy alone. In squamous NSCLC (using carboplatin plus paclitaxel as a backbone), the triplet was associated with improved overall survival (median 15.9 months vs. 11.3 months, HR 0.64), PFS (median 6.4 months vs. 4.8 months, HR 0.56), and response rate (57.9% vs. 38.4%). In non-squamous NSCLC (using pemetrexed and platinum-based drug as chemotherapy backbone), the combination improved outcomes compared with chemotherapy for OS (12-month survival 69.2% vs. 49.4%, HR 0.49), PFS (median 8.8 months vs. 4.9 months, HR 0.52), and response rate (47.6% vs. 18.9%). In both trials, toxicities were only minimally increased with the combination. Atezolizumab, in combination with carboplatin and nab-paclitaxel, was also associated with superior OS (median 18.6 months vs. 13.9 months, HR 0.79) and PFS (median 7 months vs. 5.5 months, HR 0.64) compared with chemotherapy alone. A quadruplet combination (atezolizumab, bevacizumab, carboplatin, and paclitaxel) was also superior to bevacizumab and chemotherapy (median OS 19.2 months vs. 14.7 months, HR 0.78; median PFS 8.3 months vs. 6.8 months, HR 0.62). In addition, in SCLC, atezolizumab plus chemotherapy improved survival (median 12.3 months vs. 10.3 months, HR 0.7) and PFS (median 5.2 months vs. 4.3 months, HR 0.77) compared with chemotherapy alone. Similarly, durvalumab plus chemotherapy was associated with improved OS and PFS compared with chemotherapy alone in SCLC.[29]

Other cancers have demonstrated mixed outcomes when immune checkpoint inhibitors have been combined with chemotherapy. In untreated, metastatic, PD-L1$^+$, triple-negative breast cancer, atezolizumab plus nab-paclitaxel improved PFS

(median 7.5 months vs. 5 months, HR 0.62), and OS (median 25 months vs. 15.5 months, HR 0.62) compared with nab-paclitaxel alone in a phase III trial. In the intention-to-treat analysis (not restricted to PD-L1$^+$ tumors), there was also a trend toward improved survival (median OS 21.3 months vs. 17.6 months, HR 0.84, $P = 0.08$).[30] In another portion of previously described Keynote-048 trial, pembrolizumab plus platinum doublet chemotherapy was compared against cetuximab, cisplatin, and fluorouracil in patients with SCC of the head and neck. In the total population, patients treated with pembrolizumab plus platinum chemotherapy demonstrated improved overall survival compared to patients treated with chemotherapy plus cetuximab (HR 0.77, 95% CI 0.63 to 0.93), in addition to demonstrating improved overall survival in patients with CPS of 1 or greater and CPS of 20 or greater tumors.

The superiority analysis of the Keynote-062 trial, where treatment-naïve gastric/GEJ adenocarcinoma patients with tumor CPS of 1 or greater were treated with either pembrolizumab plus platinum doublet chemotherapy or platinum doublet chemotherapy, failed to demonstrate any survival difference between the two arms (HR 0.85, 95% CI 0.7 to 1.03). No OS difference was seen between the patient groups even in patients with tumors with CPS of 10 or greater (HR 0.85, 0.62 to 1.17). Furthermore, the grade 3/4 treatment-related toxicities were significantly greater in the combination arm than in the chemotherapy arm. Additional trials are ongoing in multiple cancer types with anti-PD-1/PD-L1 plus chemotherapy.

Immune Checkpoint/Targeted Therapy Combinations

Although "targeted therapy" is a broad category that encompasses widely divergent treatments, we will focus here on antibodies or small molecules targeting mutated proteins or cell-surface molecules harbored on or within the cancer cell. The rationale for this approach is similar to that of chemotherapy, in that an effective targeted therapy may induce cell death and potentially immune infiltration, thus potentially promoting an antitumor immune response.

The first approved immune/targeted combinations are in RCC.[31] After early data suggested that axitinib and anti-PD-1/PD-L1 treatments were tolerable and had high response rates, a phase III trial comparing pembrolizumab and axitinib with sunitinib was conducted in patients with untreated metastatic RCC. The combination arm was associated with improved survival (12-month OS 89.9% vs. 78.3%, HR 0.53) and PFS (median 15.1 months vs. 11.1 months, HR 0.69). Toxicities were slightly augmented (high-grade 75.8% vs. 70.6%), although only approximately 10% of patients had to discontinue both drugs. Similarly, avelumab and axitinib were associated with improved PFS compared with sunitinib, although no OS benefit was observed. Both combinations are now approved and are commonly used options for metastatic RCC.

Among the first attempts to combine immune and targeted agents was a small study of ipilimumab and vemurafenib for BRAF mutant melanoma.[32] This combination was associated with unacceptable hepatotoxicity. While dabrafenib and ipilimumab was more tolerable, a triplet combination of ipilimumab, dabrafenib, and trametinib was associated with several colon perforations among the first few patients treated. Several studies have now tested combinations of BRAF and MEK inhibitors and anti-PD-1 or PD-L1 agents. These studies have suggested at least additive efficacy with manageable toxicity and have led to several phase III studies of BRAF/MEK/PD-1 or PD-L1 inhibition.

Endometrial cancer has been associated with responses to immune-checkpoint inhibitors, especially in the subset of patients with MSI-H tumors. However, a recent phase II study of pembrolizumab plus lenvatinib demonstrated response rates in 38% of MSI-low patients, leading to approval in metastatic endometrial cancer with at least one prior therapy. In HCC, the anti-PD-L1 agent atezolizumab has been combined with bevacizumab in treatment-naïve patients with advanced disease. In an initial phase 1b study, 26 patients were treated with the study combination. Among 21 patients evaluable for response, the response rate was 63% (all partial responses). These findings led to the first-line randomized phase III IMbrave150 trial where patients received atezolizumab plus bevacizumab or sorafenib. A recent press release suggests the combination arm demonstrated improved PFS and OS compared to sorafenib. Although the final data are pending, this will likely establish atezolizumab plus bevacizumab as the new first-line standard-of-care in advanced HCC patients. Many other immune/targeted combinations are currently being evaluated in clinical trials, including combinations with PARP inhibitors, EGFR inhibitors, and many others.

Biomarkers of Response

ICI are not effective in all cancer patients with advanced disease. A significant need exists to better identify the patients who are more likely to respond to these agents. Thus far, efforts to identify biomarkers of response have focused primarily on tissue-based biomarkers such as PD-L1, MSI-H, and tumor mutational burden (TMB) (Fig. 80.4).

Emerging areas of interest include cell-free DNA-based biomarkers and on-treatment biomarkers including immune-related adverse events (IRAEs).

Programmed Death-L1 as a Biomarker

Across current approvals for ICI, PD-L1 expression has been found to be predictive of response in only 28.9% percent of cases.[33] Among the other cases, PD-L1 expression has largely been found not to be predictive. Further complicating the issue is that different PD-L1 thresholds, PD-L1 assays and cells upon which PD-L1 is measured have been utilized in different disease sites. In gastric/GEJ adenocarcinoma and cervical cancer patients with metastatic disease, PD-L1 expression threshold of 1% or greater on tumor cells plus immune cells (by IHC 22C3) is sufficient for patients to be treated with pembrolizumab. In patients with treatment naïve metastatic NSCLC, a PD-L1 expression threshold of 50% or greater on tumor cells (by IHC

KEY CONCEPTS

Box D: Biomarkers of Response and Toxicity

- Predicting which patients will respond to treatment remains a key challenge.
- Certain clinical parameters correlate with response, including low disease burden, lung and lymph node (vs. liver and brain) involvement, and good performance status.
- Molecular correlates of response include PD-L1 expression on tumor cells or infiltrating macrophages, high tumor mutation burden, tumor cell expression of MHC class II, and T-cell infiltration
- While these factors may be clinically useful and may help with treatment prioritization in some circumstances, more reliable biomarkers are needed to improve prediction capacity.

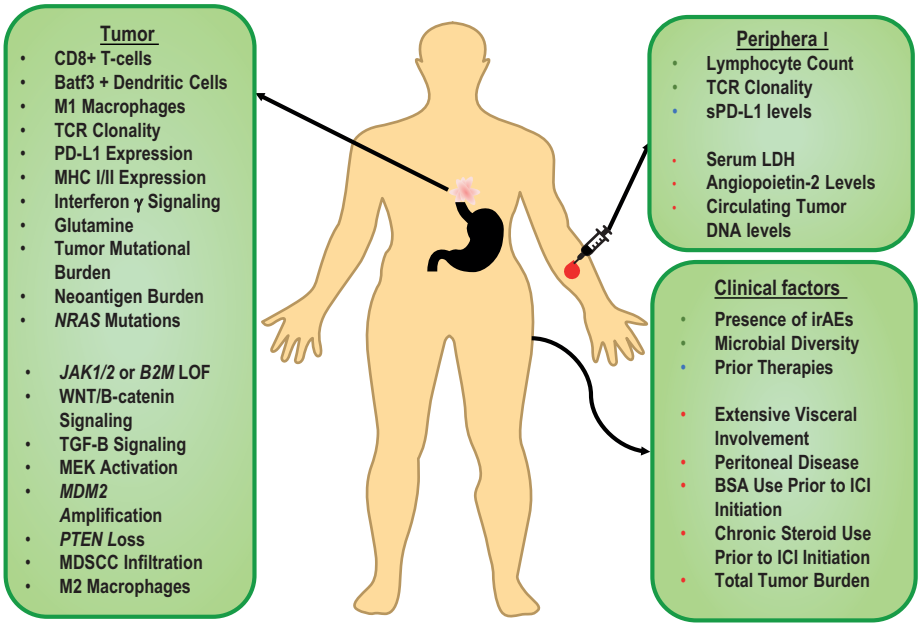


FIG. 80.4 Candidate biomarkers for response to immune-checkpoint inhibitors. *ICI*, Immune-checkpoint inhibitor; *irAEs*, immune-related adverse events; *MHC*, major-histocompatibility complex; *TCR*, T-cell receptor; *TGF-B*, transcription factor beta; *MDSCC*, myeloid-derived suppressor cell; *LDH*, lactate dehydrogenase; *BSA*, body surface area.

22C3) has been identified for the cutoff for first-line pembrolizumab therapy. In patients with cisplatin-ineligible bladder cancer, a PD-L1 expression threshold of 5% or greater on immune cells (by SP142) was used to garner approval for atezolizumab. The inter-assay and inter-disease variability make PD-L1 an inconsistent biomarker.

Microsatellite-Instability as a Biomarker

MSI-H status is perhaps the most predictive existing biomarker for ICI response. MSI-H tumors display aberrant mismatch repair, which results in spontaneous mutations; the propagation of these mutations during DNA replication leads to subsequent expression of numerous neoantigens. The volume of neoantigen expression of these tumors allows for more ready immune recognition. In the seminal publications that garnered pembrolizumab a tissue-agnostic approval across refractory advanced malignancies, the response rate in MSI-H non-colorectal cancer patients was 34.3% with a median OS of 23.5 months (95% CI, 2.4 to 4.9 months). In MSI-H colorectal cancer, ORR was 32% with a median OS not reached. Response rates with combination PD-1 and CTLA-4 blockade are even higher than with anti-PD-1 inhibitors in MSI-H colorectal cancer, although it remains to be seen whether mature OS data will also be different.[12,13]

Tumor Mutational Burden as a Biomarker

TMB is a proxy for neoantigen expression and the principle for elevated TMB predicting response to immune-checkpoint inhibitors mirrors the mechanism by which MSI-H status is thought to influence immune-therapy response. In the original study highlighting the pooled benefit of pembrolizumab in refractory MSI-H tumors in 2015, TMB was significantly associated with progression-free survival and non-statistically associated with overall survival. In a 2017 NEJM Letter to the Editor, investigators examined the relationship between TMB and response rate to anti-PD-1 or anti-PD-L1 agents in studies in 27 tumor types and found a correlation coefficient of 0.74 ($P < 0.001$).[34] Although data suggest that TMB is associated with anti-PD-1 or anti-PD-L1 response, defining an appropriate TMB cut-off and consistent means of measurement has been more problematic. In different tumor types, different TMB thresholds suggest successful prediction of checkpoint-inhibitor response. In NSCLC, this threshold has been thought to be 10 mutations/MB whereas in MSI-H colorectal cancer, this threshold may be 37 mutations/MB. Other publications suggest rather than looking at TMB by a single value across tumor types, using the top 20% of TMB values for individual cancer types can perhaps serve as a better threshold for predicting response.[35] TMB can be quantified through next-generation sequencing (NGS) based panels or through whole exome sequencing, though in clinical practice, NGS is more readily utilized. A limitation between NGS assays is the variation in number of genes sequenced by each assay and the lack of a standard conversion of TMB estimates between them. Thus, although higher TMB is suggestive of immune-checkpoint inhibitor responsiveness, the optimal means of measuring and standardizing TMB remain to be determined.[36]

Other Biomarkers

Other less prevalent biomarkers of immune-checkpoint inhibitor response include mutations in the genes encoding DNA polymerases epsilon (*POLE*) and delta 1 (*POLD1*). The polymerases play essential roles in proofreading during DNA replication, and mutations in them can lead to a hypermutated molecular phenotype. In an analysis of outcomes in patients treated with immune-checkpoint inhibitors, patients with *POLE/POLD1* mutations demonstrated significantly improved OS compared to patients without the genotype (median OS 34 months vs. 18 months, $P = 0.004$).[37] On multivariate regression, when tumor histology and MSI status were adjusted, *POLE/POLD1* mutations independently identified patients who benefit from checkpoint inhibitors.

Tumor-infiltrating lymphocytes have also been explored as a possible predictive biomarker for immune-checkpoint inhibitor response. In melanoma, an immune cell infiltrate score

(Immunoscore) from primary tumor specimens of patients who later developed metastatic disease and were treated with either anti-PD-1 or anti-CTLA-4 antibodies was assessed with checkpoint inhibitor treatment response. Among 22 patients who had primary resection specimens available, 59% had Immunoscore high tumors, while 41% had Immunoscore low tumors. Patients with Immunoscore high tumors exhibited a 5-year OS of 59.8% compared to 11.1% in patients with Immunoscore low tumors ($P = 0.024$). The predictive potential of tumor-infiltrating lymphocytes continues to be explored across other disease sites.

Toxicities from Immune-Checkpoint Inhibitors

Immune checkpoints, including PD-1, PD-L1, and CTLA-4, are key regulators of immune tolerance in the physiologic state, and thus restrain autoreactive T cells to prevent autoimmunity. When these molecules are inhibited, self-reactive T cells may be unleashed and produce inflammation. Essentially any organ system can be affected, but skin, colon, lungs, liver, and thyroid are most commonly involved. It remains unknown why particular toxicities affect individual patients. Potential triggers include: (1) shared antigens between tumor and host tissue (or perhaps mutations generating epitopes with high degrees of homology to host epitopes), (2) smoldering inflammation present in a particular organ prior to treatment (due to viral or autoimmune causes), (3) cytokine imbalances, (4) environmental exposures triggering inflammation, and (5) antibody deposition on particular organs (as has been suggested with anti-CTLA-4 hypophysitis).

The incidence of severe adverse events is approximately 20% with anti-PD-1/PD-L1 monotherapy and 50% with combination PD-1/CTLA-4 blockade; lower-grade events occur in additional patients.

Management consists of three components. First, ICIs should be discontinued (either permanently or temporarily). However, the lengthy half-lives, and still longer pharmacodynamic effects of checkpoint inhibitors, will continue to provoke further inflammation. Thus, additional immune suppression is needed for severe events. Second, glucocorticoid treatment (prednisone 1 to 2 mg/kg or equivalent) will mitigate lymphocyte-mediated inflammation. Thus far, studies have suggested that high-dose steroids do not compromise cancer-specific outcomes; however, conclusively demonstrating any effect on the cancer is challenging. One study did suggest that low-dose steroids were associated with better outcomes in patients with hypophysitis, a condition that generally does not require high-dose steroids. Toxicities that do not respond to steroids may improve with other immunosuppressants, including infliximab for colitis and mycophenolate mofetil for hepatitis. Third, disease-specific supportive management, including electrolyte and fluid replacement for colitis, oxygen supplementation for pneumonitis, and hormone replacement for endocrinopathies, may be necessary.[38] Although toxicities are usually manageable with these algorithms, between 0.35% (with anti-PD-1/PD-L1 monotherapy) and 1.2% (with combination therapy) of treated patients experience fatal toxicities.[23] While some fatalities likely stem from delays in obtaining optimal care, others are from fulminant events refractory to immunosuppression. In addition, patients with skin, joint, nerve, endocrine, and lung effects may ultimately evolve to more chronic toxicities. Most events (other than endocrinopathies), however, resolve with appropriate management. Although a full review of immune-mediated toxicities is beyond the scope of this chapter, we will review four clinically relevant toxicities that offer a window into the nature of autoimmunity generated by these agents.

KEY CONCEPTS

Box E: Toxicities of Immune-Checkpoint Inhibitors

- Clinically severe immune-related toxicities occur in 10%–20% of patients receiving anti-PD-1, and up to 50% of those with combined PD-1/CTLA-4 blockade
- These autoimmune-like phenomena may affect any organ, most commonly the skin, thyroid, colon, lungs, liver, and joints.
- Rarer events may involve the heart, brain, peripheral nervous system, and hematologic system, and may cause morbidity and mortality.
- The treatment of severe toxicities involves treatment cessation, high-dose corticosteroids, and supportive management.

Colitis

The most common serious toxicity, particularly among anti-CTLA-4 treated patients, is colitis. Although some similarities to inflammatory bowel disease exist, this entity tends to arise after several doses of therapy and is characterized by diarrhea, with much less frequent bloody stools or abdominal pain. Biopsies demonstrate infiltrating lymphocytes and neutrophils. Specific strains of microbial flora and IL-6 imbalance have been correlated with onset of colitis but have not been confirmed in larger series. Untreated colitis may lead to perforation, but this complication is usually steroid responsive. Infliximab, vedolizumab, or event fecal transplant may be effective in steroid-refractory settings.[39]

Myocarditis

Although uncommon (up to 1% to 3% with combination ipilimumab/nivolumab), clinically severe myocarditis may have fatality rates as high as 50% (Fig. 80.5). This toxicity tends to occur in the first month of therapy, frequently co-occurs with skeletal muscle inflammation (myositis) and a myasthenia-gravis-like condition, and presents with often fulminant arrhythmias.[40] The early onset and severe nature suggest that pre-existing cardiac inflammation may be the cause of this entity. Indeed, studies have suggested that shared high-frequency T-cell clones are common to both heart, skeletal muscle, and tumor. High-dose steroids may be effective but worsening and death are still common even with prompt treatment. Case reports of resolution with abatacept (CTLA-4 fusion protein) and alemtuzumab (anti-CD52) have been published, thus suggesting potentially effective rescue approaches.

Pneumonitis

Lung inflammation occurs in 3% to 5% of patients treated with anti-PD-1 agents, although subsequent series suggest up to 20% of lung cancer patients may experience pneumonitis.[41] Various radiographic appearances are possible, including interstitial thickening, groundglass opacities, and organizing pneumonia (see Fig. 80.5). Bronchoscopic evaluation reveals CD4 T cells with central memory differentiation and type 1 polarization and increased expression of IL-1B. Although pneumonitis typically improves with steroids, refractory cases causing respiratory failure do occur.

Hypophysitis

Inflammation of the pituitary gland, or hypophysitis, is a medical condition almost exclusively linked with anti-CTLA-4 therapy, which occurs in up to 10% of patients receiving combination PD-1/CTLA-4 blockade. Preclinical and correlative studies have suggested that CTLA-4 is expressed in the pituitary, particularly

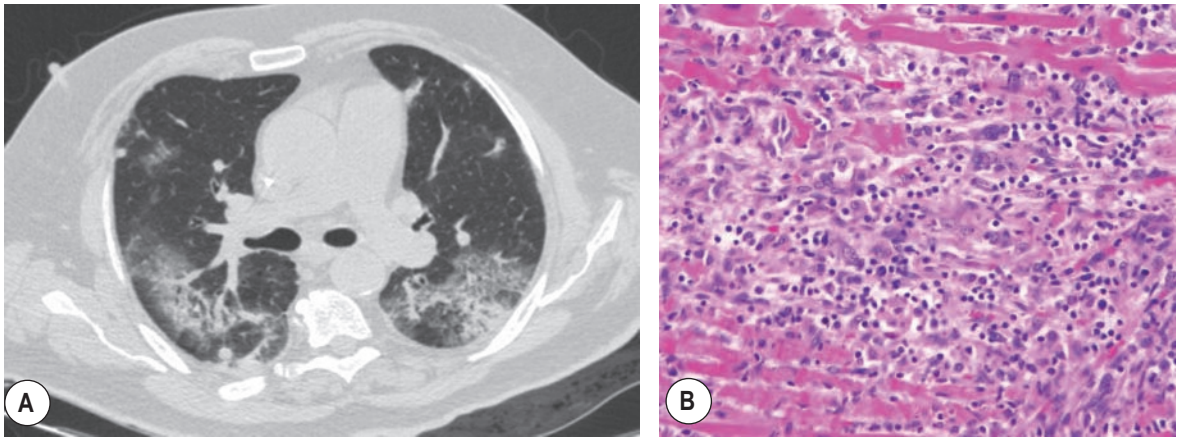


FIG. 80.5 Cross-sectional imaging showing interstitial opacities in a patient with anti-PD-1 associated pneumonitis (A). Myocardial biopsy showing extensive immune cell infiltration into myocardium in a patient with ICI-associated myocarditis (B).

on thyrotropin- and prolactin-secreting cells. With therapy, the drug is deposited on these cells, thus inducing ADCC and complement deposition similar to a type II hypersensitivity. As with other endocrine toxicities, patients may continue ICI therapy uninterrupted since the toxicity is largely irreversible due to the inflammation-induced "burnout" of all hormone-producing cells. Patients usually require lifelong glucocorticoid replacement and usually thyroid replacement.[39]

Specific Safety Considerations

Interestingly, several populations were excluded from clinical trials due to concern for either heightened immune activation and toxicity (e.g., those with pre-existing autoimmune disease or organ transplant), decreased immune function (HIV/AIDS, chronic steroid use, poor performance status), or difficulty identifying toxicities (e.g., pre-existing organ dysfunction).[42] Patients with pre-existing autoimmune disease seem to have somewhat higher risk of autoimmune adverse events, but these are manageable and associated with similar anticancer response rates as unselected patients. Patients with HIV/AIDS seem to have similar toxicity and response compared with unselected patients as well. Those with pre-existing steroid use or poor performance status have low response rates to treatment, which may reflect compromised immune function. However, these patients may respond to treatment and should not be excluded from treatment without a thoughtful consideration of risks and benefits. One near absolute contraindication to therapy is a pre-existing solid organ transplant; patients have what appears to be greater than 50% risk of graft failure with anti-PD-1 treatment. Pregnancy is also considered a very high-risk consideration, as PD-L1 is highly expressed on the placenta, and likely mediates immune exclusion from the developing fetus. Somewhat surprisingly, several case reports of healthy full-term births with excellent maternal and fetal outcomes have been reported in patients receiving anti-PD-1 treatment during pregnancy.

REFERENCES

1. Wolchok JD, Saenger Y. The mechanism of anti-CTLA-4 activity and the negative regulation of T-cell activation. *oncologist*. 2008;13(Suppl 4):2–9.
2. Callahan MK, Wolchok JD. Clinical Activity, Toxicity, Biomarkers, and Future Development of CTLA-4 Checkpoint Antagonists. *Semin Oncol*. 2015;42(4):573–586.
3. Schadendorf D, Hodi FS, Robert C, Weber JS, Margolin K, Hamid O, et al. Pooled Analysis of Long-Term Survival Data From Phase II and Phase III Trials of Ipilimumab in Unresectable or Metastatic Melanoma. *J Clin oncology: Off J Am Soc Clin Oncol*. 2015;33(17):1889–1894.
4. Wolchok JD, Hoos A, O'Day S, Weber JS, Hamid O, Lebbe C, et al. Guidelines for the evaluation of immune therapy activity in solid tumors: immune-related response criteria. *Clin cancer research: an official J Am Assoc Cancer Res*. 2009;15(23):7412–7420.
5. Ribas A, Wolchok JD. Cancer immunotherapy using checkpoint blockade. *Science*. 2018;359(6382):1350–1355.
6. Larkin J, Chiarion-Sileni V, Gonzalez R, Grob JJ, Rutkowski P, Lao CD, et al. Five-Year Survival with Combined Nivolumab and Ipilimumab in Advanced Melanoma. *N Engl J Med*. 2019;381(16):1535–1546.
7. Spagnolo F, Boutros A, Tanda E, Queirolo P. The adjuvant treatment revolution for high-risk melanoma patients. *Semin Cancer Biol*. 2019.
8. Mok TSK, Wu YL, Kudaba I, Kowalski DM, Cho BC, Turna HZ, et al. Pembrolizumab versus chemotherapy for previously untreated, PD-L1-expressing, locally advanced or metastatic non-small-cell lung cancer (KEYNOTE-042): a randomised, open-label, controlled, phase 3 trial. *Lancet*. 2019;393(10183):1819–1830.
9. Antonia SJ, Villegas A, Daniel D, Vicente D, Murakami S, Hui R, et al. Overall Survival with Durvalumab after Chemoradiotherapy in Stage III NSCLC. *N Engl J Med*. 2018;379(24):2342–2350.
10. Beckermann KE, Johnson DB, Sosman JA. PD-1/PD-L1 blockade in renal cell cancer. *Expert Rev Clin Immunol*. 2017;13(1):77–84.
11. Siefker-Radtke A, Curti B. Immunotherapy in metastatic urothelial carcinoma: focus on immune checkpoint inhibition. *Nat Rev Urol*. 2018;15(2):112–124.
12. Ganesh K, Stadler ZK, Cercek A, Mendelsohn RB, Shia J, Segal NH, et al. Immunotherapy in colorectal cancer: rationale, challenges and potential. *Nat Rev Gastroenterol Hepatol*. 2019;16(6):361–375.
13. Le DT, Durham JN, Smith KN, Wang H, Bartlett BR, Aulakh LK, et al. Mismatch repair deficiency predicts response of solid tumors to PD-1 blockade. *Science.*. 2017;357(6349):409–413.
14. Fuchs CS, Doi T, Jang RW, Muro K, Satoh T, Machado M, et al. Safety and Efficacy of Pembrolizumab Monotherapy in Patients With Previously Treated Advanced Gastric and Gastroesophageal Junction Cancer: Phase 2 Clinical KEYNOTE-059 Trial. *JAMA Oncol*. 2018;4(5):e180013.
15. Tabernero J, Cutsem EV, Bang Y-J, Fuchs CS, Wyrwicz L, Lee KW, et al. Pembrolizumab with or without chemotherapy versus chemotherapy for advanced gastric or gastroesophageal junction (G/GEJ) adenocarcinoma: The phase III KEYNOTE-062 study. *J Clin Oncol*. 2019;37(18_suppl). LBA4007-LBA.
16. El-Khoueiry AB, Sangro B, Yau T, Crocenzi TS, Kudo M, Hsu C, et al. Nivolumab in patients with advanced hepatocellular carcinoma (CheckMate 040): an open-label, non-comparative, phase 1/2 dose escalation and expansion trial. *Lancet*. 2017;389(10088):2492–2502.

17. Finn RS, Ryoo B-Y, Merle P, Kudo M, Bouattour M, Lim H-Y, et al. Results of KEYNOTE-240: phase 3 study of pembrolizumab (Pembro) vs best supportive care (BSC) for second line therapy in advanced hepatocellular carcinoma (HCC). *J Clin Oncol*. 2019;37(15_suppl):4004.
18. Cohen EEW, Bell RB, Bifulco CB, Burtness B, Gillison ML, Harrington KJ, et al. The Society for Immunotherapy of Cancer consensus statement on immunotherapy for the treatment of squamous cell carcinoma of the head and neck (HNSCC). *J Immunother Cancer*. 2019;7(1):184.
19. Ansell SM. The Highs and Lows of Immune-Checkpoint Blockade in Lymphoma. *Cancer immunology Res*. 2019;7(5):696–700.
20. Frenel JS, Le Tourneau C, O'Neil B, Ott PA, Piha-Paul SA, Gomez-Roca C, et al. Safety and Efficacy of Pembrolizumab in Advanced, Programmed Death Ligand 1-Positive Cervical Cancer: Results From the Phase Ib KEYNOTE-028 Trial. *J Clin oncology: official J Am Soc Clin Oncol*. 2017;35(36):4035–4041.
21. Motzer RJ, Tannir NM, McDermott DF, Aren Frontera O, Melichar B, Choueiri TK, et al. Nivolumab plus Ipilimumab versus Sunitinib in Advanced Renal-Cell Carcinoma. *N Engl J Med*. 2018;378(14):1277–1290.
22. Hellmann MD, Paz-Ares L, Bernabe Caro R, Zurawski B, Kim SW, Carcereny Costa E, et al. Nivolumab plus Ipilimumab in Advanced Non-Small-Cell Lung Cancer. *N Engl J Med*. 2019;381(21):2020–2031.
23. Wang DY, Salem JE, Cohen JV, Chandra S, Menzer C, Ye F, et al. Fatal Toxic Effects Associated With Immune Checkpoint Inhibitors: A Systematic Review and Meta-analysis. *JAMA Oncol*. 2018;4(12):1721–1728.
24. Pardoll DM. The blockade of immune checkpoints in cancer immunotherapy. *Nat Rev Cancer*. 2012;12(4):252–264.
25. Bentebibel SE, Hurwitz ME, Bernatchez C, Haymaker C, Hudgens CW, Kluger HM, et al. A First-in-Human Study and Biomarker Analysis of NKTR-214, a Novel IL2Rbetagamma-Biased Cytokine, in Patients with Advanced or Metastatic Solid Tumors. *Cancer discovery*. 2019;9(6):711–721.
26. Naing A, Wong DJ, Infante JR, Korn WM, Aljumaily R, Papadopoulos KP, et al. Pegilodecakin combined with pembrolizumab or nivolumab for patients with advanced solid tumours (IVY): a multicentre, multicohort, open-label, phase 1b trial. *lancet Oncol*. 2019;20(11):1544–1555.
27. Hamid O, Ismail R, Puzanov I. Intratumoral Immunotherapy-Update 2019. *Oncologist*. 2019.
28. Doroshow DB, Sanmamed MF, Hastings K, Politi K, Rimm DL, Chen L, et al. Immunotherapy in Non-Small Cell Lung Cancer: Facts and Hopes. *Clin cancer research: an official J Am Assoc Cancer Res*. 2019;25(15):4592–4602.
29. Pavan A, Attili I, Pasello G, Guarneri V, Conte PF, Bonanno L. Immunotherapy in small-cell lung cancer: from molecular promises to clinical challenges. *J Immunother Cancer*. 2019;7(1):205.
30. Schmid P, Rugo HS, Adams S, Schneeweiss A, Barrios CH, Iwata H, et al. Atezolizumab plus nab-paclitaxel as first-line treatment for unresectable, locally advanced or metastatic triple-negative breast cancer (IMpassion130): updated efficacy results from a randomised, double-blind, placebo-controlled, phase 3 trial. *Lancet Oncol*. 2019.
31. Amin A, Hammers H. The Evolving Landscape of Immunotherapy-Based Combinations for Frontline Treatment of Advanced Renal Cell Carcinoma. *Front Immunol*. 2018;9:3120.
32. Haugh AM, Johnson DB. Management of V600E and V600K BRAF-Mutant Melanoma. *Curr Treat Options Oncol*. 2019;20(11):81.
33. Davis AA, Patel VG. The role of PD-L1 expression as a predictive biomarker: an analysis of all US Food and Drug Administration (FDA) approvals of immune checkpoint inhibitors. *J Immunother Cancer*. 2019;7(1):278.
34. Yarchoan M, Hopkins A, Jaffee EM. Tumor Mutational Burden and Response Rate to PD-1 Inhibition. *N Engl J Med*. 2017;377(25):2500–2501.
35. Samstein RM, Lee CH, Shoushtari AN, Hellmann MD, Shen R, Janjigian YY, et al. Tumor mutational load predicts survival after immunotherapy across multiple cancer types. *Nat Genet*. 2019;51(2):202–206.
36. Klempner SJ, Fabrizio D, Bane S, Reinhart M, Peoples T, Ali SM, et al. Tumor Mutational Burden as a Predictive Biomarker for Response to Immune Checkpoint Inhibitors: A Review of Current Evidence. *Oncologist*. 2019.
37. Wang F, Zhao Q, Wang YN, Jin Y, He MM, Liu ZX, et al. Evaluation of POLE and POLD1 Mutations as Biomarkers for Immunotherapy Outcomes Across Multiple Cancer Types. *JAMA Oncol*. 2019.
38. Postow MA, Sidlow R, Hellmann MD. Immune-Related Adverse Events Associated with Immune Checkpoint Blockade. *N Engl J Med*. 2018;378(2):158–168.
39. Brahmer JR, Lacchetti C, Schneider BJ, Atkins MB, Brassil KJ, Caterino JM, et al. Management of Immune-Related Adverse Events in Patients Treated With Immune Checkpoint Inhibitor Therapy: American Society of Clinical Oncology Clinical Practice Guideline. *J Clin oncology: official J Am Soc Clin Oncol*. 2018;36(17):1714–1768.
40. Moslehi JJ, Salem JE, Sosman JA, Lebrun-Vignes B, Johnson DB. Increased reporting of fatal immune checkpoint inhibitor-associated myocarditis. *Lancet*. 2018;391(10124):933.
41. Naidoo J, Wang X, Woo KM, Iyriboz T, Halpenny D, Cunningham J, et al. Pneumonitis in Patients Treated With Anti-Programmed Death-1/Programmed Death Ligand 1 Therapy. *J Clin oncology: official J Am Soc Clin Oncol*. 2017;35(7):709–717.
42. Johnson DB, Sullivan RJ, Menzies AM. Immune checkpoint inhibitors in challenging populations. *Cancer*. 2017;123(11):1904–1911.

CAR T-Cell Therapy

Cliona M. Rooney

The ability of T cells to eliminate tumors was clear from early bone marrow transplants (BMTs) in 1965, when T-cell depletion to prevent graft versus host disease (GVHD) revealed that T cells also mediated a graft versus leukemia effect. The potential of adoptively transferred T cells to eliminate tumors was later demonstrated by the use of *in vitro* expanded tumor-infiltrating lymphocytes (TILs), demonstrating that contrary to earlier beliefs, T cells could be tumor specific. These studies led to the identification and validation first of tumor specific antigens and later of neoantigens derived from mutations.[1] However, tumor-specific T cells circulate with low frequency, and are difficult to selectively expand for therapy, especially when tumor biopsies are not available to provide a concentrated source of tumor-specific T cells. Hence, the idea to genetically modify T cells with tumor antigen-specific receptors arose. Although recombinant T-cell receptors (TCRs) have been cloned and used for tumor immunotherapy with some success, TCRs are MHC restricted and therefore must be customized to the HLA phenotype of each individual. Antibodies, by contrast, recognize unprocessed antigens and can target any tumor expressing the antigen of interest on its cell surface. Hence chimeric antigen receptors (CARs), in which antibody variable regions linked to TCR signaling domains were developed. CARs can render any T cell tumor-specific, regardless of its native TCR specificity and without the constraints of HLA restriction.

KEY CONCEPTS

Capacity of T Cells to Eliminate Tumors

- T cells
  - Eliminate tumors with high specificity and low toxicity
  - Migrate to tumor sites
  - Immense capacity to expand and persist long-term
- Any T cell can be rendered tumor-specific by expression of a chimeric antigen receptor transge

The first CARs combined the variable heavy (V_H) and light (V_L) domains of antibodies with the constant (C)α and Cβ domains of the TCR, making a two-chain CAR. The following year, Gideon Gross and Zelig Eshhar published the CAR design that is most common today.[2] In this design, the V_H and V_L antibody domains were linked in a single chain and combined with the zeta chain of the TCR to generate a cytotoxic CAR (Fig. 81.1A). Over the ensuing 30 years, CARs have been extensively modified and optimized, producing second and third generation CARs that include costimulatory endodomains that trigger proliferation and cytokine secretion, as well as tumor-cell killing (see Fig. 81.1B).[3] Of note, each CAR T cell also expresses its own TCR (Fig. 81.1C). Polyclonally activated CAR T cells comprise T cells with tens of thousands of different TCRs, the specificities of which are usually unknown. The specificity of a TCR may profoundly affect the function of the CAR if, for example, it is specific for an infecting virus.

In 2013, CAR T cells were labeled "the breakthrough of the year" by Science magazine after CAR T cells targeting CD19 produced spectacular clinical activity in patients with B-cell acute lymphocytic leukemia (ALL).[4,5] In 2017, the FDA approved CD19.CARs for the treatment of B-ALL and diffuse large B-cell lymphoma (DLBCL). CARs targeting other malignancies have been less successful, and CAR T-cell therapies face many challenges as investigators seek to increase potency without increasing toxicity. It is clear that anti-tumor efficacy requires extensive expansion of CAR T cells after infusion, migration to tumor sites, resistance to the many inhibitory cytokines and ligands encountered within the tumor microenvironment (TME), functional persistence despite a general lack of positive signals, and means to overcome heterogeneity of tumor antigen expression. These challenges will be discussed below. We will not discuss potential tumor target antigens since these have been described extensively in other reviews.[6]

KEY CONCEPTS

Tumors Develop by Avoiding the Immune Response

Tumors

- Downregulate major histocompatibility complex (MHC) molecules and target antigens
- Lack costimulatory ligands and cytokines
- Recruit inhibitory cells that express inhibitory ligands and cytokines
- High metabolism competes with T cells for glucose and amino acids
- Tumor-specific T cells are rendered anergic at the tumor site

CAR STRUCTURE

T-cell activation and proliferation requires at least three signals. The first is provided by ligation of the TCR (or CAR), the second by engagement of costimulatory receptors on T cells by complementary ligands expressed by professional antigen presenting cells (pAPCs), and the third by cytokines, produced by T cells and pAPCs. Lack of signals 2 and/or 3 can result in T-cell anergy or tolerance, and neither signal is present in most TMEs.

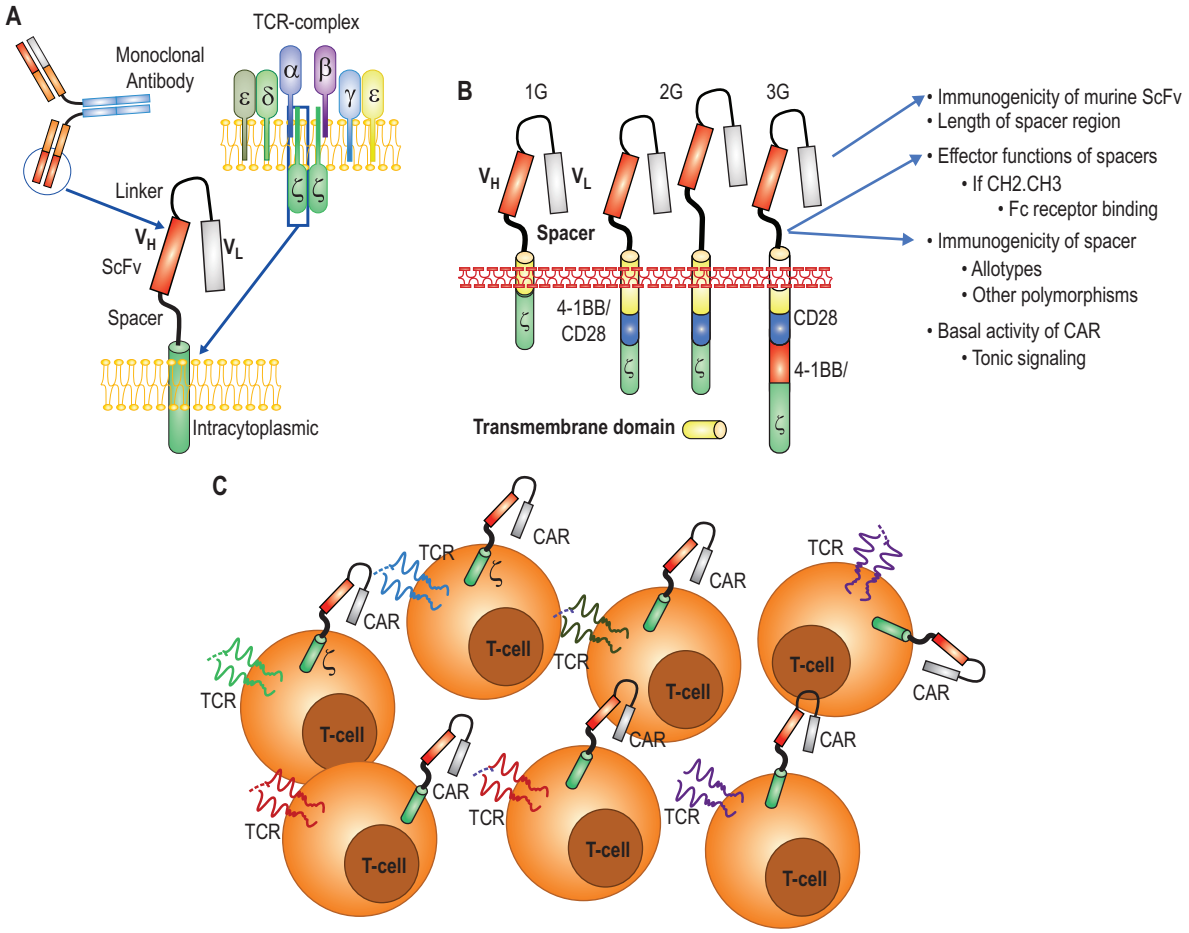


FIG. 81.1 Chimeric Antigen Receptor *(CAR)* Structure. (A) A CAR comprises an extracellular antigen-binding domain, separated by a spacer from a transmembrane domain and the intracellular ζ chain of the T-cell receptor *(TCR)*. The antigen-binding region usually combines the variable domains of the heavy and light chains of an antibody into a single-chain variable fragment *(ScFv)*. The spacer is often derived from the constant CH2 and/or CH3 regions of an antibody. The transmembrane domain shown here derives from the ζ chain. As can be seen, the structure of the CAR is very different from the structure of the TCR and, although the CAR can form a functional immune synapse, its signaling is likely to be very different from that of the TCR. (B) CAR variants. Each component of a CAR has almost infinite variants. The extracellular antigen-binding domain may be a ligand for a tumor antigen or an ScFv. The spacer may be long or short, flexible or rigid, to allow access to the cognate epitope on the targeted antigen and production of a functional immune synapse. The transmembrane domain can be substituted to improve CAR function, and may derive from the costimulatory molecule or the spacer. One or more endodomain from costimulatory molecules may be placed between the transmembrane domain and the TCR ζ chain. (C) CAR-modified T cells have dual specificity for the antigens targeted by the TCR and the CAR. There may be hundreds of thousands or more different TCRs (clonotypes) within a population of CD3 and CD28 antibody-activated CAR T cells, depending on the starting number of T cells transduced. Different TCR sequences are indicated by colors.

Inherent Problems of the Chimeric Antigen Receptor (CAR) Molecule. Most CAR ScFvs are derived from murine antibodies that are immunogenic, and may cause immune elimination. Humanization of the framework or the use of phage display libraries of human antibodies can solve this problem. Many spacers are derived from antibody constant regions that may be polymorphic, and pre-existing antibodies to constant region allotypes may limit CAR T-cell persistence. Constant regions may also bind to Fc receptors on myeloid cells, resulting in activation leading to CRS or apoptosis. The length of the spacer is important, as it determines antigen binding and may affect the avidity of the CAR. Finally, antigen-independent CAR signaling may induce T-cell differentiation and exhaustion, particularly if CARs dimerize at the cell surface or recognize T-cell antigens. *TCR*, T-cell receptor.

The first generation (1 G) of CARs combines an antigen-binding domain, usually from an antibody (but sometimes from a natural ligand), with the cytotoxic zeta chain of the TCR via a transmembrane domain and a hinge or spacer to facilitate antigen binding (see Fig. 81.1B). 1 G CAR T cells can kill tumor cells, but do not proliferate or produce cytokines in response to CAR ligation and, as a result, have been ineffective in clinical studies despite the infusion of massive T-cell numbers. Second generation (2 G) CARs include costimulatory endodomains, usually from CD28 or 4-1BB, that mediate cytokine production and proliferation and, as a result, show enhanced expansion and persistence. However, second generation CARs induce insufficient cytokines for sustained autonomous proliferation. Their true potential for both expansion and clinical efficacy is observed only when they are infused after lymphodepleting

(LD) regimens,[6, 7] most commonly cytoxan and fludarabine, that liberate the homeostatic cytokines IL7 and IL15, which maintain normal T- and B-cell numbers in the circulation.[8] After LD, these cytokines are available to infused CAR T cells, dramatically increasing their *in vivo* expansion and clinical efficacy. However, endogenous hematopoietic cells usually recover within a few weeks, reducing the availability of IL7 and IL15, so that CAR T-cell expansion peaks at one or 2 weeks, often followed by rapid decline in numbers.

The CAR is a non-physiological molecule that may present problems for its host T cell. The choice of costimulatory endodomain has been the subject of many preclinical studies. The CD28 endodomain produces more rapid effector functions than the 4-1BB domain, but the latter promotes longer T-cell persistence. Third generation CARs attempt to combine the benefits of both, and some studies indicated some clinical benefit.[7] The transmembrane and spacer domains also affect CAR function, and each single-chain variable fragment (ScFv) must be tested in different CAR backbones to optimize its function. Other factors that can limit the function and persistence of a CAR include its ability to engage the antigen and form a robust synapse with the tumor. This is affected by the avidity of the CAR for its antigen, as well as the length and rigidity of the spacer. Spacer domains that are derived from IgG CH2.CH3 domains may express allotypic antigens bound by pre-existing antibodies and may bind Fc receptors on NK cells and macrophages, resulting in activation or apoptosis of either partner. Other components of the CAR are immunogenic, such as the murine ScFv. The tendency of some CARs to dimerize, particularly if highly expressed, results in tonic signaling that can lead to T-cell differentiation and exhaustion (see Fig. 81.1B).[9] Finally, spatial and temporal differences in signaling through the TCR and the CAR are likely to affect the long-term function of CAR T cells.

Our strategy for optimizing a new ScFv is first to assess its ability to support T-cell proliferation, to kill tumor cells specifically, and to produce cytokines in response to the targeted antigen. Second, we measure the ability of the CAR to support continued T-cell proliferation and killing after repeated challenge with fresh tumor cells. Finally, the CAR is tested for its ability to eliminate tumors in murine xenograft models. Parallel experiments in which either the tumor or the T cells are labeled with bioluminescent dyes, such a luciferase, allow serial measurement of tumor responses, while bioluminescent T cells provide information of T-cell migration, expansion, and persistence, and may give an indication of tumor activity. However, preclinical models used for CAR optimization cannot predict the best outcome in patients, and even humanized mouse models cannot recreate the complex TME.

CAR T-CELL TOXICITY

Cytokine release syndrome (CRS) and immune effector cell-associated neurotoxicity syndrome (ICANs). The efficacy of CAR T cells comes at a cost. CRS is the major common severe toxicity that accompanies the logarithmic CAR T-cell expansion seen in patients with a high tumor burden and who have received LD chemotherapy. The cytokines responsible for CRS are produced not from the expanding T cells, but from macrophages that are likely activated by cytokines like GM-CSF, TNFα and IFNγ produced by activated T cells, as well as by costimulatory interactions through CD40 and CD40 ligand. Macrophage activation syndrome is often observed. CRS can be ameliorated by antibodies to IL6 (tocilizumab) and steroids. Initially, there was concern that blocking IL6 too early would inhibit the anti-tumor activity of the CAR T cells, but this has proved not to be the case; CRS is now considered an expected and manageable toxicity. ICANs is a heterogeneous group of neurotoxicities that sometimes follows CRS but does not respond to tocilizumab. Preclinical models of CRS and neurotoxicity suggested that the IL1 receptor α antagonist, anakinra, could block both CRS and neurotoxicity.[10] IL1Rα acts upstream of IL6, and anakinra blocks the production of both cytokines from macrophages. When used preemptively, this did not inhibit the anti-tumor activity of CD19.CARs. In a small clinical study at MD Anderson, anakinra alleviated ICANS in 4 of 6 patients who had received a CD19.BBz CAR for relapsed or refractory large B-cell lymphoma. It was suggested that earlier administration would likely produce better responses.[11] Several clinical trials are currently evaluating anakinra as a treatment for CRS and ICANs.

Despite the possibility of preventing CRS and ICANs, there are extensive efforts to prevent their occurrence. Ying *et al*. altered the structure of a clinically effective CD19-BBz CAR so that it produced lower levels of cytokines and proliferated less in response to CAR ligation, while retaining killing activity.[12] The authors treated 25 patients with relapsed or refractory lymphoma using escalating doses of T cells expressing the modified CAR. Six of the 11 patients on the highest dose levels receiving 2 to 4×10^8 cells achieved complete responses (CRs), two had partial responses (PRs), none of the 25 patients infused had greater than grade 1 CRS, and none had neurotoxicity.

On-target, off-tumor toxicity. Finding an antigen that is expressed on tumors but not on normal tissues is not easy. Several studies have instead targeted antigens that are overexpressed on tumor cells, hoping that normal antigen-positive tissues will escape recognition, as can be the case for therapeutic antibodies. Patients can survive complete loss of normal $CD19^+$ B cells, often with gamma globulin supplementation, and although B cells recover as CD19.CAR levels decrease, this may take more than 2 years. Other CAR target tissues are not so forgiving; fatalities have occurred in patients receiving CARs targeting HER2neu, EGFR variant 3, and carbonic anhydrase IX (CAIX).[13] A patient receiving a high dose of HER2.CAR T cells for metastatic colon cancer at the NCI experienced rapid pulmonary edema 15 minutes after infusion, progressing to multiorgan failure and death. This was considered a result of cytokine storm induced by CAR T-cell recognition of low levels of HER2 on normal lung tissue. At our center, patients with glioblastoma multiforme (GBM) and osteosarcoma received T cells expressing a HER2.28ζ CAR with a different ScFv without experiencing toxicities, but also without major clinical efficacy.[6] None of these patients received preinfusion LD, which seems to be essential for effective CAR T-cell therapy. In a follow-up phase I trial (NCT00902044), this group infused T cells modified with the same HER2.CAR after LD, and reported one CR for over 20 months in a child with metastatic rhabdomyosarcoma who received seven doses of HER2.CAR T cells.[14] This patient experienced grade 1 CRS after each of the three infusions that were preceded by LD, but no pulmonary or cardiac toxicity despite the presence of CAR T cells in the circulation for over 200 days.

Many preclinical strategies to overcome on-target, off-tumor toxicity have been described, including dialing down CAR T-cell avidity (so that only antigens overexpressed on tumor cells are recognized), or using Boolean gating, in which "and" "or" "not" gates are used to allow tumor-cell killing only when a predetermined pattern of antigens is present.[13] In such strategies, the T cell expresses two or more receptors for antigens differentially expressed on tumors and normal tissues. A cytotoxic CAR might recognize and kill target cells only in the absence of second antigen present on a healthy cell that would deliver the inhibitory signal via an inhibitory CAR. In a second strategy, a 1 G CAR is combined with a non-cytotoxic costimulatory CAR that delivers signal 2 so that the dual CAR T-cell proliferation occurs only if a tumor-specific combination of antigens is present. Many of these strategies rely on cis expression of antigens, and it is not clear if co-inhibition by local or distant healthy cells could prevent tumor-cell killing. Roybal's group designed a synthetic notch receptor that on ligation of an extracellular antigen-binding domain by an antigen in the TME, cleavage of the notch transmembrane domain releases a synthetic intracellular transcription factor to trigger expression of genes designed to enhance the anti-tumor activity of the CAR T cells.[15]

As the potency of CAR T cells is increased genetically, suicide genes that can be activated rapidly become increasingly important.[15] Suicide genes were first evaluated for the treatment of GVHD mediated by allogeneic donor T cells after hematopoietic stem cell transplantation (HSCT). Among the first was the herpes simplex thymidine kinase (*TK*) gene that could convert a nontoxic prodrug (ganciclovir) into a toxic metabolite. The problems with the viral TK enzyme are, first, that it is immunogenic, leading to unwanted immune elimination of transgenic T-cells, and second, that it precludes the use of ganciclovir as an antiviral agent. To solve these problems, we developed an inducible human caspase 9 (icasp9), in which the native dimerization domain of human caspase 9 was replaced by a modified human FK binding domain that binds with high affinity to a small molecule chemical inducer of dimerization. Administration of the dimerizer led to rapid elimination of icasp9-expressing T cells both in preclinical models and in HSCT recipients who had received icasp9-modified donor T cells to enhance immune reconstitution. The dimerizer was infused in patients who developed GVHD, resulting in a 90% decrease of transduced cells within 30 minutes and an additional one log depletion in 24 hours. Symptoms of GVHD resolved within 24 to 48 hours, and did not recur.[16]

ANTIGEN LOSS AND HETEROGENEITY

The heterogeneity of the targeted antigen expression and CAR T-cell-driven antigen loss are major challenges when a single epitope in a single antigen is targeted. Unless the antigen and indeed the epitope are crucial for tumor survival, they may easily be downregulated, and antigen loss variants can be selected in the face of CAR T-cell killing. CD19 expression is essential for B-cell survival and is expressed on all resting and activated normal and malignant B cells during all phases of development, until it is downregulated on plasma cells. As a result, 70% to 90% of patients with relapsed or refractory B-cell acute lymphoblastic leukemia (B-ALL) enter CR in response to CD19.CAR T cells. Although 30% to 60% of these patients will relapse, only 10% to 20% relapse with CD19-negative disease. Interestingly, a significant proportion of these are associated with myeloid lineage switch.

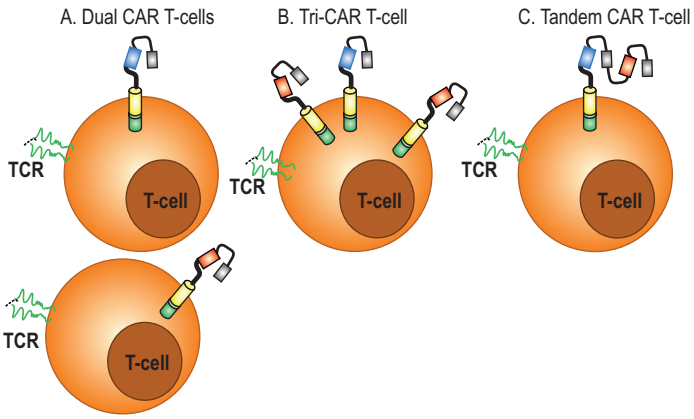


FIG. 81.2 Overcoming Tumor Heterogeneity. (A) Multiple populations of T cells, each expressing a different chimeric antigen receptor *(CAR)*, can be infused to limit antigen heterogeneity or antigen loss. This strategy is limited only by the expense of the clinical grade vectors. (B) Shows a Tri-CAR T cell that recognizes three different tumor antigens. This strategy can also increase the avidity of a weak CAR for antigens expressed at low levels. If expressed from different vectors, this strategy may be challenged by FDA limits on vector copy number. This can be overcome in (C), in which two ScFvs are linked in series in the same molecule. Tandem CARs must be designed based on biophysical properties to allow two different antigens to be targeted by a single molecule. *TCR*, T-cell receptor.

Few CAR target antigens are as stable or homogeneously expressed as CD19, and the outgrowth of antigen-negative tumor cells after CAR T-cell treatment has been shown repeatedly in preclinical tumor models. To overcome this problem, investigators have combined CARs with different specificities, expressed on the same or different T-cell populations in separate or single vectors. In innovative strategies, ScFvs have been combined with linkers to generate tandem CARs in single molecules that have proved effective in increasing the avidity of the CAR for tumors expressing low levels of antigens (Fig. 81.2). Hegde *et al.* showed that a tandem CAR with specificity for both HER2 and IL13Rα2 could engage both target antigens and produced greater control of GBM in xenograft models than T cells expressing both HER2.CAR and IL13Rα2.CAR. The same group[17] showed that while HER2, IL13 receptor α2, and ephrin-A2 were differentially expressed on primary GBM samples, a tricistronic vector encoding CARs specific for each of the three antigens was able to target most of the cells in primary tumors and showed improved tumor control in xenograft models, compared to mono-or bi-specific CAR T cells.[18]

TARGETING HEMATOLOGIC MALIGNANCIES

CAR T cells targeting B-cell malignancies have so far surpassed other CAR T cells in efficacy, likely in part because B cells are professional antigen-presenting cells that express costimulatory molecules and cytokines that help CAR T-cell persistence and expansion, and in part because they share the B-cell lymphoid niche. Although costimulatory molecules are not expressed by all B-cell tumors, they are expressed by normal B cells in lymphoid tissues and may be upregulated on tumor cells exposed to cytokines, such as IFNγ, TNFα, and GM-CSF, that are produced by

CAR T cells on ligation of the CAR. T cells naturally circulate in lymphoid tissues in which B-cell tumors are located, and therefore are more likely to access tumor tissues; lymphoid tissues provide a supportive niche for T-cell expansion. CD19.CARs have shown outstanding success for the treatment of B-ALL, B-cell lymphoma, and chronic lymphocytic leukemia, while BCMA. CARs for the treatment of multiple myeloma are also promising, producing tumor responses in up to 85% of patients including CRs in 45%.[19] CD30.CARs have also been highly effective for B-cell lymphoma, and sequential trials at our institution clearly demonstrate the added value of lymphodepletion.[7]

T-cell and myeloid-cell malignancies are more challenging, as neither population is dispensable. Most antigens expressed on T-cell malignancies are also expressed on healthy T cells, so it is often difficult even to grow the CAR.T cells. Gomes-Silva *et al.* overcame this problem by knocking out CD7 on CD7.CAR T-cells to prevent fratricide, allowing CD7.CAR T cells to target both T-cell malignancies and AML.[20] In contrast, CD5.CAR- and CD30.CAR T cells solved this problem for themselves. CD5 is downregulated on CD5 CAR T cells and CD30 is masked in *cis* by the CD30.CAR. However, without a suicide strategy, CARs targeting normal T cells may be limited to use as a bridge to transplant, which can be vital for patients who do not respond to standard therapies and hence are not eligible for HSCT. CD30 expression, in contrast, is limited to activated T cells; CD30.CAR T cells have been safe and effective for both T- and B-cell malignancies, notably without an increase in viral infections.[21] However, this might be because CD30.CAR T cells peaked at one week after infusion, and by 8 weeks represented less than 0.1% of circulating PBMCs.

CARs targeting AML antigens like CD123, CD33, and CLL1 are currently in clinical trials at multiple centers throughout the world. However, most are also expressed on healthy myeloid cells or their progenitors, so again are used as a bridge to transplant, combined with suicide switches, or even with hematopoietic stem cells knocked out for the target antigen. AML also suffers from heterogeneous antigen expression, and multitargeted CARs will likely be required to sustain remissions.

SOLID TUMORS AND THEIR IMMUNOSUPPRESSIVE MICROENVIRONMENT

CAR T cells have had less success for the treatment of solid tumors. Access to solid tumors, in contrast with hematologic malignancies, requires T cells to extravasate into non-lymphoid tissues, which, in the absence of specific chemokine signaling, is slow. CAR T cells arriving at the tumor site are often found outside the tumor core, likely blocked by the multiple inhibitory components of the TME, which includes suppressive myeloid cells, T regulatory cells, and tumor associated fibroblasts (Fig. 81.3). Within the tumor core, T cells are subject to suppressive ligands such as PD-L1, galectin 9, ceacam 1, cytokines such as TGF-β and IL-10, and metabolic inhibitors like arginase 1 and indoleamine dioxygenase (IDO), while costimulatory ligands and growth-promoting cytokines are lacking. Clinical trials targeting solid tumors have been disappointing, but PRs and rare CRs lend promise to their potential. It is now accepted that effective CAR T cells will require concomitant strategies to counter the suppressive TME. Many such strategies have been evaluated preclinically and are making their way to the clinic. These strategies include additional genetic modifications of the CAR T cells and combination with monoclonal antibodies, small molecules that inhibit suppressive myeloid cells, and oncolytic viruses.

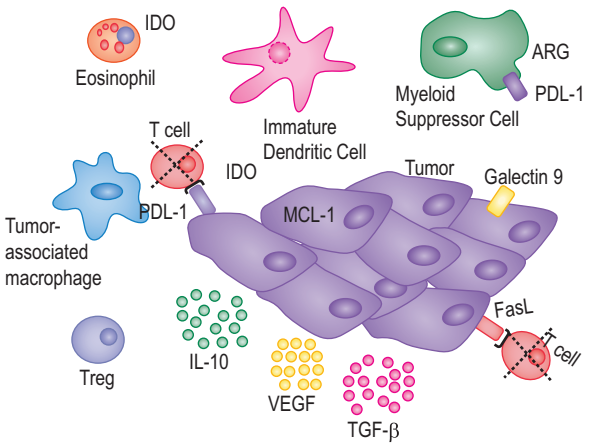


FIG. 81.3 Suppressive Components of the Tumor Microenvironment. This figure shows just some of the immunosuppressive components that promote tumor growth and metastasis, increase angiogenesis, and inhibit the infiltration and function of tumor-specific T cells. Targeting just one of these inhibitory factors may be insufficient to relieve T-cell inhibition, and inhibitors that interfere with multiple factors may be required. Multiple factors may be tackled at once, for example, by eliminating suppressive myeloid cells that exert multiple inhibitory actions. *IDO*, Indoleamine dioxygenase; *IL*, interleukin; *TGF*, transforming growth factor beta, *VEGF*, vascular endothelial growth factor.

KEY CONCEPTS

Suppressive Cells of the Tumor Microenvironment

- T regulatory cells
- Suppressive myeloid cells
  - Tumor-associated macrophages, myeloid suppressor cells, immature dendritic cells
- Express inhibitory and cytotoxic ligands
  - PD1, galactin 9, CD48, CD122, CD155, RCAS1, Fas ligand
- Produce metabolic inhibitors
  - Indoleamine dioxygenase, ARG1, adenosine, lactic acid, hypoxia

COMBATING THE TME BY GENE MODIFICATION OF CAR T CELLS

Multiple strategies to improve CAR T-cell proliferation after infusion, homing to, and functional persistence at the tumor site have been evaluated preclinically; several are in clinical trials. Homing of CAR T cells to the tumor site can be improved by co-expression of transgenic receptors for chemokines expressed within the TME, either by tumor cells or by tumor stroma. Maintaining CAR T-cell persistence after the effects of lymphodepletion have waned may be achieved by the systemic administration of cytokines, but since these have dose-limiting toxicities, several trials are evaluating CAR T cells producing cytokines either constitutively or in response to CAR ligation. To prevent excessive production of transgenic cytokines and their use by unwanted cell types, such as Tregs, constitutively active cytokine receptors can be expressed in CAR T cells (Fig. 81.4C).

TGFβ, produced by many tumors and by immunosuppressive tumor components, is an important homeostatic cytokine that potently inhibits T cells. We expressed a dominant-negative

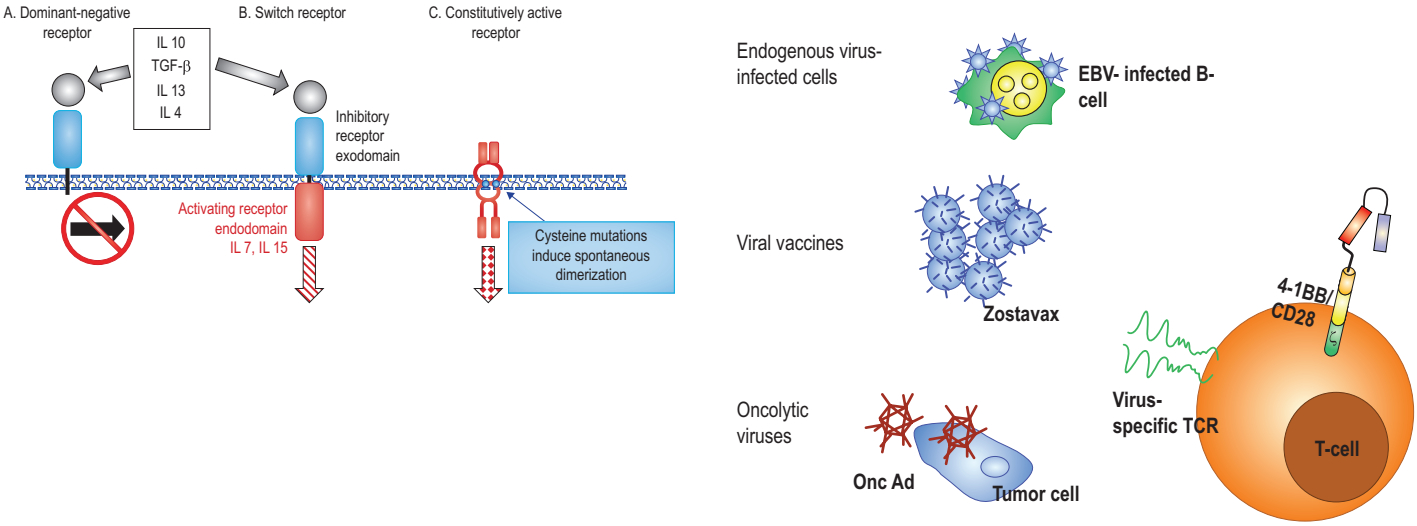


FIG. 81.4 Enhancing Chimeric Antigen Receptor (CAR) T-Cell Signaling. (A) A dominant-negative receptor for an inhibitory ligand in which the intracellular signaling domain is deleted, so the cell is not inhibited on ligand binding. In (B), the extracellular domain of an inhibitory receptor is linked to an intracellular signaling domain of an activating receptor, so activating signal can convert the inhibitory signal into a stimulatory signal. Both receptors can bind their ligand and deplete it from the environment. (C) A cysteine mutation in the transmembrane domain of the IL7 receptor leads to constitutive IL7 receptor signaling, leading to Jak1-STAT5 signaling and supporting cytokine-independent T-cell proliferation. In (D), CARs can be introduced into virus-specific T cells so that additional signaling can be induced via the virus-specific TCR. Viruses potently stimulate both adaptive and innate immunity via pattern recognition receptors on infected cells and innate responders. Hence, CAR virus-specific T cell (VST) stimulation may be provided by endogenous viruses such as EBV, viral vaccines such as the live-attenuated VZV vaccine, zostavax, or oncolytic viruses, such as adenoviruses that can also recruit CAR VSTs to the infected tumor site. *IL*, Interleukin; *TCR*, T-cell receptor.

KEY CONCEPTS

Gene Modifications to Improve CAR T-Cell Function and Persistence

- Growth-promoting cytokines or constitutive cytokine receptors
- Dominant-negative or switch receptors for inhibitory ligands
- Multiple chimeric antigen receptors to avoid antigen heterogeneity and downregulation
- Immune modulatory payloads, such as mini-antibodies to checkpoints or IL12
- Suicide strategies

TGFβ receptor (TGFβDNR) on EBV-specific T-cells (EBVSTs) and showed they were resistant to TGFβ *in vitro* and in preclinical tumor models (see Fig. 81.4A). The removal of a homeostatic mechanism from T cells was a safety concern since TGFβ knockout mice developed T-cell leukemia. However, these leukemias were restricted to IL2-independent pre-thymic T cells and the DNR was not oncogenic in mature post-thymic T cells. Indeed, TGFβDNR-EBVSTs were safe and produced CRs in patients with multiply relapsed or refractory EBV-positive lymphoma and could be detected in patients for up to 5 years.[15] Since then, the TGFβDNR has been used safely in HER2.CAR T cells, melanoma-specific TILs, and HPV-specific T cells. Clearly, there are multiple checkpoints for T-cell proliferation; removing a single one is likely to be safe and worth testing in patients with few, if any, other options. Switch receptors take this concept one step further and combine the extracellular domain for inhibitory cytokines, such as TGF-β, with the intracellular signaling domains of activating cytokines, such as IL-7 to convert an inhibitory signal into an activating signal (see Fig. 81.4B).[6] Other strategies use CAR T cells to deliver immune modulating payloads such as mini-antibodies targeting checkpoint molecules.

Sitkovsky's group has long been a proponent of targeting hypoxia and adenosine in the TME.[22] Hypoxia in the TME drives the expression of the ectoenzymes CD39, that converts extracellular ATP and ADP into AMP, while CD73 produces adenosine from AMP. Adenosine activates A2 adenosine receptors (A2AR) on T cells, activating cyclic AMP and inhibiting T-cell-receptor signaling. Several groups have demonstrated the antitumor activity of A2R antagonists on their own and in combination with other immune modulators such as checkpoint inhibitors.

KEY CONCEPTS

Combination Strategies to Improve CAR T Cells

- Molecules that eliminate suppressive myeloid cells
  - All-trans retinoic acid, Rank ligand inhibitors, bisphosphonates
- Epigenetics molecules that upregulate tumor antigens
  - Histone deacetylase inhibitor (HDAC) inhibitors, methylation inhibitors
- Oncolytic viruses that debulk tumor and produce immunomodulatory transgenes
- Checkpoint inhibitors to disable the suppressive tumor microenvironment.

THE DIFFERENTIATION STATE AND SUBSET ORIGIN OF THE INFUSED T CELLS IS IMPORTANT

Generation of CAR T cells usually involves stimulating T cells with CD3 and CCD28 antibodies, followed by transduction with vectors expressing the CAR of interest, then expansion with cytokines until sufficient T cells for the patient dose, quality control (QC) testing for eligibility, and other analyses have been obtained. Within these parameters, there are multiple variables in addition to the structure and specificity of the CAR: the source of T cells may be peripheral blood or apheresis units that may be processed into PBMCs, purified $CD3^+$ T cells, or more defined subsets. The vector may lentiviral, retroviral, or transposon and the media used may include fetal bovine serum, human serum, human platelet lysate, or serum-free formulations, containing a range of cytokines and sometimes incorporating drugs such as Akt inhibitors or IL21 to inhibit T-cell differentiation. T cells may be expanded for just a few days or for up to 2 weeks in culture ware ranging from 24 well plates, to G-Rex flasks, the Wave bioreactor, or the Prodigy. Finally, cells may be infused fresh or after cryopreservation. The relative benefits of most of these variables have not been directly compared in clinical trials, and preclinical *in vitro* or *in vivo* studies may have limited value. The commonly held theory is that the least differentiated T-cell phenotype is desirable, and that cells must proliferate and kill tumor cells over multiple *in vitro* stimulations and be effective in murine xenograft models.

It has been suggested in preclinical studies that CD19 .BB. CAR T cells with a defined T-cell subset composition of 50% central memory (cm)-derived $CD8^+$ T cells plus 50% $CD4^+$ T cells from unseparated T cells had better preclinical antitumor activity than other subset combinations. This combination was later studied in a clinical trial for adult patients with heavily pretreated NHL,[23] with half of patients receiving the CD19.BB.CAR after cytoxan and fludarabine preconditioning achieving CR. However, this study did not determine if this subset combination was more effective than unseparated subsets. To identify the characteristics of T cells with the greatest persistence and expansion after infusion, this group compared $CD8^+$ T-cell clonotypes in the infusion product with the cells detected in patient blood at different times after infusion using single cell RNA seq. T cells in the infusion product fell into four groups associated with T-cell activation, effector function, metabolism, and cell cycle, and clonotypes that persisted fell into all four groups, but most commonly derived from groups with enhanced cell cycle and effector functions. The efficacy of memory-enriched CAR T cells is currently under evaluation in several clinical trials for B-cell malignancies and brain tumors, some in combination with PD-L1 blockade (NCT02706405).

We have evaluated CARs in virus-specific T cells (VSTs) that derive from the T-cell memory compartment. We hypothesized that CAR VSTs could be boosted *in vivo* by endogenous viruses, such as EBV, viral vaccines such as VZV, or oncolytic viruses (see Fig. 81.4D), and that if infused after HSCT, they could protect against both viral infections and leukemic relapse. An intrapatient comparison of a first generation GD2.CAR expressed in EBVSTs or CD3-activated T-cells (ATCs) showed that GD2. CAR-EBVSTs circulated with higher frequency in the 6 weeks after infusion than GD2.CAR ATCs and produced CRs in 3 of 11 patients. However, neither population (CAR EBVSTs or CAR ATCs) expanded after infusion and persistence even at low levels of either population correlated positively with time to progression. We also used CD19.28.CAR T cells specific for CMV, adenovirus, and EBV (CD19.CAR-VSTs) as adjuvant therapy after HSCT for children with high risk B-ALL.[24] Patients did not receive lymphodepletion, were not lymphopenic at the time of infusion, and had circulating normal B cells. Virus reactivation correlated with much greater expansion of CAR-VSTs and loss of normal B cells, showing that viruses can expand CAR T cells via the TCR. This study also showed that in the absence of lymphodepletion, normal B cells were unable to induce substantial CD19. CAR.T-cell expansion.

INTRAPATIENT COMPARATIVE STUDIES CAN BEST IDENTIFY OPTIMAL CAR T CELLS

There are many different CAR structures and manufacturing strategies, and it remains unclear which is the best. Intra-patient comparisons, in which CAR T cells are introduced into the same patient environment, provide powerful means to compare post-infusion expansion and persistence. In addition to comparing GD2.CAR EBVSTs with GD2.CAR ATCs within patients, our group compared first (1 G) and second generation (2 G) CD19. CARs, and then 2 G with 3 G CD19.CARs. Each patient received two cell products simultaneously, and the relative expansion of each could be compared using quantitative PCR. As expected, the 2 G CD19.CARs expanded to a greater degree than 1 G CD19.CARs in patients with B-cell lymphoma, as measured by quantitative PCR of PBMCs. Although none of the six patients had CRs, the cells were infused without prior lymphodepletion. In the second study, 3 G CARs infused after lymphodepletion expanded to a greater degree than 2 G CARs, and persisted for longer, although both peaked at 2 weeks after infusion.[7] Three of 11 patients with disease at the time of infusion had CRs and three had PRs. While powerful, these studies cannot determine precisely which product had the greatest clinical benefit, since T-cell expansion and persistence within the tumor site and lymphoid tissues could not be measured and may not be accurately reflected by the cells detected in the circulation. Hence, there is a great need for improved *in vivo* imaging and biopsy analyses.

OFF-THE-SHELF CAR T CELLS

The generation of autologous CAR T cells for every patient is an expensive and lengthy process, and patients cannot always wait the 4 months usually required to generate and QC test their product. Furthermore, individual CAR T-cell products cannot be standardized, and many fail the product eligibility criteria. Allogeneic off-the-shelf CAR T cells can be generated in bulk to treat multiple patients in the shortest possible time frame. Unfortunately, alloreactive T cells in the infusion product may cause GVHD, while host alloreactive T cells can rapidly eliminate the infused cells. Despite this problem, partially HLA-matched, allogeneic VSTs have been almost as effective as stem cell donor-derived VSTS at controlling viral infections after HSCT, without causing GVHD. However, graft rejection is far more likely in fully immunocompetent hosts, and investigators have used gene-editing approaches to solve both problems. To prevent GVHD, they have deleted the TCR, and to prevent rejection, they have prevented the cell surface expression of HLA class I and II molecules. Since HLA-negative T cells are sensitive to NK-cell killing, NK-cell-inhibitory molecules such as HLA E must also be expressed. However, NK-cell-inhibitory

receptors are heterogenous, and HLA E may not block all NK cells. Our group has evaluated alternative strategies such as allodefense receptors (ADRs), which comprise CARs that recognize molecules upregulated on activated T- and NK-cells. Co-expression of an ADR of CD19.CAR T cells has proved effective in humanized xenograft models.[25] We are also evaluating VSTs as hosts for ADRs, since allogeneic VSTs with their more limited array of TCRs have not caused GVHD in multiple clinical trials.

SUMMARY

CAR T cells hold great promise as a therapy with high tumor specificity that causes few, if any, long-term toxicities. If their anti-tumor efficacy in the immunosuppressive TME can be improved, CAR T cells would be infinitely preferable to most standard chemoradiotherapies and surgery. To achieve this goal, combination strategies that enhance the potency of CAR T cells and reverse the immunosuppressive TME will likely be required. Finally, an affordable and feasible model to bring cellular immunotherapies into clinical practice is required.

REFERENCES

1. Yang JC, Rosenberg SA. Adoptive T-cell therapy for cancer. *Adv Immunol*. 2016;130:279–294.
2. Eshhar ZT, Waks Gross G. The emergence of T-bodies/CAR T cells. *Cancer J*. 2014;20(2):123–126.
3. Dotti G, Gottschalk S, Savoldo B, et al. Design and development of therapies using chimeric antigen receptor-expressing T cells. *Immunol Rev*. 2014;257(1):107–126.
4. Kalos M, Levine BL, Porter DL, et al. T cells with chimeric antigen receptors have potent antitumor effects and can establish memory in patients with advanced leukemia. *Sci Transl Med*. 2011;3(95):95ra73.
5. Brentjens RJ, Riviere I, Park JH, et al. Safety and persistence of adoptively transferred autologous CD19-targeted T cells in patients with relapsed or chemotherapy refractory B-cell leukemias. *Blood*. 2011;118(18):4817–4828.
6. Wagner J, Wickman E, DeRenzo C, et al. CAR T cell therapy for solid tumors: bright future or dark reality? *Mol Ther*. 2020;28(11):2320–2339.
7. Ramos CA, Rouce R, Robertson CS, et al. In vivo fate and activity of second- versus third-generation CD19-Specific CAR-T cells in B Cell Non-Hodgkin's lymphomas. *Mol Ther*. 2018;26(12):2727–2737.
8. Lowe KL, Mackall CL, Norry E, et al. Fludarabine and neurotoxicity in engineered T-cell therapy. *Gene Ther*. 2018;25(3):176–191.
9. Long AH, Haso WM, Shern JF, et al. 4-1BB costimulation ameliorates T cell exhaustion induced by tonic signaling of chimeric antigen receptors. *Nat Med*. 2015;21(6):581–590.
10. Norelli M, Camisa B, Barbiera G, et al. Monocyte-derived IL-1 and IL-6 are differentially required for cytokine-release syndrome and neurotoxicity due to CAR T cells. *Nat Med*. 2018;24(6):739–748.
11. Strati P, Ahmed S, Kebriaei P, et al. Clinical efficacy of anakinra to mitigate CAR T-cell therapy-associated toxicity in large B-cell lymphoma. *Blood Adv*. 2020;4(13):3123–3127.
12. Ying Z, Huang XF, Xiang X, et al. A safe and potent anti-CD19 CAR T cell therapy. *Nat Med*. 2019;25(6):947–953.
13. Andrea AE, Chiron A, Bessoles S, et al. Engineering next-generation CAR-T cells for better toxicity management. *Int J Mol Sci*. 2020;21(22):8620.
14. Hegde M, Joseph SK, Pashankar F, et al. Tumor response and endogenous immune reactivity after administration of HER2 CAR T cells in a child with metastatic rhabdomyosarcoma. *Nat Commun*. 2020;11(1):3549.
15. Bollard CM, Tripic T, Cruz CR, et al. Tumor-specific T-cells engineered to overcome tumor immune Evasion induce clinical responses in patients with relapsed Hodgkin lymphoma. *J Clin Oncol*. 2018;36(11):1128–1139.
16. Zhou X, Brenner MK. Improving the safety of T-cell therapies using an inducible caspase-9 gene. *Exp Hematol*. 2016;44(11):1013–1019.
17. Hegde M, Mukherjee M, Grada Z, et al. Tandem CAR T cells targeting HER2 and IL13Ralpha2 mitigate tumor antigen escape. *J Clin Invest*. 2019;129(8):3464.
18. Han X, Wang Y, Wei J, et al. Multi-antigen-targeted chimeric antigen receptor T cells for cancer therapy. *J Hematol Oncol*. 2019;12(1):128.
19. Yang Y, Li y, Gu H, et al. Emerging agents and regimens for multiple myeloma. *J Hematol Oncol*. 2020;13(1):150.
20. Scherer LD, Brenner MK, Mamonkin M. Chimeric antigen receptors for T-cell malignancies. *Front Oncol*. 2019;9:126.
21. Ramos CA, Grover NS, Beaven AW, et al. Anti-CD30 CAR-T cell therapy in relapsed and refractory Hodgkin lymphoma. *J Clin Oncol*. 2020;38(32):3794–3804.
22. Hatfield SM, Sitkovsky M. A2A adenosine receptor antagonists to weaken the hypoxia-HIF-1alpha driven immunosuppression and improve immunotherapies of cancer. *Curr Opin Pharmacol*. 2016;29:90–96.
23. Turtle CJ, Hanafi LA, Berger C, et al. CD19 CAR-T cells of defined CD4+:CD8+ composition in adult B cell ALL patients. *J Clin Invest*. 2016;126(6):2123–2138.
24. Lapteva N, Gilbert M, Diaconu I, et al. T-Cell receptor stimulation enhances the expansion and function of CD19 chimeric antigen receptor-expressing T cells. *Clin Cancer Res*. 2019;25(24):7340–7350.
25. Mo F, Watanabe N, McKenn MK, et al. Engineered off-the-shelf therapeutic T cells resist host immune rejection. *Nat Biotechnol*. 2021;39(1):56–63.

Immunoglobulin Therapy: Replacement and Immunomodulation

Mark Ballow

Over 70 years ago, a cold ethanol fraction of plasma that contained an enriched fraction of gammaglobulin was used intramuscularly as passive immunotherapy for the treatment and protection of infectious pathogens and subsequently as antibody replacement therapy for patients with primary immune deficiency diseases (PIDDs). This Cohn ethanol plasma fraction of immunoglobulin G (IgG) remained the principal form of therapy until 1981 when an intravenous (IV) preparation (i.e., intravenous immunoglobulin [IVIG]) became available. Subsequently, a Swiss physician observed that thrombocytopenia resolved when patients with immune deficiency were treated with IVIG. This observation led to the use of IVIG in patients with autoimmune idiopathic thrombocytopenic purpura (ITP), and an expansion on the use of IVIG as immunomodulatory therapy in several US Food and Drug Administration (FDA) approved autoimmune disorders (Table 82.1). Most IVIG preparations are derived from plasma by the Cohn ethanol fractionation method or the Cohn-Oncley modification. IVIG products are modified to prevent the formation of IgG aggregates and to make IVIG suitable for the IV route. Excipients, such as sugars (maltose or D-sorbitol) or amino acids (glycine or L-proline), stabilize the IgG molecules from aggregation. Cold ethanol fractionation, as the first step in plasma processing, inactivates human immunodeficiency virus (HIV). Treatment with low pH, solvent and detergent, pasteurization or nanofiltration/depth filtration in combination, depending on the Ig product, is used as further five to seven steps for viral inactivation and removal. Commercial lots of IVIG are derived from approximately 15,000 donors (not to exceed 60,000 donors according to the FDA and the Plasma Protein Therapeutics Association). Each lot must contain adequate levels of antibody to measles. These products may vary slightly among manufacturers and from lot to lot; however, they are generally comparable with regard to clinical efficacy, but perhaps not tolerability. The newer Ig formulations are iso-osmolar, low-sodium, and low-IgA liquid products. The characteristics of IVIG preparations available in the United States are reviewed elsewhere (*Immune Deficiency Foundation. Immunoglobin Products. Available at: https://primaryimmune.org/about-primary-immunodeficiencies-treatments/immunoglobulin-products [Accessed on March 23, 2020]*).

TABLE 82.1 FDA-Approved Indications for Intravenous Immunoglobulin

- Primary humoral immunodeficiency disease, as replacement therapy.
- Immune thrombocytopenic purpura, to prevent severe and/or control bleeding
- Children with HIV/AIDS and recurrent infections, to prevent serious bacterial infections
- B-cell chronic lymphocytic leukemia with recurrent infections and humoral immune deficiency, to prevent bacterial infection
- Kawasaki disease, to prevent coronary artery aneurysms
- Bone marrow transplantation, to decrease risk of infection, interstitial pneumonia, and graft-versus-host disease in the first 100 days after transplantation
- Chronic inflammatory demyelinating polyneuropathy, to improve neuromuscular impairment and for maintenance therapy to prevent relapse
- Multifocal motor neuropathy, to improve neuromuscular impairment and maintenance therapy to prevent relapse

AIDS, Acquired immunodeficiency syndrome; *FDA*, US Food and Drug Administration; *HIV*, human immunodeficiency virus.

KEY CONCEPTS

Properties of Intravenous Immunoglobulin and Subcutaneous Immunoglobulin

- Plasma fractionation (first step) by cold ethanol/Cohn-Oncley modification (fraction II)
- >98% IgG; >90% monomeric IgG
- Traces of other immunoglobulins (i.e., IgA and IgM) and serum proteins
- Addition of sugar or amino acids, stabilizes IgG from aggregation
- Multiple viral inactivation/removal steps:
  - Cold ethanol fraction
  - Chromatography
  - Solvent/detergent treatment
  - Caprylate fractionation
  - Nanofiltration
  - Depth filtration
  - Pasteurization
- Intact Fc receptor important for biological function:
  - Opsonization and phagocytosis
  - Complement activation
  - Antibody-dependent cytotoxicity
- Normal half-life comparable to serum IgG
- Normal proportion of IgG subclasses
- Broad spectrum of antibodies to bacterial and viral agents

In this chapter, the application of IVIG as replacement therapy in patients with primary immune deficiency (PID) and the potential mechanisms of action of Ig therapy in the treatment of autoimmune and inflammatory diseases are reviewed. For more information on the use of Ig therapy in specific diseases, the reader is referred to an evidence-based review of the topic elsewhere.[1]

REPLACEMENT THERAPY WITH INTRAVENOUS IMMUNOGLOBULIN

Several studies have shown that pulmonary abnormalities are the most important factors associated with morbidity and mortality in patients with PIDs. The number of infections, days on antibiotic therapy, days missed from school or at work, and hospitalized days may not be sufficient indicators of adequate treatment. Therefore the improvement or maintenance of pulmonary function is an important measure of the success of therapy. Orange et al.[2] examined the impact of serum IgG trough levels on pneumonia incidence in patients with PIDDs receiving replacement IVIG therapy in a meta-analysis of published clinical studies. The results of their analysis showed that trough IgG levels increased by 121 mg/dL for every 100 mg/kg increase in dose, which resulted in a decrease in pneumonia incidence by 27%. There was a strong correlation between increasing trough IgG levels and a decrease in pneumonia. Clinicians should choose a "biological" trough level for the serum IgG related to the patient's clinical course (i.e., frequency of infections) instead of arbitrarily choosing a trough value of ≥500 mg/dL. A longitudinal study of patients with common variable immune deficiency (CVID) and X-linked agammaglobulinemia with a follow-up of 22 years showed that the dose of IVIG needed to keep a patient free of infection varied among patients, emphasizing that the goal of replacement therapy should be improvement of clinical outcome, not a specific IgG trough level.

The overall consensus among clinical immunologists is that an IVIG dose of 400 to 600 mg/kg/month is a good *starting point*. Catabolism of infused Ig varies among individuals. The rate of elimination of IgG may be higher during a period of active infection. For persons with a higher catabolism of infused IgG or more frequent infections, infusions every 3 weeks with smaller doses may be more efficacious. In the final analysis, trough levels are only a general guide, and the clinical well-being of the patient is a more important parameter. Clearly, with higher trough levels, the incidence of pneumonia and comorbid conditions (e.g., bronchiectasis and meningoencephalitis) is reduced. Pulmonary abnormalities are among the most frequent comorbid conditions associated with morbidity and mortality in patients with antibody immunodeficiencies. Periodic pulmonary function testing and careful use of high-resolution chest computed tomography should be used to monitor pulmonary function and anatomy.

For replacement therapy in patients with PID, all brands of IVIG are probably equivalent for efficacy, although there are differences in viral inactivation processes (see *Immune Deficiency Foundation. Immunoglobin Products. Available at: https://primaryimmune.org/about-primary-immunodeficiencies-treatments/immunoglobulin-products [Accessed on March 23, 2020]*). The choice of brands may be dependent on the hospital, payor, or home care formulary; the distributor availability and cost; and, most importantly, patient tolerability. The dose, manufacturer, and lot number should be recorded for each infusion to facilitate review procedures for adverse events or other consequences. It is crucial to record all side effects that occur during the infusion. It is also recommended that liver and renal functions be monitored once or twice a year. Changes in weight or growth in children or changes in clinical response or adverse events may necessitate a more frequent laboratory evaluation. More frequent serum IgG levels may be necessary for patients on home therapy, especially for those receiving subcutaneous immunoglobulin (SCIG) products to evaluate compliance.

In patients with active infection, especially patients with CVID, the initial (first) dose can be halved (i.e., 200 to 300 mg/kg) and the dose repeated 2 weeks later to achieve a full dose. To minimize cost and inconvenience, self-administration and home treatment have been used successfully. For home therapy, patients need to be selected carefully. Infusions should be done only in the presence of a responsible adult who is ready to provide assistance. Home treatment has been reported to be as effective as hospital treatment in terms of the frequency of infections, days missed from school or work, antibiotics used, and Ig level achieved. Patients receiving home treatment should be seen regularly (i.e., approximately every 6 months) to monitor clinical status, liver function, and serum IgG level.

The first description on the use of the subcutaneous (SC) route for Ig replacement therapy was published in 1980. It was reported to be safe, well tolerated, and effective in achieving adequate serum IgG levels. In a multicenter study of patients with hypogammaglobulinemia receiving SC infusions at home, a significant reduction in adverse systemic reactions with SC administratiown, compared with the intramuscular (IM) or IV administration, was observed. Anaphylactoid reactions did not occur. Studies have shown enhancements in quality-of-life measurements.[3] Local tissue reactions occur with SCIG infusions, including swelling, soreness, redness, induration, local heat, itching, and bruising. There are advantages and disadvantages for each of the routes of administration of replacement Ig therapy.

SCIG is a suitable alternative to IVIG and may present certain opportunities for optimizing care for patients with PIDD at home. SCIG is an excellent alternative to IVIG for those patients unable to tolerate the IV route of Ig replacement due to systemic side effects with IVIG administration. SCIG also leads to higher serum IgG steady-state levels and may eliminate the wear-off effects of IVIG replacement therapy.

KEY CONCEPTS

Treatment of Patients With Primary Immune Deficiency With Intravenous Immunoglobulin

- Starting dose 400–600 mg/kg every 4 weeks.
- Maintain a serum IgG trough level or steady-state serum IgG level that keeps the patient free of infections (i.e., biological trough level).
- Recent studies indicate that most patients may require a trough or steady-state level of 750–850 mg/dL.
- Higher trough levels (>800 mg/dL) may be necessary to prevent chronic pulmonary disease and enteroviral meningoencephalitis.
- Subcutaneous immunoglobulin (SCIG) offers an advantage in achieving higher steady-state plasma IgG levels.
- It may take ≥3 months to achieve a steady state after a change in dose.
- Monitor serum blood urea nitrogen (BUN) and creatinine and perform liver function tests every 12 months.
- Keep a log of dose, manufacturer, lot number, and reactions for each infusion. Reactions from IVIG are commonly due to rate of infusion, change of product, concurrent infections or inflammation or other preexisting conditions (migraine headaches).
- For patients with rate-related adverse side effects consider pretreatment with:
  - Acetaminophen
  - Diphenhydramine
  - Nonsteroidal antiinflammatory drugs (NSAIDs)
  - Corticosteroids
- Consider alternate administration of SCIG in patients with frequent systemic adverse reactions to IVIG.

CLINICAL PEARLS

Comparison of Routes of Administration of Immunoglobulin Therapy in Patients With Primary Immune Deficiency

Intravenous Route

Advantages

- Can achieve rapid plasma levels.
- Can use this route in patients with bleeding disorders.
- Can give large volumes per infusion allowing intermittent dosing (every 21–28 days).

Disadvantages

- Requires venous access and trained personnel in most situations.
- Interrupts patient's schedule for 3–5-h period.
- Often needs patient to come to a hospital or infusion center for treatment.
- May have more frequent systemic side effects in some patients than subcutaneous (SC) route.
- Large shift in immunoglobulin (Ig) levels during dosing may cause adverse effects at or just after peak and during low troughs, e.g., "wear-off" effects.

Subcutaneous Route

Advantages

- Avoids need for Intravenous (IV) access for patients with poor venous access.
- Eliminates trough levels
- Achieves a stable serum level of IgG
- May eliminate 3rd–4th week fatigue ("wear off" effect) prior to next infusion.
- Has fewer systemic adverse effects than IV route.
- Provides more flexibility for patient's (and/or parent's) schedule
- Convenient for those living at a distance from infusion center or hospital
- Helpful for young adults going to college.

Disadvantages

- Requires frequent dosing due to relatively small volume per infusion.
- Has common minor local reactions at the site of infusion.
- Depends on patient reliability.
- May require multiple infusion sites.
- Some equipment is complex, with need for an infusion pump, sterile technique.
- Loss of dexterity in the elderly may make it difficult to set up equipment and/or draw Ig out of bottles.

SCIG products are 10%, 16.5%, or 20% formulations, the former being similar in composition to IVIG products. The calculated dose for SC administration is generally 100 to 150 mg/kg weekly. Depending on the weight or body mass of the patient and the concentration of the SCIG (i.e., 10% or 20%), infusions may have to be given more frequently than every 7 days. Pharmacokinetics studies in clinical trials have suggested that upward adjustments by 37% of the IVIG dose may be needed to achieve comparable bioavailability, defined as the area under the serum concentration curve (AUC). This adjustment in dosing between IVIG and SCIG has not been mandated by European regulatory agencies, and in the United States some clinical immunologists select dosing based on optimization for prevention of infections, as noted above for the IVIG dosing. The SCIG schedule should be started 1 week after the last dose of IVIG, or in a new patient loaded with four or six doses of SCIG. Before home treatment, patients need to be instructed on the correct technique, observed to apply it under close supervision, and educated about the recognition of possible side effects. SCIG infusion is safer, better tolerated, and preferred by some patients. More detailed information about SCIG administration has been published by Wasserman.[4] Studies in Europe and Canada have shown that healthcare costs of SCIG therapy are lower compared with IVIG. It should be considered as an alternative, especially in those patients with systemic adverse reactions from IV administration. Another method for the administration of Ig by the subcutaneous route has been approved and has been referred to as *facilitated SC infusion*.[5] In this approach, human recombinant hyaluronidase is used to enhance the volume of 10% liquid Ig that can be infused into the SC space, allowing for monthly doses of Ig to be administered in one or two sites. The hyaluronidase depolymerizes the hyaluronan temporarily, allowing the Ig greater access to the lymphatics of the SC space and facilitating the absorption of Ig. This method of administration of 10% Ig has been shown in pharmacokinetic studies to result in bioavailability of 93% of the IVIG dose and thus does not require an upward adjustment factor in calculating the dose of Ig replacement.

ADVERSE EVENTS ASSOCIATED WITH INTRAVENOUS IMMUNOGLOBULIN THERAPY

Rate-Related Adverse Events

Typical rate-related adverse reactions with IVIG include tachycardia, dyspnea, chest tightness, back pain, arthralgia, myalgia, hypertension or hypotension, headache, pruritus, rash, and low-grade fever. Mild to moderate reactions occur in 5% to 15% of infusions; severe reactions occur in less than 1% of patients. Reactions may be delayed and occur several hours after the infusion. Of course, in patients with autoimmune disorders, reaction rates are higher with higher IgG doses. Patients with more profound immunodeficiency or patients with active infections also tend to have more severe reactions. Other factors that contribute to adverse reactions include change of IVIG products, concomitant infections, higher concentrations or lyophilized products, and rapid infusion rates (reviewed in the report by Stiehm).[6] The cause of the reactions is thought to be related to the anticomplementary activity of IgG aggregates in the IVIG in which immune complexes form between infused antibodies and antigens of infectious agents in the patient. The other possible mechanism is that the formation of oligomeric or polymeric IgG complexes that interact with Fc receptors and triggers the release of inflammatory mediators. These rate-related reactions occur less frequently with the newer IVIG products that are liquid and iso-osmolar. Headaches are the most frequent symptom associated with IVIG infusions, occurring in 5% to 20% of infusions and one-third of patients. Slowing the infusion rate or discontinuing therapy until symptoms subside may diminish the reaction. Pretreatment with nonsteroidal antiinflammatory drugs (NSAIDs), acetaminophen (10–15 mg/kg/dose in a child), diphenhydramine (1 mg/kg/dose in a child; 25–50 mg 12 years and older), and/or hydrocortisone (6 mg/kg/dose, maximum 100 mg) 1 hour before the infusion may prevent adverse reactions. Oral hydration prior to the infusion is often helpful. Switching IgG products may also lead to adverse events in 15% to 18% of patients and is discouraged. Persistence of side effects with IVIG infusions despite rate changes, premedication, or even product changes should be a consideration for switching to SCIG.

Central Nervous System–Related Adverse Events

Aseptic meningitis has been reported as one of the complications of IVIG, especially with large doses and rapid infusions, and in the treatment of patients with autoimmune disease.

CLINICAL PEARLS

Adverse Events Associated With Immunoglobulin Therapy

Rate-Related
- Infusion site erythema, swelling, pain, itching
- Headache
- Myalgia, back pain, arthralgia
- Malaise, fatigue
- Chills, fever
- Pruritus
- Rash, urticaria
- Nausea, vomiting
- Tachycardia
- Dyspnea, chest pain, or tightness
- Hypotension
- Hypertension

Central Nervous System
- Severe headaches
- Trigger migraine headaches
- Aseptic meningitis*

Renal
- Azotemia
- Renal failure

Thromboembolic Events[a]
- Thrombosis, cerebral infarction
- Myocardial infarction
- Pulmonary embolism

Anaphylaxis From Anti-Immunoglobulin E Antibodies to Immunoglobulin A

Other (Isolated Reports)
- Cardiac rhythm abnormalities
- Coagulopathy
- Serum sickness
- Hemolysis, related to alloantibodies to blood type A/B
- Cryoglobulinemia
- Neutropenia
- Alopecia
- Uveitis
- Noninfectious hepatitis
- Progressive neurodegeneration

*See text for predisposing risk factors

[a]See Clinical Pearls Risk Factors for Adverse Events box.

Interestingly, this rarely occurs in subjects with immunodeficiencies. Symptoms, including headache, stiff neck, and photophobia, usually develop within 24 hours after completion of the infusion and may last 3 to 5 days. Spinal fluid pleocytosis occurs in most patients. Long-term complications are minimal. The etiology of aseptic meningitis is unclear, but migraine headache has been reported as a risk factor and may be associated with recurrence despite the use of different IVIG preparations and slower rates of infusion.

Renal Adverse Events

Acute renal failure is a rare but a significant complication of IVIG treatment and has been associated in the past with products containing sucrose. Histopathological findings of acute tubular necrosis, vacuolar degeneration, and osmotic nephrosis were suggestive of osmotic injury to the proximal renal tubules. Most patients (95%) had received large doses for treatment of autoimmune diseases. The majority of the cases were treated successfully with conservative therapy, but deaths were reported in 17 patients, all of whom had serious underlying conditions. Most cases of this adverse event were associated with IVIG products containing sucrose as a stabilizer. Risk factors for this adverse reaction include pre-existing renal insufficiency, diabetes mellitus, dehydration, patient age less than 65 years, sepsis, paraproteinemia, and concomitant use of nephrotoxic agents. Newer IVIG products are using alternative stabilizers (e.g., amino acids) instead of sucrose; currently there are no Ig products in the United States that contain sucrose.

CLINICAL PEARLS

Risk Factors for Adverse Events

- Infusion issues:
  - Prior history of an infusion reaction with an immunoglobulin (Ig) product
  - First infusion in a patient with active infection or inflammation
  - Changing Ig products
  - Rapid infusion and/or large dose
- Patient factors:
  - Preexisting renal insufficiency
  - Prior history of thrombotic event
  - Autoimmune disorder
  - Diabetes mellitus
  - Older age
  - Hyperlipidemia or elevated cholesterol
  - Dehydration with volume depletion
  - Hypercoagulable state
  - Indwelling catheters
  - Paraproteinemia or other causes of hyper-viscosity
  - Cardiac or peripheral vascular disorders
  - Estrogen use
  - Smoking

Thromboembolic Events

This adverse effect was observed mainly in patients receiving large doses of IVIG for autoimmune diseases, although there had been reported cases in patients with PIDD. Patients with elevated serum viscosity (e.g., cryoglobulinemia, hypergammaglobulinemia, and hypercholesterolemia) are at risk for developing a critical increase in serum viscosity with IVIG, especially high doses that predispose them to thromboembolic events, such as myocardial infarction, stroke, deep vein thrombosis, or pulmonary embolism. Recently, a contaminating procoagulant factor (e.g., factor XIa) has been implicated in these thromboembolic events. Patients at risk are older (>65 years), on multiple drugs, and have comorbidities, such as diabetes or hypertension. Patients at risk should be well hydrated, IVIG should be administered at lower rates, and products with low sodium and an osmolality in the physiological range should be selected. Ig manufacturers have taken steps to remove or reduce procoagulant activity from their products.

Transfusion Reaction Caused by Antibodies Against Immunoglobulin A

True anaphylaxis is rare in patients with selective IgA deficiency and CVID who develop IgE antibodies to IgA after treatment with Ig; this adverse event appears to be much less frequent than originally thought. In various studies 10% to 22% of patients with CVID have IgG antibodies to IgA, but there is no correlation between the presence of these antibodies and adverse reactions. Patients with anti-IgA antibodies who have had reactions to IVIG have tolerated SCIG.[7]

Hemolysis Caused by Isoagglutinins (Anti-A and Anti-B Antibodies) in the Immunoglobulin Products

Hemolysis following the administration of Ig products is uncommon and occurs in approximately 1.6% of patients and more commonly in patients receiving large doses of Ig, e.g. autoimmune diseases. It usually does not lead to clinical adverse effects; however, if extensive, it can cause renal failure. IVIG infusion–associated hemolysis occurs from 12 hours to 10 days after the first infusion. Patients with blood type A or AB are particularly at risk. Labs include Coombs' positivity, decreases in the hemoglobin and haptoglobin levels, increase in lactate dehydrogenase, and increases in bilirubin and reticulocyte counts.[8] Currently, isoagglutinins levels in Ig products have a limit of ≤1:64 by regulatory agencies. Additional manufacturing processes have been added by some manufacturers to reduce the level of isoagglutinins in their Ig products.

Other Adverse Reactions

A number of other adverse reactions have been reported in association with IVIG infusions, as noted in Clinical Pearls 2. These side effects are usually less common and are discussed in more detail elsewhere.[6,8]

Summary: Immunoglobulin Replacement in Treatment of Immune Deficiency

Ig replacement is the mainstay of treatment for primary humoral immune deficiency. The goal of treatment is to provide a broad spectrum of antibodies to prevent infections and chronic long-term comorbid conditions such as chronic lung disease. The usual dose is 400 to 600 mg/kg/month, but this may vary among individuals, and some patients may require higher doses. A serum trough level above 500 mg/dL has been shown to be effective in the prevention of severe bacterial infections. However, recent studies have suggested that even higher doses (e.g., a "biological" trough level and achieving trough levels of IgG in the range of 750 to 900 mg/dL) may be desirable.[9] Ig can be given intramuscularly, intravenously, or subcutaneously. SCIG administration has been proven to be safe and is a good choice in some patients, especially those experiencing side effects of administration by the IV route. Generally, Ig replacement therapy is considered safe in the majority of patients. Side effects are usually mild and treatable with premedication. Good manufacturing practices, improved screening of plasma donors, testing of the source plasma, and additional viral inactivation and removal steps have made Ig a better and safer plasma-derived product. To optimize the care of patients with PIDDs, the Immune Deficiency Foundation has a helpful chart to facilitate the discussion between the patient and the healthcare provider (Table 82.2).

TABLE 82.2 Immunoglobulin Replacement Treatment Options

	Intravenous Immunoglobulin	Subcutaneous Immunoglobulin	Hyaluronidase-Facilitated Immunoglobulin
Who?	Indicated for adult and pediatric patients with PI	Indicated for adult and pediatric patients with PI	Indicated for adult patients with PI
How?	Usually administered by a nurse	Self-administered	Either self-administered or given by a nurse
Where does it go?	Infused directly into the bloodstream through a vein	Infused or injected under the skin into the subcutaneous tissues of the arms, belly, outer buttock or the thighs	Infused under the skin into the subcutaneous tissues of the belly, outer buttock or the thighs
When?	Usually given every 3–4 weeks	Can be given on a flexible schedule from daily to every 2 weeks	Can be given every 3–4 weeks
How long?	Can take 2–6 h to infuse	Can take 5 min to 2 h to infuse or inject	Can take 1–2 h to infuse
Where is it given?	Can be infused at home, in a hospital or an outpatient infusion center depending on insurance and patient preference	Usually administered in a home setting after the patient is trained to be independent	Can be infused at home or in an outpatient infusion center depending on insurance and patient preference
Side effects?	Patients can have side effects that are often related to the rate of infusion and can be treated and prevented with other medications, given before or after the treatment	Skin can be red and irritated at the site of injections. This often improves with each injection	Skin can be red and irritated at the site of injections. This often improves with each injection. The volume per injection is larger than standard subcutaneous (under the skin) injection, so the volume is more visible under the skin, and may take 48–72 h to totally absorb

PI, primary immunodeficiency. Epland, K., Perez, E.E. IDF *Guide to Ig Therapy—Immunoglobulin Replacement Therapy for People Living with Primary Immunodeficiency Diseases*. 4th ed. Immune Deficiency Foundation, 2018.

MECHANISMS OF ACTION OF IMMUNOGLOBULIN THERAPY IN AUTOIMMUNE AND INFLAMMATORY DISEASES

Although we still do not understand *all* the mechanisms by which IVIG has immunomodulatory effects, knowledge of the actions of IVIG in these diseases will allow for definition of appropriate indications and schedules of administration of IVIG and the design of a new generation of IVIG products that are better able to target the immune perturbations in autoimmune and inflammatory processes. Furthermore, additional multicenter placebo-controlled clinical trials are needed to confirm clinical efficacy. A more detailed discussion of the evidence-based treatment of autoimmune and inflammatory disorders can be found elsewhere.[1] This chapter does not review the indications for IVIG therapy in these disorders; the reader is directed to other works, including comprehensive reviews and practice-based guidelines, for more details about the various indications for IVIG therapy in autoimmune and inflammatory

disorders. This section adresses the possible mechanisms of actions of Ig therapy; more than one mechanism may be important in the immune modulation of these autoimmune disorders.

KEY CONCEPTS

Mechanisms of Action of Intravenous Immunoglobulin in Autoimmune and Inflammatory Diseases

- Blockade of Fc receptors on macrophages of the reticuloendothelial system of liver and spleen.
- Restoration of idiotypic/anti-idiotypic network.
- Suppression or neutralization of cytokines by specific antibodies in the intravenous immunoglobulin (IVIG); augmenting immune modulating cytokines, e.g. interleukin (IL)-1RA, IL-10.
- Block the binding of adhesion molecules on leukocytes to vascular endothelium.
- Inhibition of complement deposition and formation of the C5b-C9 membrane-attack complex on endomysial capillaries in dermatomyositis, and interaction with effector fragments C3b and C4b of the complement cascade.
- Neutralization of microbial toxins (superantigens).
- Inhibit Fas-mediated keratinocyte apoptosis by anti-Fas antibodies (toxic epidermal necrolysis).
- Induction of apoptosis with anti-Fas antibodies at high concentrations of IVIG.
- Neutrophil apoptosis by anti-Siglec-9 antibodies in IVIG.
- IVIG saturation of FcRn receptors to accelerate degradation of IgG autoantibodies.
- Induction of inhibitory FcγRIIb receptors on effector macrophages by the binding of sialylated IgG to DC-SIGN on antigen-presenting cells via IL-33 induction of IL-4 by basophils.
- B-cell modulation through the IL-22 receptor, FcγRIIB receptor, and neuralization of BAFF cytokine (e.g., B cell–activating factor [BAFF]).
- Inhibition of T-cell proliferative responses.
- Expansion and/or activation of a population of regulatory T cells (Tregs) by enhancement of the cyclooxygenase 2 (COX-2) pathway and increasing prostaglandin E_2 (PGE_2) from dendritic cells (DCs).
- Inhibition of leukocyte recruitment into inflammatory tissues by inhibiting selectin and integrin binding.
- Downregulation of the T-helper 17 (Th17) pathway.
- Modulation of DC function through C-type lectin receptors (DCLRs): the DC-specific intercellular adhesion molecule-grabbing nonintegrin (DC-SIGN; spleen) and DCLR (lung and lymph nodes).

Blockade of Fc Receptors of the Reticuloendothelial System

ITP results from accelerated platelet destruction attributable to an immunological process resulting in bleeding that may be life threatening. Studies have shown that IVIG leads to the rapid reversal of the thrombocytopenia in ITP. Autoantibody-opsonized platelets are destroyed in the spleen and the liver by FcγR-mediated phagocytic clearance. One of the first mechanisms proposed was that IVIG blocks the FcγRs on monocytes, macrophages, and the cells of the reticuloendothelial systems to reduce immune mediated destruction of antibody-coated platelets.

Interactions of Idiotype and Anti-Idiotype as Immune Modulation

Idiotype–anti-idiotype interactions between antiplatelet glycoprotein IIb (GPIIb)/IIIa autoantibodies and IVIG may be another mechanism by which IVIG could affect autoantibody production and effector function in ITP and may be playing a role in the long-term immune effects in ITP. IVIG contains anti-idiotypic antibodies to a number of disease-associated autoantibodies including anti-DNA, anti-factor VIII, antineutrophil cytoplasmic antibody (ANCA) autoantibodies, antithyroid autoantibodies, acetylcholine receptor, and others. A proposed mechanism of action of IVIG in ANCA-positive vasculitis is binding or neutralization of the ANCA autoantibodies by anti-idiotypic antibodies in IVIG. Studies have shown that IVIG contains antibodies with idiotypic specificities that can bind and neutralize potentially pathogenic autoantibodies in autoimmune neurological diseases, such as Guillain-Barré syndrome (GBS), chronic inflammatory demyelinating polyneuropathy (CIDP), and myasthenia gravis (MG). Anti-idiotypic antibodies in IVIG directed against idiotopes on the anti-GM_1 Ig molecule block the binding of the anti-GM_1 antibodies to its target antigen. Support for a similar mechanism in acquired factor VIII deficiency comes from the fact that IgG or $F(ab')_2$ fragments in IVIG preparations are capable of binding to autoantibodies to factor VIII. Thus the anti-idiotypic antibodies in IVIG may be beneficial by restoring the idiotypic control network in these autoimmune diseases.

The Role of the FcRn Receptor on Immune Modulation

The FcRn receptor (neonatal Fc receptor) was identified as the receptor responsible for protecting IgG from catabolism in the endocytic vesicles of the endosome, and this explains the relatively long half-life of this plasma protein. IVIG may accelerate the catabolism of IgG autoantibodies by saturating these protective receptors in direct proportion to the relative concentration of exogenous plasma levels of IgG from IVIG.[10] In a rat model of immune thrombocytopenia IVIG enhanced the clearance of antiplatelet antibodies in a dose-dependent manner by saturation of the FcRn receptor. Approximately 50% of the overall effect of IVIG in ITP may be attributed to IVIG-mediated acceleration of the elimination of antiplatelet antibodies by the FcRn saturation mechanism. In FcRn knock-out mice, IVIG failed to increase the clearance of antiplatelet antibodies. Although very little data are available in human autoimmune disease to support this mechanism, recent developments of monoclonal antibodies to the FcRn receptor and high-affinity FcRn binding engineered Fc fragments are now in clinical trials.[11]

KEY CONCEPTS

Proposed Mechanisms of Action of Intravenous Immunoglobulin in Autoimmune Idiopathic Thrombocytopenic Purpura

- Fc receptor blockade of reticuloendothelial system.
- Fcγ receptor downregulation.
- Idiotype–anti-idiotype interaction between antiplatelet GPIIb/IIa autoantibodies and the anti-idiotypic antibodies in intravenous immunoglobulin.
- Activation of inhibitory receptor FcγRIIB.
- Saturation of FcRn receptor to accelerate the catabolism of antiplatelet autoantibodies.

Modulation of Immunoregulatory Function Through the Fc Receptor

The FcγRIIB receptor provides an inhibitory signal to cells through a pathway mediated by an immunoregulatory tyrosine-based inhibition motif (ITIM). Studies have shown that IVIG stimulates these inhibitory FcγRIIB receptors found on a variety of cell types,

including macrophages, B cells, and a subpopulation of T cells. Samuelsson et al.[12] using a murine model of ITP, showed that the administration of IVIG prevented platelet destruction by a pathogenic monoclonal autoantibody. Protection was associated with the induction of the expression of FcγRIIB receptors on splenic macrophages. This inhibitory FcγRIIB receptor was required for the protection of the animals against the monoclonal autoantibody, since disruption of the receptor by either genetic deletion or a blocking monoclonal antibody (mAb) reversed the therapeutic effects of IVIG. Kaneko et al.[13] showed that the inhibitory properties of IVIG were linked to the sialylation of the glycan component of the Fc fragment. The important glycan moiety in the IgG molecule is located at the asparagine (Asn^{297}) site in the second domain of the constant region. The sialylated fraction accounts for approximately 10% of the total IgG. Using a K/BxN serum-induced arthritis model in mice, Kaneko et al.[13] showed that IVIG at 1 g/kg inhibited the inflammatory arthritic process. De-glycosylated or neuraminidase-treated IVIGs were unable to inhibit this inflammation. IVIG enriched for the sialylated glycan moiety had comparable inhibitory effects on the inflammatory process at only 10% of the dose used with intact IVIG. This inhibitory activity was dependent on FcγRIIB expression on effector macrophages. An engineered recombinant/sialylated human IgG1 Fc protein had a 35-fold enhanced immune-modulating activity compared with native IVIG. A splenic marginal zone macrophage expressing the C-type lectin receptor (i.e., SIGN-R1) was required for the anti-inflammatory activity of IVIG in concert with its ability to bind to sialylated Fc domains. The interaction between sialylated IgG Fc and SIGN-R1 led to the upregulation of the inhibitory FcγRIIB receptor on effector cells. DC-SIGN (dendritic cell [DC]-specific intercellular adhesion molecule-grabbing nonintegrin), the human orthologue of SIGN-R1, may have a comparable role in the anti-inflammatory effects of IgG Fc fragment on human macrophages and DCs. Anthony et al.[14] presented data in their murine model that this immune-modulating pathway may be mediated by the production of interleukin-33 (IL-33) by DCs, which, in turn, produces IL-4 from basophils, leading to the increased expression of the FcγRIIB receptor on effector macrophages. Others, however, have questioned whether this pathway of immune modulation is important in humans.[15,16] Studies by von Maddur et al.[17] showed that basophil expansion did not occur with IVIG treatment of patients with autoimmune disease, and that IVIG did not induce IL-33 production by DC-SIGN$^+$ APC in humans. Furthermore, the effects of IVIG on the activations of DC-SIGN$^+$ human cells were independent of IgG-Fc sialylation.[15] Nevertheless, these exciting studies define, at least in mice, an important mechanism by which IVIG modulates immune processes mediated through sialylated Fc on the IgG molecule and the receptors on DCs and effector macrophages (i.e., SIGN-R1 and FcγRIIB) involved in this antiinflammatory process.

Neutralizing Antibody Activity in Intravenous Immunoglobulin Against Bacterial Toxins

Kawasaki syndrome (KS), an acute multisystem disease of unknown etiology, primarily affects infants and young children. Although KS occurs worldwide in children of all racial groups, it is most prevalent in Japan and in children of Japanese ancestry. Although the acute illness is generally self-limiting, coronary artery abnormalities related to a generalized inflammation and immune activation of small and medium-sized blood vessels develop in up to 25% of untreated patients. Although the etiology remains unknown, the clinical features and laboratory findings suggest an infectious or post-infectious process. The administration of high-dose IVIG, together with aspirin, is the standard of care in the treatment of KS.

KS is associated with marked activation of T cells, monocytes, and macrophages. On the basis of immunological and clinical features of KS overlapping with those of bacterial toxic shock-like syndrome, studies were carried out to determine if KS is associated with exposure to a superantigen, such as a bacterial toxin. Acute KS is associated with marked immune activation and increased circulating cytokine levels. Some of these cytokines elicit proinflammatory and prothrombotic responses by inducing the expression of leukocyte adhesion molecules, which localize inflammatory cells to vascular endothelial cells. The expression of endothelial–leukocyte adhesion molecules has been demonstrated in acute KS, and its downregulation correlates with favorable response to IVIG treatment. The magnitude and persistence of proinflammatory cytokine synthesis have been reported to constitute a risk for the development of coronary artery abnormalities.

IVIG has been shown to contain high titers of specific antibodies inhibitory to the activation of T cells by staphylococcal and streptococcal superantigens. These findings may account for the observation that treatment of acute KS with IVIG results in a marked reduction of macrophage and T-cell activation. In this regard, the efficacy of IVIG in suppressing the immune activation associated with KS and, more importantly, its ability to prevent the development of coronary artery abnormalities in this illness may relate to the neutralizing antibody activity of IVIG against these bacterial toxins. Toxin neutralization is unlikely to be the only beneficial effect of IVIG in KS. Blocking or neutralizing cytokines by the anti-cytokine antibodies in IVIG may modulate the local inflammatory responses of blood vessels in KS by modifying leukocyte adhesion after increasing the expression of cell-surface determinants on vascular endothelial cells.

IVIG may have therapeutic value in the treatment of patients with toxic shock syndrome secondary to *Staphylococcus aureus* or *Streptococcus pyogenes* exotoxins. In an open study by the Canadian Streptococcal Study Group IVIG appeared to be beneficial in patients with streptococcal toxic shock syndrome. In a meta-analysis of IVIG treatment of neonatal sepsis, there was a six-fold decrease in mortality. IVIG inhibits *Staphylococcus* exotoxin-induced T-cell activation and contains antibodies against exotoxins responsible for toxic shock syndrome. Great variations in neutralizing activity against streptococcal pyrogenic exotoxins can be found in different brands and even among different lots of IVIG. However, these findings suggest that it is possible to select an IVIG preparation that contains high levels of neutralizing activity against a wide variety of group A streptococcal superantigens, which could be used in the treatment of streptococcal toxic shock syndrome. The neutralizing capacity of IVIG against these bacterial superantigens is important because of their potential to stimulate the production of proinflammatory cytokines that lead to clinical disease. A number of in vitro studies have shown that IVIG can inhibit the production of, or bind to and neutralize, a number of cytokines and growth factors from various cell types.[18,19] Thus IVIG may exert its antiinflammatory effects in many different types of inflammatory diseases by interrupting or modifying a number of different steps in the inflammatory cascade, from the inhibition of effector cell function to reduction in cytokine-induced endothelial cell activation, or the "neutralization" of proinflammatory cytokines.

Modulation of Adhesion Molecules on Endothelial Cells and Antibodies in Intravenous Immunoglobulin to Cell-Surface Receptors

IVIG contains a number of natural autoantibodies that may have immune-modulating activities (e.g., antibodies to CD4, major histocompatibility complex [MHC] class I molecules, cytokines, adhesion molecules, and Siglec molecules on leukocytes and other cell-surface molecules). The "natural" antibodies in IVIG have also been shown to bind to a number of plasma and tissue proteins, including B cell–activating factor (BAFF), granulocyte macrophage–colony-stimulating factor (GM-CSF), liver antigens, and beta-amyloid peptide.[19] The binding of native IgG can be significantly increased by mild denaturing conditions (e.g., mild pH treatment and cold ethanol precipitation used during the manufacturing of IVIG).

IVIG contains specific antibodies to a 10-peptide sequence, including the RGD (Arg-Gly-ASP) motif that is expressed in adhesion molecules on a variety of cell surfaces. Most integrins bind to this RGD sequence. IVIG can inhibit leukocyte recruitment into inflammatory tissues by inhibiting selectin and integrin binding. In a mouse model of sickle cell vaso-occlusive crisis, IVIG was shown to inhibit neutrophil adhesion to the vascular endothelium, resulting in an increase in capillary blood flow and reversal of the vessel occlusion.[20] IVIG could modulate this cytokine-mediated endothelial cell activation by neutralizing the effects of the cytokines, inhibiting endothelial cell responses to the cytokines, or inhibiting the production of cytokines and growth factors. These mechanisms of IVIG may be playing an important role in preventing coronary artery abnormalities in patients with KS.

Toxic epidermal necrolysis (TEN) and Stevens-Johnson syndrome (SJS) are severe drug-induced skin diseases. TEN results in apoptotic epidermal cell death, in which there is separation of large areas of the skin at the epidermal junction, producing the appearance of scalded skin. Keratinocyte apoptosis that precedes epidermal detachment is an early event in TEN. A number of drugs, including sulfonamides, anticonvulsants, and NSAIDs, can cause TEN and SJS. The mortality rate can be as high as 30%. Viard et al.[21] studied serum samples from patients with TEN and found that the sera of these patients had very high levels of soluble Fas ligand (sFasL). Keratinocytes normally express the death receptor Fas. The keratinocytes of patients with TEN also express very high levels of active Fas ligand. In in vitro studies, IVIG completely inhibited Fas-mediated keratinocyte apoptosis. This effect was related to the presence of naturally occurring Fas-blocking antibodies in IVIG, which inhibit Fas-mediated keratinocyte cell death. In a large multicenter retrospective study, early infusion of high-dose IVIG (mean total dose 2.7 g/kg) was well tolerated and effective in improving the survival of patients with TEN.

Modulation of Complement Effector Function

The principal inflammatory mechanism in dermatomyositis (DM) is complement (C)-dependent microangiopathy with activation of C3 and deposition of the complement C5b-9 membrane attack complex (MAC) on the endomysial capillaries. In patients with DM who were treated with IVIG, C3 deposition was reduced with corresponding decreases in complement expression on endomysial capillaries. IVIG prevents the uptake of complement components and formation of the MAC on the endomysial capillaries in the muscle tissues of patients with DM. Consequently, IVIG allowed neovascularization to occur with reversal of the ischemic process, resulting in muscle tissue healing. This effect of IVIG on complement deposition may be relevant to other autoimmune neurological diseases, such MG, GBS, and CIDP, in which a complement may be playing a role in the tissue damage.[22] IVIG may also protect the brain against acute ischemic injury mediated by a complement.[23]

Effects of Immunoglobulin on the Regulatory T-Cell Pathways

In a mouse model of multiple sclerosis (MS), IVIG has been shown to expand and enhance the function of $FoxP3^+$ regulatory T cells (Tregs) while inhibiting the differentiation of T-helper 17 (Th17) and Th1 cells. This protective effect of IVIG was lost in mice that were depleted of Tregs.[24] These changes were independent of FcγRIIB and the Fc domain, since $F(ab')_2$ fragments led to similar changes in Th17 cells, Tregs, and clinical efficacy in this experimental allergic encephalomyelitis (EAE) model. De-sialylated IVIG had the same immune-modulating effects as "native" IVIG.[25] Investigations by Trinath et al.[26] suggested that the mechanism by which IVIG induces Tregs was the enhancement of the cyclo-oxygenase 2 (COX-2) pathway via increased expression of prostaglandin E_2 (PGE_2) from human DCs.

IVIG therapy in an ovalbumin (OVA)-sensitized mouse model of asthma markedly attenuates lung inflammation, decreased bronchial hyperresponsiveness to methacholine, and suppressed the Th2 pathway.[27] The draining pulmonary lymph nodes of IVIG-treated mice showed a significant increase in $CD4^+CD25^+FoxP3^+$ Treg cells. IVIG-primed DCs on adoptive transfer to OVA-sensitized and challenged mice abrogated airway hyperresponsiveness and induced Tregs cells. In their model system, Massoud et al.[28] reported that sialylated IgG bound to a novel C-type lectin receptor (i.e., dendritic cell immunoreceptor [DCIR]) led to the induction of Tregs cells. Thus, a number of studies have demonstrated the importance of the induction of $FoxP3^+$ Tregs by IVIG in modulating the autoimmune/anti-inflammatory process. In contrast, Ig therapy in this murine model downregulated the Th-17 pathway. The effects of IVIG on Th17 cells are not due to neutralization of the Th17 cytokines but mediated through STAT3.[29] In patients with KS and GBS, clinical improvement with IVIG therapy correlated with increased number and function of Tregs.[30] B cells also express the inhibitory FcγRIIB receptor. Of note, observations in patients with CIDP treated with high-dose IVIG demonstrated upregulation of the FcγRIIB receptor on peripheral blood B cells and monocytes.[31] IVIG also inhibits IL4/CD40-, Toll-like receptor (TLR)-, and B-cell receptor (BCR)-mediated activation of B cells. IVIG contains antibodies to BAFF, an important regulatory cytokine for B-cell survival. IVIG treatment of patients with CIDP decreased elevated serum levels of BAFF.

Effects of Intravenous Immunoglobulin on B Cells

The effects of IVIG on B cells is mainly mediated through interaction with the CD22 receptor, a receptor belonging to the Siglec superfamily. The CD22 receptor is important in modulating signaling thresholds for B-cell activation. Sialylated IgG binds to CD22 on B cells to affect B-cell inhibition through several BCR signaling pathways.

Summary: Intravenous Immunoglobulin in Treatment of Autoimmune and Inflammatory Diseases

Clearly, IVIG (IgG) has a number of immune-modulating effects and has been found to be an effective treatment for a wide spectrum of autoimmune and inflammatory diseases.[19] At present, IVIG is FDA approved for only a few autoimmune and inflammatory diseases. In autoantibody-mediated disease, the Fc domain appears to be the important IgG moiety that leads to immune modulation. The importance of sialylation of the Fc fragment remains controversial, as do the mediators involved (e.g., IL-33 and IL-4). Differences in animal models, IVIG source, route and timing of the administration of the IVIG, mouse strain, and other variables may account for the differences in the experimental observations of laboratories. In T-cell-mediated animal models of disease, such as EAE, there is strong evidence that there is upregulation of Tregs and inhibition of the Th17 pathway. These effects may be mediated by the F(ab′)2 portion of the IgG molecule, and not the Fc domain. Furthermore, there is also controversy over the receptor on antigen-presenting cells (APCs), such as macrophages and DCs, which are involved in the immune-modulating process mediated by IVIG. In certain animal models, the SIGN-R1 (or in humans DC-SIGN) is important, and in other models (e.g., murine asthma), a novel C-type lectin receptor (DCIR) appears to be important. These differences undoubtedly relate to the disease model. Two studies have demonstrated the importance of the PGE_2 pathway on IVIG-induced immune modulation mediated by the F(ab′)2 portion of the IgG molecule. These observations may point to the possibility of alternative treatment regimens that employ the selective increase in PGE_2 in certain autoimmune or inflammatory disorders to increase Tregs and inhibit the production of certain cytokines. Further clarification in human disease models using in vitro human cells is important. These mechanisms are not mutually exclusive, and probably more than one mechanism is playing a role in the efficacy of Ig therapy in an autoimmune disease process. A better understanding of the pathogenic mechanisms involved in these diseases will undoubtedly lead to a more effective therapy with IVIG and more specific, modified forms of this biological product.

ON THE HORIZON

Translational Research Opportunities Related to Immunoglobulin Therapy

- Elucidation of mechanism(s) of action will lead to more precise bioengineered molecular products to treat autoimmune and inflammatory diseases.
- Clarification of the mechanisms of action of intravenous immunoglobulin (IVIG) in autoimmune and inflammatory diseases should lead to better understanding of the pathobiology of these diseases.
- Enhanced product purification with further identification of the minor components (e.g., procoagulant factors, isohemagglutinins) in Ig products that may contribute to adverse reactions should lead to improved manufacturing processes and improved product tolerability and safety for patients.
- Development of innovative approaches to the delivery of Ig products, such as pre-filled syringes, to patients will enhance safety and compliance.

REFERENCES

1. Perez EE, Orange JS, Bonilla F, et al. Update on the use of immunoglobulin in human disease: A review of evidence. *J Allergy Clin Immunol*. 2017;139(3s):S1–S46.
2. Orange JS, Grossman WJ, Navickis RJ, Wilkes MM. Impact of trough IgG on pneumonia incidence in primary immunodeficiency: a meta-analysis of clinical studies. *Clin Immunol*. 2010;137(1):21–30.
3. Peshko D, Kulbachinskaya E, Korsunskiy I, et al. Health-related quality of life in children and adults with primary immunodeficiencies: a systematic review and meta-analysis. *J Allergy Clin Immunol Pract*. 2019;7(6):1929–1957.e5.
4. Wasserman RL. The nuts and bolts of immunoglobulin treatment for antibody deficiency. *J Allergy Clin Immunol Pract*. 2016;4(6):1076–1081.e3.
5. Wasserman RL, Melamed I, Stein MR, et al. Long-term tolerability, safety, and efficacy of recombinant human hyaluronidase-facilitated subcutaneous infusion of human immunoglobulin for primary immunodeficiency. *J Clin Immunol*. 2016;36(6):571–582.
6. Stiehm ER. Adverse effects of human immunoglobulin therapy. *Transfus Med Rev*. 2013;27:171–178.
7. Rachid R, Bonilla F. The role of anti-IgA antibodies in causing adverse reactions to gamma globulin infusion in immunodeficient patients: a comprehensive review of the literature. *J Allergy Clin Immunol*. 2012;129:628–634.
8. Guo Y, Tian X, Wang X, Xiao Z. Adverse effects of immunoglobulin therapy. *Front Immunol*. 2018;9:1299.
9. Lucas M, Lee M, Lortan J, et al. Infection outcomes in patients with common variable immunodeficiency disorders: relationship to immunoglobulin therapy over 22 years. *J Allergy Clin Immunol*. 2010;125(6):1354–1360 e4.
10. Yu Z, Lennon VA. Mechanism of intravenous immune globulin therapy in antibody-mediated autoimmune diseases.[see comment]. *N Engl J Med*. 1999;340(3):227–228.
11. Bayry J, Kaveri SV. Kill 'em all: efgartigimod immunotherapy for autoimmune diseases. *Trends Pharmacol Sci*. 2018;39(11):919–922.
12. Samuelsson A, Towers TL, Ravetch JV. Anti-inflammatory activity of IVIG mediated through the inhibitory Fc receptor. *Science*. 2001;291(5503):484–486.
13. Kaneko Y, Nimmerjahn F, Ravetch JV. Anti-inflammatory activity of immunoglobulin G resulting from Fc sialylation. *Science*. 2006;313(5787):670–673.
14. Anthony RM, Kobayashi T, Wermeling F, Ravetch JV. Intravenous gammaglobulin suppresses inflammation through a novel T(H)2 pathway. *Nature*. 2011;475(7354):110–113.
15. Bayry J, Bansal K, Kazatchkine MD, Kaveri SV. DC-SIGN and alpha2,6-sialylated IgG Fc interaction is dispensable for the anti-inflammatory activity of IVIg on human dendritic cells. *Proc Natl Acad Sci U S A*. 2009;106(9):E24;author reply E5.
16. Yu X, Vasiljevic S, Mitchell DA, et al. Dissecting the molecular mechanism of IVIg therapy: the interaction between serum IgG and DC-SIGN is independent of antibody glycoform or Fc domain. *J Mol Biol*. 2013;425(8):1253–1258.
17. Maddur MS, Stephen-Victor E, Das M, et al. Regulatory T cell frequency, but not plasma IL-33 levels, represents potential immunological biomarker to predict clinical response to intravenous immunoglobulin therapy. *J Neuroinflammation*. 2017;14(1):58.
18. Bayry J, Negi VS, Kaveri SV. Intravenous immunoglobulin therapy in rheumatic diseases. *Nat Rev Rheumatol*. 2011;7(6):349–359.
19. Ballow M. Mechanisms of immune regulation by IVIG. *Curr Opin Allergy & Clin Immunol*. 2014;14(6):509–515.
20. Chang J, Shi PA, Chiang EY, Frenette PS. Intravenous immunoglobulins reverse acute vaso-occlusive crises in sickle cell mice through rapid inhibition of neutrophil adhesion. *Blood*. 2008;111(2):915–923.
21. Viard I, Wehrli P, Bullani R, et al. Inhibition of toxic epidermal necrolysis by blockade of CD95 with human intravenous immunoglobulin. *Science*. 1998;282(5388):490–493.
22. Lunemann JD, Nimmerjahn F, Dalakas MC. Intravenous immunoglobulin in neurology—mode of action and clinical efficacy. *Nat Rev Neurol*. 2015;11(2):80–89.

23. Thom V, Arumugam TV, Magnus T, Gelderblom M. Therapeutic potential of intravenous immunoglobulin in acute brain injury. *Front Immunol*. 2017;8:875.
24. Othy S, Hegde P, Topcu S, et al. Intravenous gammaglobulin inhibits encephalitogenic potential of pathogenic T cells and interferes with their trafficking to the central nervous system, implicating sphingosine-1 phosphate receptor 1-mammalian target of rapamycin axis. *J Immunol*. 2013;190(9):4535–4541.
25. Othy S, Topcu S, Saha C, et al. Sialylation may be dispensable for reciprocal modulation of helper T cells by intravenous immunoglobulin. *Eur J Immunol*. 2014;44(7):2059–2063.
26. Trinath J, Hegde P, Sharma M, et al. Intravenous immunoglobulin expands regulatory T cells via induction of cyclooxygenase-2-dependent prostaglandin E2 in human dendritic cells. *Blood*. 2013;122(8):1419–1427.
27. Massoud AH, Guay J, Shalaby KH, et al. Intravenous immunoglobulin attenuates airway inflammation through induction of forkhead box protein 3-positive regulatory T cells. *J Allergy Clin Immunol*. 2012;129(6) 1656–1665.e3.
28. Massoud AH, Yona M, Xue D, et al. Dendritic cell immunoreceptor: a novel receptor for intravenous immunoglobulin mediates induction of regulatory T cells. *J Allergy Clin Immunol*. 2014;133(3):853–863.e5.
29. Maddur MS, Sharma M, Hegde P, et al. Inhibitory effect of IVIG on IL-17 production by Th17 cells is independent of anti-IL-17 antibodies in the immunoglobulin preparations. *J Clin Immunol*. 2013;33(Suppl 1):S62–S66.
30. Maddur MS, Othy S, Hegde P, et al. Immunomodulation by intravenous immunoglobulin: role of regulatory T cells. *J Clin Immunol*. 2010;30(suppl 1):S4–S8.
31. Tackenberg B, Jelcic I, Baerenwaldt A, et al. Impaired inhibitory Fcgamma receptor IIB expression on B cells in chronic inflammatory demyelinating polyneuropathy. *Proc Natl Acad Sci U S A*. 2009;106(12):4788–4792.

Glucocorticoids

Anthony J. Frew[†] *and David B. Corry*

Glucocorticoids are among the most commonly prescribed drugs and are used for a wide range of medical conditions.[1] More than 60 years after their introduction into clinical practice, they remain the most important and most frequently used class of antiinflammatory drug, and their use continues to increase especially with the introduction of intranasal glucocorticoids that do not require prescription. Community survey data suggest that 1.2% to 3% of the general population and up to 7% of the elderly are taking oral glucocorticoids.[2,3] Between 56% and 68% of patients with rheumatoid arthritis (RA) are more or less continuously treated with glucocorticoids.[4] Although glucocorticoids are relatively inexpensive, total market volume is about US$10 billion per year.[2] Glucocorticoids are widely used because they are the most effective (and cost-effective) antiinflammatory and immunomodulatory drugs available. However, glucocorticoids can cause serious adverse effects, especially when used incorrectly.

MECHANISMS OF ACTION

The way in which glucocorticoids are used in different clinical conditions is essentially empirical, as there is only limited evidence to support current practice in specific clinical settings.[4] In general, glucocorticoid dosages are increased in parallel with the clinical activity and severity of the disease under treatment. The rationale for this approach is that higher dosages increase glucocorticoid receptor saturation in a dose-dependent manner (Table 83.1), which intensifies the therapeutically relevant, *genomic* actions of glucocorticoids. Moreover, with increasing dosages, additional and qualitatively different, *nonspecific, nongenomic* actions of glucocorticoids come into play (see Table 83.1).

KEY CONCEPTS

Characteristics Applying to Genomic Actions

- Physiologically relevant
- Therapeutically effective at all dosages, even very small ones (low-dose therapy).
- Slow; significant changes in the regulator protein concentrations are not seen within less than 30 min because of the time required for cGCR activation/translocation, transcription, and translation effects.
- The GC-induced synthesis of regulator proteins can be prevented by inhibitors of transcription (e.g., actinomycin D) or translation (e.g., cycloheximide).

Genomic Actions of Glucocorticoids

The antiinflammatory and immunomodulatory effects of glucocorticoids (GCs) are mainly mediated by genomic mechanisms (Figs. 83.1 and 83.2). Their lipophilic structure and low molecular mass allow GCs to pass easily through the cell membrane and bind to cytosolic glucocorticoid receptors (cGCRs). Ultimately this either induces the synthesis of regulatory proteins ("*transactivation*") or inhibits their synthesis ("*transrepression*").[5] Between 10 and 100 genes per cell are directly regulated by glucocorticoids but many other genes are regulated indirectly, through interaction with transcription factors and co-activators (see below).[6] It has been estimated that glucocorticoids influence the transcription of approximately 1% of the entire genome.

Structure of the Cytosolic Glucocorticoid Receptor

The unactivated cGCR (cGCRα) is a 94 kDa protein held in the cytoplasm as a multiprotein complex, consisting of several heat shock proteins (hsp), including hsp90, hsp70, hsp56, and hsp40 (chaperones) (Fig. 83.3). The cGCR interacts with immunophilins, p23. and several kinases of the mitogen-activated protein kinase (MAPK) signaling system, including Src, which also act as molecular (co)chaperones (see Figs. 83.1 and 83.3).[1,7] The general function of molecular (co)chaperones is to bind and stabilize proteins at intermediate stages of folding, assembling, translocation, and degradation. With regard to cGCR, they also regulate cellular signaling, which includes (1) stabilizing a high-affinity conformational state of cGCR; (2) opening the glucocorticoid binding cleft to access by glucocorticoids; and (3) stabilizing the binding of GCR to the promoter.[1]

The first step in assembling the multiprotein cytosolic complex is ATP- and hsp40(YDJ-1)-dependent formation of a cGCR-hsp70 complex that primes the receptor for subsequent ATP-dependent activation by hsp90, Hop, and p23.[5] The glucocorticoid receptor consists of different domains with distinct functions: an N-terminal domain, a DNA-binding domain (DBD), and a ligand-binding domain (LBD) (see Fig. 83.3). The N-terminal domain harbors transactivation functions, especially within the "τ1" region. The zinc finger motif, a sequence common to many DNA-interacting proteins, is found twice within the DBD. The LBD consists of 12 α helices, several of which help form a hydrophobic ligand-binding pocket.[6] The cGCR contains another major transactivation region ("τ2") that can interact with the above-mentioned cofactors (see Fig. 83.3). Following GC/cGCR binding, the hsp90 molecules and other molecular chaperones are rapidly shed. This allows translocation into the cell nucleus, where the GC/cGCR complex binds as a homodimer to consensus palindromic DNA sites (GC-responsive elements [GREs]).[7]

Translocation Into the Nucleus

Nuclear translocation of the GC/cGCR complex occurs within 20 minutes. This may be caused by hormone-directed

[†] Deceased.

TABLE 83.1 Current Knowledge on the Relationship Between Clinical Glucocorticoid Dosing and Cellular Glucocorticoid Actions

Terminology (mg Prednisone Equivalent per Day)	Clinical Application	Genomic Actions (Receptor Saturation)	Unspecific Nongenomic Actions	cGCR-Mediated Nongenomic Actions
Low dose (≤7.5)	Maintenance therapy for many rheumatic diseases	+ (< 50%)	–	?
Medium dose (>7.5–≤ 30)	Initially given in primary chronic rheumatic diseases	++ (> 50–<100%)	(+)	(+)
High dose (>30–≤100)	Initially given in subacute rheumatic diseases	++ (+) (almost 100%)	+	+
Very high dose (>100 mg)	Initially given in acute and/or potentially life-threatening exacerbations of rheumatic diseases	+++ ([almost] 100%)	++	+ (+?)
Pulse therapy (≥250 mg for 1 or a few days)	Particularly severe and/or potentially life-threatening forms of rheumaticdiseases	+++ (100%)	+++	+ (++?)

cGCR, Glucocorticoid receptor.
From Buttgereit F, Straub RH, Wehling M, Burmester GR. Glucocorticoids in the treatment of rheumatic diseases. An update on mechanisms of action. *Arthritis Rheum.* 2004;50:3408–3417, with permission.

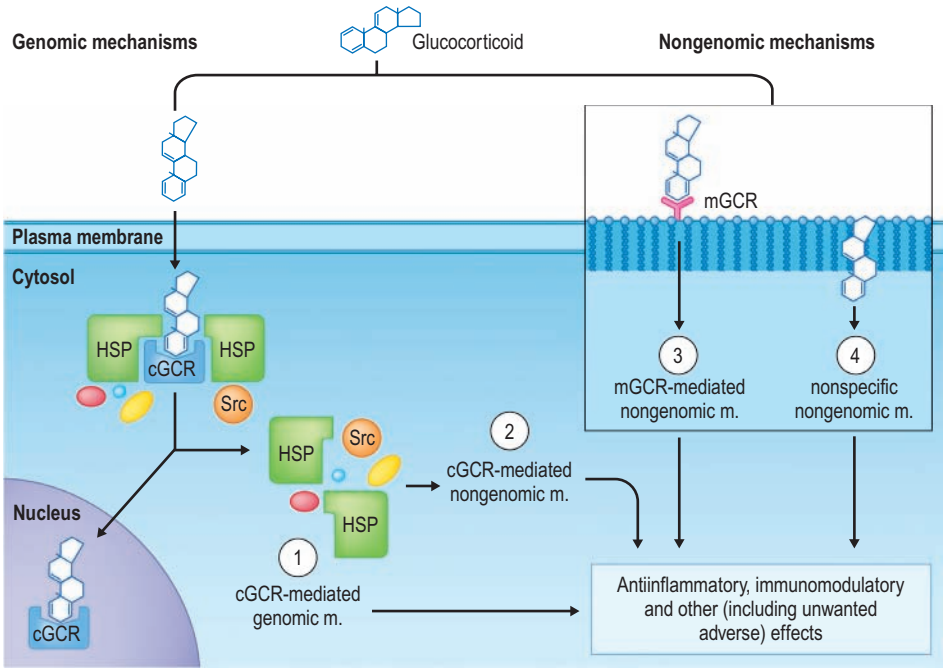


FIG. 83.1 Mechanisms of the Cellular Action of Glucocorticoids. As lipophilic substances, glucocorticoids pass very easily through the cell membrane into the cell, where they bind to ubiquitously expressed cytosolic glucocorticoid receptor *(cGCR)*. This is followed by either the classic cGCR-mediated genomic effects *(1)* or by cGCR-mediated nongenomic effects *(2)* Moreover, the glucocorticoid is very likely to interact with cell membranes, either specifically via membrane-bound glucocorticoid receptor *(mGCR) (3)* or via nonspecific interactions with cell membranes *(4)*. *HSP, heat shock protein.* (From Buttgereit F, Straub RH, Wehling M, Burmester GR. Glucocorticoids in the treatment of rheumatic diseases. An update on mechanisms of action. *Arthritis Rheum.* 2004;50:3408–3417.)

recruitment of an immunophilin (FKBP51, FKBP52, or CyP-40) or the protein phosphatase PP5 and dynein to the GCR.[7] Depending on the target gene, transcription is then either activated (*transactivation* via positive GRE) or inhibited (*transrepression* via negative GRE) (see Fig. 83.2, A and B).

Interactions With Transcription Factors

As well as the interactions of GC/cGCR complexes with GREs, a further important genomic mechanism of GC action is the interaction of activated cGCR monomers with transcription factors. Accordingly, although the GC/cGCR complex does not inhibit their synthesis, it modulates the activity of AP-1 (activator protein-1), NF-κB (nuclear factor-kappa B) and NF-AT (nuclear factor for activated T cells). This leads to inhibition of nuclear translocation and/or function of these transcription factors, and hence to inhibition of expression of many immunoregulatory and inflammatory cytokines. Possible mechanisms include:[8]

- Synthesis of IκB (a specific inhibitor of NF-κB) induced through GC/cGCR complex–GRE interaction (see Fig. 83.2, A).
- Protein–protein interaction of the GC/cGCR complex with transcription factors through binding to their subunits (see Fig. 83.2, C), which prevents their DNA binding.
- Competition for nuclear co-activators between the GC/cGCR complex and transcription factors (see Fig. 83.2, D).

Inhibition of transcription factor function and the resultant inhibition of protein expression are referred to as *transrepression*. Numerous genes are regulated by this mechanism. Many adverse effects of GC are caused by *transactivation* (induced synthesis of regulator proteins), whereas most antiinflammatory effects are mediated by *transrepression* (inhibited synthesis of regulator proteins). This differential molecular regulation underlies current drug-discovery programs aimed at developing dissociated cGCR-ligands.[9]

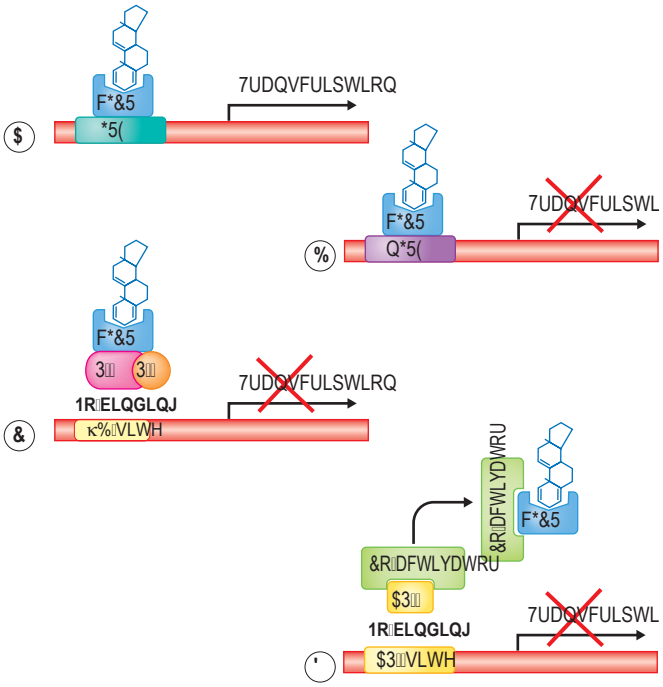


FIG. 83.2 (A–D) **Genomic Mechanisms of Glucocorticoids.** (A–D) The different mechanisms by which the activated glucocorticoid receptor complex leads to the induction or to the inhibition of transcription and finally translation/synthesis of specific regulator proteins. Details are given in the text. *cGCR*, Cytosolic glucocorticoid receptors; *GRE*, glucocorticoid-responsive elements. (From Buttgereit F, Straub RH, Wehling M, Burmester GR. Glucocorticoids in the treatment of rheumatic diseases. An update on mechanisms of action. *Arthritis Rheum.* 2004;50:3408–3417.)

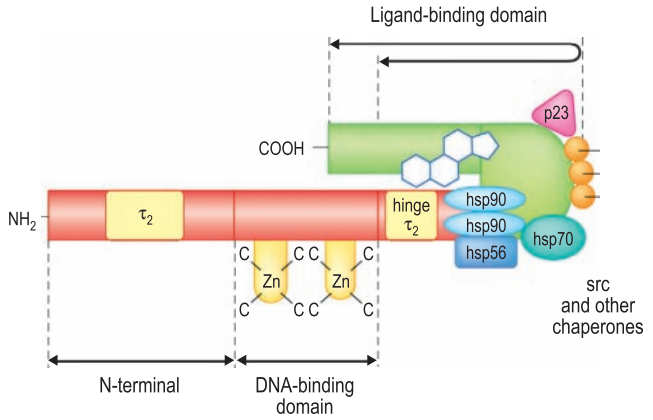


FIG. 83.3 **Structure of the Cytosolic Glucocorticoid Receptor (cGCR).** The unactivated (unligated) cGCR is a 94 kDa protein retained in the cytoplasm as a multiprotein complex consisting of several heat shock proteins *(hsp)*, including hsp90, hsp70, hsp56, and hsp40 (chaperones). Furthermore, the cGCR interacts with immunophilins, p23, and several kinases of the mitogen-activated protein kinase (MAPK) signaling system, including Src, which also act as molecular (co-)chaperones. An important function of molecular (co)chaperones is to stabilize a specific conformational state of the GC which binds ligand with high affinity (see text). The receptor protein itself consists of different domains: an N-terminal, a DNA-binding domain (DBD), and a ligand-binding domain (LBD). The N-terminal domain harbors transactivation functions, especially within the so-called τ1 region. Another major transactivation region is τ2, which can interact with the above-mentioned cofactors. (From Buttgereit F, Straub RH, Wehling M, Burmester GR. Glucocorticoids in the treatment of rheumatic diseases. An update on mechanisms of action. *Arthritis Rheum.* 2004;50:3408–3417.)

The Cytosolic Glucocorticoid Receptor β Isoform

The cGCRβ isoform is an alternative splicing variant of cGCRα that does not bind GC or activate gene expression. This isoform is thought to function as a negative inhibitor of cGCRα and it may play a role in the clinical phenomenon of GC-resistance. cGCRβ lacks the GC-binding domain, which is needed for activation, and as it does not undergo ligand-dependent downregulation, it has a longer half-life than the active form (cGCRα). It is thought that the likely mechanism of the dominant negative activity of cGCRβ is through the formation of inactive heterodimers with cGCRα.[8]

Post-Transcriptional and Post-Translational Mechanisms

Glucocorticoids also act through post-transcriptional and post-translational mechanisms, including reduction of the half-life of cytokine mRNA and downregulation of GCR, via reduced mRNA levels and reduced stability of the GCR protein.

KEY CONCEPTS

Glucocorticoid Effects on Immune Cells

- Inhibit leukocyte traffic and access of leukocytes to the site of inflammation.
- Interfere with functions of leukocytes, fibroblasts, and endothelial cells.
- Suppress the production and actions of humoral factors involved in the inflammatory process.

Glucocorticoid Receptor Resistance

Several different mechanisms may explain the clinical finding of GCR resistance, among them alterations in the number, binding affinity, or phosphorylation status of GCR. Other possible explanations are polymorphic changes and/or overexpression of (co-)chaperones, increased expression of inflammatory transcription factors, overexpression of the GCRβ isoform, the multidrug resistance pump, and altered mGCR expression.[10]

Nongenomic Actions of Glucocorticoids

Some regulatory effects of glucocorticoids occur within seconds or minutes. These are too rapid to result from genomic actions, and therefore must be due to *nongenomic* mechanisms of action. Three different rapid nongenomic actions of glucocorticoids have been described.

Cytosolic Glucocorticoid Receptor–Mediated Nongenomic Actions

Dexamethasone can rapidly inhibit epidermal growth factor-stimulated $cPLA_2$ (cytosolic phospholipase A_2) activation with subsequent arachidonic acid release.[11] This effect is thought to be mediated by occupation of cGCR, rather than changes in gene transcription, as the observed effect is RU486-sensitive (i.e., glucocorticoid receptor dependent) but actinomycin-insensitive (i.e., transcription independent). Chaperones or co-chaperones of the multiprotein complex may act as signaling

components to mediate this effect. Following glucocorticoid binding, the cGCR is released from this complex to mediate classic genomic actions. However, there is also a rapid release of Src and other (co-)chaperones of the multiprotein complex, which may cause rapid inhibition of arachidonic acid release. Similarly, dexamethasone has been reported to have cardiovascular protective effects which are neither genomic (because they occurred too quickly and were actinomycin insensitive) nor nonspecific-nongenomic effects (because they occurred at a very low concentration (100 nM)).[12] These may involve binding of glucocorticoids to the cGCR, leading to nontranscriptional activation of phosphatidylinositol 3-kinase, protein kinase Akt, and endothelial nitric oxide synthase.

Nonspecific Nongenomic Actions

Glucocorticoids are sometimes administered at very high doses. Systemic daily dosages greater than 100 mg prednisone equivalent are regarded as "very high dose." "Pulse therapy" is the daily administration of ≥250 mg prednisone equivalent for one or a few consecutive days[13] (see Table 83.1). At a daily dose of 100 mg prednisone equivalent, almost all cGCR are completely saturated, which implies that specificity (i.e., the exclusivity of receptor-mediated effects) is lost. Nonspecific nongenomic actions occur in the form of physicochemical interactions with biological membranes, which probably contribute to the therapeutic effect.[4] Intercalation of glucocorticoid molecules into cell membranes is thought to alter cell function by influencing cation transport and increasing mitochondrial proton leak. The resulting inhibition of calcium and sodium cycling across the plasma membrane of immune cells is thought to contribute to rapid immunosuppression and to reduced inflammation.

Such high GC doses are only used in a few clinical specialties, and this practice has been criticized by endocrinologists and pharmacologists. Unfortunately, there are no randomized controlled trials of high-dose glucocorticoid therapy, but it is often used with clinical success in acute exacerbations of life-threatening diseases and various clinical conditions resistant to other therapies. For example, pulsed intravenous (IV) methylprednisolone is effective in the treatment of systemic lupus erythematosus (SLE) and rapidly immunosuppresses patients with organ- and/or life-threatening manifestations of SLE. However, the standard regime (1 g/day for 3 consecutive days) is associated with significant risks of infection, and lower doses may be just as effective.[14]

High-dose glucocorticoids are also often used in immune thrombocytopenia associated with SLE, although comparative studies are lacking[15] It has been calculated that, in situations like these, the concentrations achieved in vivo are high enough (around 10^{-5} mol/L) to cause immediate nonspecific nongenomic effects on immune cells.[4] Intraarticular injections also bring high concentrations of glucocorticoids into contact with inflammatory cells, although it is difficult to assess locally achieved concentrations because crystal suspensions are most often used.

Specific Nongenomic Actions

Glucocorticoids can also induce specific nongenomic actions, mediated through membrane-bound glucocorticoid receptors (mGCRs). The existence and function of membrane-bound receptors have been demonstrated for various steroids (including mineralocorticoids, gonadal hormones, vitamin D, and thyroid hormones).[4,16] Small numbers of mGCRs can be demonstrated by immunofluorescence on human peripheral blood mononuclear cells (monocytes and B cells) obtained from healthy controls.[16] The monoclonal antibody used to detect mGCR also recognizes cGCR, suggesting that mGCRs are probably variants of cGCRs produced by differential splicing or promoter switching. Immunostimulation with lipopolysaccharide increases the percentage of $mGCR^+$ monocytes and this can be prevented by inhibiting the secretory pathway with brefeldin A. This suggests that mGCRs are actively upregulated and transported through the cell following immunostimulation. These in vitro findings are consistent with observations that the frequency of $mGCR^+$ monocytes is increased in patients with rheumatic disorders and is positively correlated with disease activity in RA.[16] It remains unclear whether mGCRs are involved in pathogenesis. Alternatively, and perhaps they may more likely, cause negative feedback regulation.

GLUCOCORTICOID EFFECTS ON IMMUNE CELLS

Through the above mechanisms, glucocorticoids mediate a wide range of antiinflammatory and immunomodulatory effects, with virtually all primary and secondary immune cells affected to some extent (Table 83.2).[8]

KEY CONCEPTS

Definition of Conventional Terms for Glucocorticoid Dosages

Low dose	≤7.5 mg prednisone equivalent per day
Medium dose	>7.5 mg, but ≤30 mg prednisone equivalent per day
High dose	>30 mg, but ≤100 mg prednisone equivalent per day
Very high dose	>100 mg prednisone equivalent per day
Pulse therapy	≥250 mg prednisone equivalent per day for 1 or a few days

TABLE 83.2 Important Effects of Glucocorticoids on Primary and Secondary Immune Cells

Monocytes/Macrophages
- ↓ number of circulating cells (↓ myelopoiesis, ↓ release)
- ↓ expression of MHC class II molecules and Fc receptors
- ↓ synthesis of proinflammatory cytokines (e.g., IL-2, IL-6, TNF-α) and prostaglandins

T Cells
- ↓ number of circulating cells (redistribution effects)
- ↓ production and action of IL-2 (most important)

Granulocytes
- ↓ number of eosinophil and basophil granulocytes
- ↑ number of circulating neutrophils

Endothelial Cells
- ↓ vessel permeability
- ↓ expression of adhesion molecules
- ↓ production of IL-1 and prostaglandins

Fibroblasts
- ↓ proliferation
- ↓ production of fibronectin and prostaglandins

IL, Interleukin; *MHC*, major histocompatibility complex; *TNF-α*, tumor necrosis factor alpha.

From Buttgereit F, Saag K, Cutolo M, et al. The molecular basis for the effectiveness, toxicity, and resistance to glucocorticoids: focus on the treatment of rheumatoid arthritis. *Scand J Rheumatol.* 2005;34:14–21, with permission.

THE ROLE OF ENDOGENOUS GLUCOCORTICOIDS IN INFLAMMATORY ARTHRITIS

Exogenous (therapeutic) and endogenous (physiological) glucocorticoids differ in several respects. The most important differences are in their relative mineralocorticoid and glucocorticoid (anti-inflammatory) activities. Exogenous and endogenous glucocorticoids also differ in their pharmacological characteristics, such plasma kinetics, metabolism, biological half-life, lipophilicity, drug-receptor interactions, and nongenomic potencies.[17]

The actions of exogenous glucocorticoids as described above are well established. In contrast, we know relatively little about the role of endogenous glucocorticoids in arthritis. While glucocorticoid actions on target tissues are thought to be determined by glucocorticoid plasma concentrations and the tissue-specific density of glucocorticoid receptors, it seems that endogenous glucocorticoids are subject to extensive pre-receptor metabolism. Within target cells or tissues, glucocorticoid action depends not only on plasma hormone levels, receptor expression, and receptor-effector coupling but also on local glucocorticoid metabolism. Specifically, 11β-hydroxysteroid dehydrogenases appear to govern access of glucocorticoids to their cognate receptors by changing the balance between active and inactive glucocorticoids within the cell (reviewed in Buttgerit *et al.*[17]). Thus, the predominant reductase activity of 11β-hydroxysteroid dehydrogenase *type 1* (11βHSD1) catalyzes the formation of bioactive cortisol from inactive cortisone (in humans) and corticosterone from 11-dehydrocorticosterone (in rodents). This $NADP^+(H)$-dependent enzyme is present in many tissues and usually increases the intracellular availability of active glucocorticoids. In contrast, 11β-hydroxysteroid dehydrogenase *type 2* (11βHSD2) only possesses dehydrogenase activity: it inactivates active glucocorticoids and therefore decreases the intracellular concentration of bioactive glucocorticoids.

Proinflammatory cytokines such as interleukin-1β (IL-1β) and tumor necrosis factor alpha (TNF-α) stimulate 11βHSD1 and downregulate 11βHSD2 expression. Hence, specific proinflammatory cytokines can modulate local intracellular glucocorticoid metabolism, which may affect their own proinflammatory effects. More recently, substantial glucocorticoid metabolism has been shown in joints since TNF-α and IL-1β induce 11βHSD1 activity in primary cultures of synovial fibroblasts isolated from synovial tissue biopsies from rheumatoid arthritis (RA) patients.[18]

In a rodent model of immune-mediated arthritis, targeted disruption of glucocorticoid signaling in osteoblasts attenuates joint inflammation and cartilage destruction.[19] These results suggest that, under the control of endogenous glucocorticoids, osteoblasts modulate immune-mediated inflammatory responses, and, as a consequence, inflammation-induced cartilage damage and bone integrity. These findings are supported by recent evidence suggesting that the effects of glucocorticoids follow a dose–response curve with permissive or even stimulatory effects at physiological concentrations, and suppressive effects at pharmacological concentrations.[18]

THERAPEUTIC USE

A wide range of GC molecules are available for clinical use: the common basic structure has been modified to improve their usefulness in various clinical applications (Fig. 83.4). Despite their widespread use, the designation of GC treatment regimens is often imprecise. Recommendations for a standardized nomenclature for GC therapy are summarized below.[13]

Terminology

Although the term "*steroids*" is widely used to describe this class of drugs, it is too broad, as it simply describes chemical compounds characterized by a common multiple-ring structure (including cholesterol, vitamin D, and sex hormones). Similarly, the terms "*corticosteroids*" or "*corticoids*" are not sufficiently precise, as the adrenal cortex synthesizes not only glucocorticoids but also mineralocorticoids and androgens. For these reasons, the terms "*glucocorticoid*" or "*glucocorticosteroid*" are preferred, but "*glucocorticoid*" is the more widely used term.

When describing the use of glucocorticoids, it is necessary to define the drug, the dosage, the route of administration, and the timing, frequency, and duration of treatment.

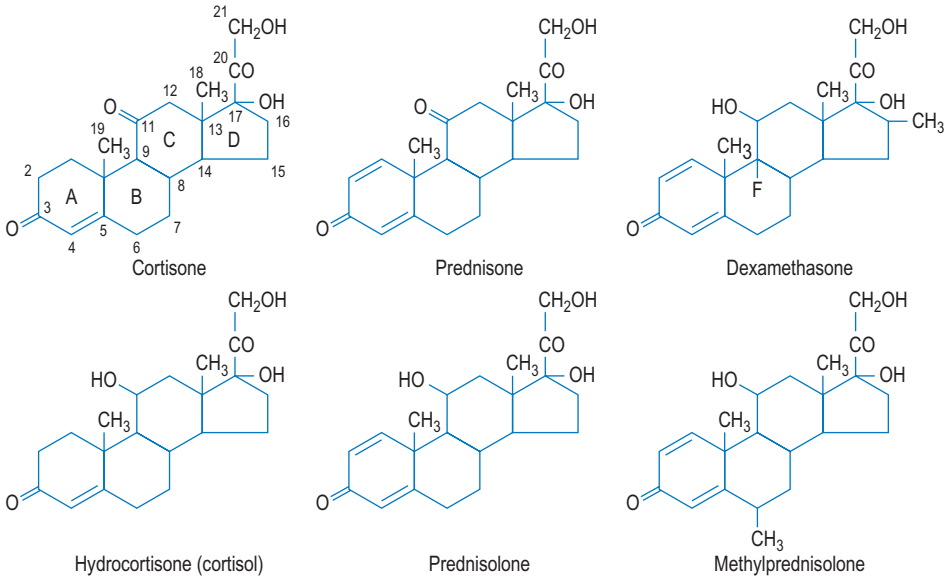


FIG. 83.4 Molecular Structures of Cortisone and Commonly Used Glucocorticoids. Carbon and ring nomenclature is indicated on the cortisone molecule.

TABLE 83.3 Drug Potencies of Selected Glucocorticoids

Cortisol	1 (per definition)
Prednyliden	3.5
Predniso(lo)ne	4.0
Methylprednisolone	5.0
Dexamethasone	25
Betamethasone	25

Drug potencies describe the potency of the respective drug to produce classical genomic (antiinflammatory) effects relative to cortisol. These potencies provide the basis for calculating equivalent dosages.
Data from Lipworth BJ. Therapeutic implications of non-genomic glucocorticoid activity. *Lancet.* 2000;356:87–89.

Different glucocorticoid drugs have different potencies and they differ in their ability to produce the distinct therapeutic effects discussed above. Drug potencies are usually described by the equivalent dosages (relative potencies) to produce classic genomic effects (Table 83.3). These values have been used for decades, although the experimental and clinical evidence for their precise relative potency is weak. In practice, relative potencies are useful as a general therapeutic guideline in daily clinical practice, as long as they are not used dogmatically. It has therefore been suggested that we should continue to use relative potencies until more exact data are available, and that doses of different glucocorticoids should be expressed in terms of "prednisone equivalent": that is, doses of different glucocorticoids are expressed as equivalent to milligrams of prednisone (or prednisolone, as the potency of prednisone is equal to that of prednisolone).

However, recent data indicate that the concept of equivalent dosages is only valid for doses less than 100 mg prednisone equivalent, because nongenomic effects come into play at higher doses. This reason is important because the relative potencies of different glucocorticoids to produce these nongenomic effects are completely different from their classic genomic effects (see Table 83.3). For example, methylprednisolone is used for pulse therapy of exacerbations of immunologically mediated disorders. Prednisolone and methylprednisolone have similar genomic potency, but when used for high-dose therapy the nonspecific nongenomic effect of methylprednisolone is more than three times stronger. This may explain the apparent superiority of high-dose methylprednisolone. In contrast, betamethasone has very low nongenomic potency, which may be why this drug is considered less effective systemically, even though it has the same genomic potency as dexamethasone. In summary, the clinical usage of different glucocorticoids is clearly determined by the magnitude of their clinical efficacy, but another important factor in selecting which one to use is their nongenomic potency.

GLUCOCORTICOID TREATMENT REGIMENS: GENERAL ASPECTS

For many decades there has been confusion surrounding the terms used to describe dosage (very low, low, mild, mild to moderate, moderate, high, very high, ultra-high, and mega) and by loose usage of terms such as "low-dose therapy," "high-dose therapy," and "pulse therapy." A 2002 consensus statement has clarified this situation, partly to achieve scientific consistency and partly to recognize that glucocorticoid actions are strongly dose-dependent, both quantitatively and qualitatively. The following standardized nomenclature for glucocorticoid dosages and glucocorticoid treatment regimens is now recommended.[13]

Low Dose

Treatment with doses of up to 7.5 mg prednisone equivalent per day is considered low-dose glucocorticoid therapy because these doses occupy less than 50% of the receptors. Such courses are frequently used for maintenance therapy; and have relatively few adverse effects (such as osteoporosis). As there may be relative hypocortisolism in chronic inflammatory conditions such as RA and polymyalgia rheumatica, low-dose glucocorticoids act in part as replacement therapy for reduced adrenal glucocorticoid production.

Medium Dose

Glucocorticoid doses of more than 7.5 mg but less than 30 mg prednisone equivalent per day are considered to be medium-dose therapy because they lead to a significantly higher receptor engagement between 50% and less than 100%. These doses are effective in modulating disease activity in various rheumatic diseases but may have considerable and dose-dependent adverse effects if given for longer periods.

High Dose

Treatment with doses of more than 30 mg and up to 100 mg of prednisone equivalent per day is considered high-dose therapy because these doses significantly increase receptor saturation, in a dose-dependent manner. At approximately 100 mg prednisone equivalent per day, receptor saturation is almost complete and it is likely that genomic glucocorticoid effects are fully exerted (see Table 83.1). High-dose therapy can be successfully given as initial treatment for subacute diseases such as non-life-threatening exacerbations or visceral complications of RA or other connective tissue diseases but cannot be administered long term, because of the danger of severe adverse effects.

Very High Dose

Doses above 100 mg of prednisone equivalent per day are considered "very high." At this level, there is virtually 100% saturation of cytosolic receptors, so any further increase in dose may affect the pharmacodynamics (e.g., receptor off-loading and re-occupancy), receptor synthesis, and expression. At these doses nongenomic effects may deliver additional therapeutic benefit, although it remains unclear whether these effects have any direct therapeutic relevance. Experimental data suggest that these differential effects come increasingly into play above ~100 mg/day (see Table 83.1). Doses of greater than 100 mg of prednisone equivalent per day are frequently (and successfully) given as initial treatment for acute or life-threatening exacerbations of connective tissue diseases, vasculitis, and RA. These doses cannot be administered long term because of their severe adverse effects.

Pulse Therapy

Pulse therapy involves administration of ≥250 mg prednisone equivalent per day (usually IV) for a short period of time

(typically 1 to 5 days, rarely longer). At such high doses, the nongenomic potencies of glucocorticoids come into play. It is likely that these explain the success of very high doses and pulse therapy in acute exacerbations of immunologically mediated diseases. The immediate effects of very high doses may be additive to the genomic effects mediated by cytosolic GC receptors. These additional effects may make a crucial contribution to the therapeutic effect by helping to terminate acute exacerbations. Circumstances where very high doses or pulse therapy can be successful include acute episodes or particularly severe forms of rheumatic diseases such as systemic lupus erythematosus, vasculitis, polymyositis, and RA (see below).

Alternate-Day Regimens

Alternate-day regimens were introduced for long-term oral GC therapy with the aim of minimizing undesirable adverse effects, such as suppression of the hypothalamo-pituitary–adrenal (HPA) axis. Instead of daily dosing, a single dose is administered every other morning, usually in a dose equivalent to, or somewhat higher than, twice the usual daily dose. The idea behind this regimen is to allow the HPA axis to remain active by exposing the body to exogenous GC on alternate days instead of suppressing it every day. This strategy only works if the HPA axis is still active, and, unfortunately, patients often experience breakthrough symptoms on the second day of treatment. Alternate-day regimens are used rarely these days except in patients with juvenile idiopathic arthritis, in whom alternate-day glucocorticoid regimens cause less inhibition of body growth.

Glucocorticoid Withdrawal Regimens

Because of their significant adverse effects, GCs are usually reduced or stopped as soon as the disease is under control. This needs to be done carefully in order to avoid a relapse in disease activity and to permit recovery of adrenal function. There are no controlled comparative studies to support a specific regimen for weaning patients off GCs, as this process needs to be adjusted according to disease activity, dose/duration of therapy, and clinical response. When patients with rheumatoid arthritis are treated with up to 10 mg/day of prednisone, the daily dose can be reduced by 2.5 mg every month until 5 mg/day is reached. Thereafter, doses can often be reduced by 1 mg per month. When higher doses have been used, a reduction by 5 mg every 1 to 2 weeks down to 20 mg/day is often well tolerated, followed by further reductions of 1 to 2.5 mg/day every 2 to 3 weeks. Addition of immunomodulatory drugs such as methotrexate and azathioprine may allow further dose reductions. Short courses of GC for conditions such as asthma may be ceased without tapering if the total treatment course was less than 14 days in duration. With longer periods of treatment, the dose should be reduced gradually to allow the HPA axis to recover, but the rate of reduction is usually quicker than for rheumatological conditions.

Glucocorticoids in Rheumatoid Arthritis: An Example

Glucocorticoids are crucially important in the management of rheumatoid arthritis and are used in various dosages at different disease stages. RA is therefore a useful example to discuss GC therapy in more detail.

Low-Dose Maintenance Therapy

In early RA, GC in doses less than 10 mg are highly effective for relieving symptoms in patients with active arthritis. Many patients are functionally dependent on this low-dose therapy and continue it long term.[20] Improvement has been documented in all clinical parameters, including pain scales, joint scores, morning stiffness, and fatigue, but also in acute inflammatory markers such as erythrocyte sedimentation rate (ESR) and C-reactive protein (CRP). After 6 months of therapy the beneficial effects of GC seem to diminish, but if therapy is then tapered and stopped, patients often experience an exacerbation of symptoms within a few months.

The disease-modifying properties of GC were first described in 1995, in a 2-year trial of 7.5 mg prednisolone in patients with RA of short/intermediate disease duration who also were treated with nonsteroidal antiinflammatory drugs (NSAIDs) (95%) and disease-modifying antirheumatic drugs (DMARDs) (71%). Presently, GC are considered to have disease-modifying properties in early RA, but it is less clear whether they inhibit the progression of erosions in RA of longer duration. A meta-analysis of 15 studies that included 1414 patients concluded that GC given in addition to standard therapy can substantially reduce the rate of erosion progression in RA.[21] Another meta-analysis of 70 randomized placebo- or drug-controlled studies found similar effects of DMARDs, glucocorticoids, and biologicals on radiographic progression in RA.[22]

Glucocorticoid Pulse Therapy

GC pulse therapy is used to treat some serious complications of RA and to induce remission in active disease: for example, when initiating second-line antirheumatic treatment. Pulse therapy with methylprednisolone 1 g/day, dexamethasone 200 mg/day, or equivalent, given intravenously for 1 to 3 days, was effective in most studies with a beneficial effect that generally lasts for about 6 weeks, albeit with large variations. This means it is not sensible to apply pulse therapy in active RA, unless the therapeutic strategy is changed (e.g., DMARD treatment introduced to maintain the remission induced by pulse therapy).

Intraarticular Glucocorticoid Injections

Intraarticular injections with GC are often used in RA. The benefit achieved varies according to several factors, such as the joint treated (size, weight-bearing, or non-weight-bearing), inflammatory activity, the volume of synovial fluid in the joint, whether or not synovial fluid was aspirated before the injection, the choice and dose of GC preparation, injection technique, and whether the joint is rested after the injection.[23] In order to prevent GC-induced joint damage, it is recommended that intraarticular GC injections

KEY CONCEPTS

Glucocorticoid Therapy in Rheumatoid Arthritis

The risk–benefit ratio of low-dose GC has been shifted in the last few years:

- GCs can now be considered as disease-modifying antirheumatic drugs, especially in early RA.
- Adverse effects of low-dose GCs are less abundant and less severe than previously suggested, and some (e.g., osteoporosis) can be well managed.
- The goal of antirheumatic treatment in early RA is to induce remission of disease by aggressive management: GCs are part of this aggressive strategy.

should not be repeated more often than once every 3 to 4 weeks, and no more than 3 to 4 times a year in a weight-bearing joint.

Adverse Effects

Studies of GC toxicity tend to be retrospective and observational. This can make it difficult to differentiate unfavorable outcomes attributable to GCs from those occurring due to the underlying disease or other comorbidities. Furthermore, there is a strong selection bias for GC use, as physicians are more inclined to use them in patients with more severe disease. Frequent, but less serious adverse effects (such as skin thinning, Cushingoid appearance) may be of great concern to patients, whereas more debilitating toxicities such as osteoporosis, cataracts, and GC-induced hypertension may initially go unrecognized or be asymptomatic. Interpretation of toxicity data is further confounded by the use of GCs at variable points in the disease course, limited data defining "threshold" doses for particular adverse events, and the fact that toxicity reports cover a heterogeneous group of GC-treated diseases.

Compared to other antirheumatic agents, GCs have a low incidence of short-term symptomatic toxicity and patients rarely discontinue therapy for these reasons. Despite over 60 years of use, we still lack robust data on the longer-term toxicities of glucocorticoids from large randomized controlled trials with long-term follow-up. The commonest GC toxicities are summarized below.

Some progress has been made by formulating recommendations on which adverse effects of glucocorticoid treatment should be monitored in rheumatoid arthritis, how to monitor them, and how often.[24] Two levels of monitoring GC adverse events have been proposed. (1) For routine clinical practice, details are given on how to identify adverse events in a systematic and practical way. This should result in preventive and therapeutic measures in order to minimize the risks of glucocorticoid therapy. (2) For clinical trials, recommendations have been made on how to accurately assess the frequency and severity of a wider range of adverse events.[25]

Osteoporosis

Glucocorticoid-induced osteoporosis (GIOP) is the most important potential complication of prolonged GC therapy. Chronic GC treatment results in rapid and profound reductions in bone mineral density, with most bone loss occurring during the first 6 to 12 months of treatment.[9]

GIOP initially affects trabecular bone, but cortical bone is also affected with more chronic use, at sites such as the femoral neck. Precisely how GCs affect bone remains obscure. GCs decrease calcium absorption, increase renal calcium loss, diminish sex and growth hormone production, induce muscle wasting, and modulate RANKL/OPG, NF-κB, and AP-1 signaling in bone.[9] All of these changes lead to enhanced osteoclast function and life span, and hence to increased bone resorption. Consequently, markers of bone resorption are often increased in patients treated with GCs.[9] However, reduced bone formation due to reduced osteoblast function is likely to be a more important effect of GCs on skeletal health. Oral doses of prednisone as low as 2.5 mg/day have been shown to suppress serum osteocalcin, a marker of bone formation. Histologically, mean wall thickness is reduced, reflecting the reduced amount of bone replaced in each remodeling unit. In vitro, osteoblasts and their precursors are highly GC responsive. Here, the predominant effect is to promote osteoprogenitor proliferation, lineage commitment, and osteoblast differentiation, resulting in the formation of bone nodules of increased size and numbers. However, GCs also inhibit type I collagen expression, and reduce pre-osteoblastic replication. Finally, GCs promote apoptosis of osteoblasts and osteocytes. The inhibitory effects of GCs on bone formation may be partly due to downregulation of insulin-like growth factor-1 (IGF-1) expression by osteoblasts. Fortunately, we now have effective strategies for the prevention and treatment of GC-induced osteoporosis, using calcium, vitamin D, and specific osteotropic agents such as bisphosphonates or parathyroid hormone.

Osteonecrosis

Osteonecrosis has long been considered an important consequence of high-dose glucocorticoid use. In a Japanese study of femoral head osteonecrosis, 35% cases were related to GC treatment. Higher average dose may be a more important predictor of avascular necrosis of bone than cumulative dose. Osteonecrosis is particularly noted in SLE, but rarely occurs in RA patients receiving low-dose therapy, affecting less than 3% of patients. Osteonecrosis rarely occurs in SLE patients on prednisone doses less than 20 mg/day.

Myopathy

As with osteonecrosis, GC-induced myopathy is rare in patients receiving low-dose glucocorticoids. In small studies, myopathy appears more closely associated with fluorinated GC preparations such as triamcinolone than with prednisone. Notably, myopathy has been reported after only 3 months' treatment with triamcinolone 8 mg/day. In general, myopathy attributable to prednisone only occurs after higher doses and longer durations of treatment.

Cardiovascular Adverse Effects

Glucocorticoid-induced hypertension seems to be, at least in part, mediated via fluid retention (as a result of mineralocorticoid effects); it is dose related and less likely with medium- or low-dose therapy. Individual variation in susceptibility and other factors, such as the starting level of blood pressure, dietary salt intake, functional renal mass, associated diseases, and drug therapy may also play a role.

Another troublesome potential toxicity of low-dose GCs is the development of premature atherosclerotic vascular disease. This has proven difficult to investigate: studies evaluating the effects of GCs on lipids and atherosclerosis in RA patients have yielded mixed results, with some studies suggesting that GCs may actually reverse unfavorable lipid changes. At present, there is no evidence of a strong association between low-dose glucocorticoids and cardiovascular disease in RA, even though atherosclerotic vascular disease is known to be accelerated in patients with Cushing disease. Systemic GC use during pregnancy may also have epigenetic effects, leading to adult hypertension in the next generation.[26]

Dermatological Adverse Effects

Skin thinning and ecchymoses are common adverse events with glucocorticoids, even at low doses. Cutaneous atrophy results from catabolic effects of GC on keratinocytes and fibroblasts. Purpura and easy bruising in GC-treated patients are probably due to decreased vascular structural integrity. A Cushingoid appearance is very troubling to patients but is uncommon

at doses below the physiuological range. One study reported facial fullness ("moon facies") in 13% of patients receiving 4 to 12 mg triamcinolone for up to 60 days. These adverse effects are observed in over 5% of patients exposed to ≥5 mg prednisone equivalent for ≥1 year. The incidence of iatrogenic Cushing syndrome is dose-dependent and generally becomes evident after greater than 1 month of GC therapy. Alternate-day therapy may reduce the incidence, although there are only limited data supporting this concept. Glucocorticoid acne, hirsutism, and striae are other undesirable dermatological effects that occur even with lower doses.[27]

Gastrointestinal Adverse Effects

Glucocorticoids are considerably less toxic to the upper gastrointestinal (GI) tract than NSAIDs. If GCs independently increase the risk of GI events (e.g., gastritis, ulceration, bleeding), the effect is slight, with estimated relative risks varying from 1.1 (not significant) to 1.5 (marginally significant). There have also been anecdotal reports of intestinal rupture, diverticular perforation, and pancreatitis attributed to low-dose GCs. Glucocorticoids are frequently used concurrently with NSAIDs in RA, and meta-analyses confirm that the combination of the two drugs synergistically increases the risk of adverse GI events. In a large-scale study based on the UK General Practice Research Database, the risk of upper GI complications was 1.8 times higher for GC users than for nonusers (95% confidence interval [CI], 1.3 to 2.4). The risk tended to be greater for higher GC doses, but the dose gradient was not statistically significant. The risk was greater than 12 times higher for those taking both glucocorticoids and NSAIDs than for those not using either. No studies have yet looked at the GI effects of combining GCs with cyclooxygenase-2 (COX-2)-selective NSAIDs.[27]

Infectious Diseases

Medium- to high-dose GC therapy may increase the risk of serious infections leading to hospitalization or surgery, particularly when administered for prolonged periods of time. However, there is no evidence that infection rates are increased in patients on doses of prednisone below 10 mg/day or cumulative doses below 700 mg. In those taking higher doses, the risk of infection appears to be lessened with alternate-day therapy.

In GC-treated patients, physicians should be aware of the risk of infections with typical and atypical organisms, recognizing that GCs may blunt the classic clinical features and delay diagnosis. *Pneumocystis jiroveci* infections can occur with doses as low as 16 mg/day of prednisone for 8 weeks.[28] Herpes zoster is also more common in RA patients treated with immunosuppressive agents.[24] However, it is difficult to separate the independent effects of GC use from those of other commonly used antirheumatic agents, such as methotrexate and anti-TNF-α agents. At present, the independent role of GCs in facilitating herpes zoster infection in patients with RA remains uncertain.

Other Adverse Effects

Adverse effects on the hypothalamo–pituitary axis and on glucose metabolism are reviewed elsewhere, together with neuropsychiatric and ophthalmological adverse effects.[24,27,29]

TIMING OF GLUCOCORTICOID ADMINISTRATION MATTERS

In patients with rheumatoid arthritis, major symptoms such as pain, inflammation and stiffness vary across the day, usually with the highest severity in the morning hours. These symptoms are preceded by elevated levels of proinflammatory cytokines. Based upon these considerations it has been proposed that altering the timing of glucocorticoid administration may help to optimize RA therapy.[17]

NEW GLUCOCORTICOID RECEPTOR LIGANDS ON THE HORIZON

The various mechanisms of GC action provide interesting opportunities for developing optimized GCs and GC-receptor ligands.

Selective Glucocorticoid Receptor Agonists

The genomic component mechanisms of transactivation and transrepression offer the opportunity to develop GC-receptor ligands that predominantly cause transrepression rather than transactivation. This concept is based on the proposition that the antiinflammatory properties of GCs are mostly due to repression of AP-1- and NF-κB-stimulated synthesis of inflammatory mediators, whereas their adverse effects are associated with transactivation of genes involved in metabolic processes. Investigators have therefore sought novel GCR ligands with high transrepression activity but low effect on transactivation. One such compound, A276575, exhibits a high affinity for the GCR and potently represses IL-1α-induced IL-6 production, similar to dexamethasone. However, unlike dexamethasone, A276575 induces little aromatase activity. Other novel nonsteroidal GCR ligands are being developed that possess high repression activities against inflammatory mediator production but have lower transactivation activities than traditional GCs. Substances that cause a receptor conformation preferring a GCR/protein interaction as opposed to a GCR/DNA-binding-dependent mechanism are called selective glucocorticoid receptor agonists (SEGRAs) or "dissociated glucocorticoids." The SEGRA concept has, however, recently been challenged by studying a mouse knock-in strain with a dimerization-deficient GCR, which demonstrated that some inflammatory processes can be suppressed by glucocorticoids while others cannot.[30] Also, these mice exhibited classical adverse effects of glucocorticoids such as glucocorticoid-induced osteoporosis. Thus, depending on the process being treated, SEGRAs could be therapeutically more effective or less effective; moreover, not all adverse effects of glucocorticoid therapy may be reduced.[30,31] Nevertheless, from the clinical perspective it is disappointing that the sound underlying theory and the promising initial data have not led yet to a therapeutic breakthrough. As of now, it remains uncertain whether SEGRAs will become relevant in clinical practice.

Long-Circulating Liposomal Glucocorticoids

The antiinflammatory efficacy of GCs can be improved by the additional benefits of nongenomic actions at high GC concentrations. This has led to the use of long-circulating liposomal GCs in experimental models. In rats with experimental autoimmune encephalitis, GC-containing liposomes accumulate at sites of inflammation, reaching concentrations greater than

10^{-5} mol/L for ≥18 hours. These liposomes may be therapeutically superior to conventional intravenous high-dose GC therapy, as evidenced by their successful use in rats with adjuvant-induced arthritis. A single injection of 10 mg/kg liposomal prednisolone phosphate resulted in complete remission of the inflammatory response for almost a week. In contrast, the same dose of unencapsulated prednisolone phosphate did not reduce inflammation, and had only a slight effect after repeated daily injections. It may be that preferential delivery of GC to the site of inflammation leads to very high GC concentrations at the inflamed joint, but lower plasma concentrations, with a consequent lower rate of adverse effects. These are very promising developments which exploit the broad spectrum of therapeutically relevant genomic and nongenomic GC actions at the site of inflammation.[32]

In conclusion, glucocorticoids are extremely valuable antiinflammatory agents with a range of actions that are useful in the management of inflammatory conditions, including arthritis and asthma. Recent research on GCs has highlighted the effects of cytosolic GCRs on intracellular signaling, transcription processes, and gene expression. Exploration of membrane-bound glucocorticoid receptors, dose–effect relationships, and the timing of glucocorticoid administration have stimulated intensive research activity aimed at improving the efficacy:risk ratio. Modified-release prednisone has already been approved for the treatment of rheumatoid arthritis and the associated morning stiffness. Further clinical developments seem likely to follow, such as new GCR ligands and liposome encapsulation, with a view to improving the risk–benefit ratio of GC therapy and the well-being of patients.

ON THE HORIZON

- Selective glucocorticoid receptor agonists (SEGRAs)
- Modified-release steroids
- Liposomal encapsulation
- Optimized combinations with biologicals

REFERENCES

1. Shimba A, Ikuta K. Control of immunity by glucocorticoids in health and disease. *Semin Immunopathol*. 2020:20.
2. Laugesen K, Jorgensen JOL, Petersen I, Sorensen HT. Fifteen-year nationwide trends in systemic glucocorticoid drug use in Denmark. *European Journal of Endocrinology*. 2019;181:267–273.
3. Overman RA, Yeh JY, Deal CL. Prevalence of oral glucocorticoid usage in the United States: a general population perspective. *Arthritis Care Res (Hoboken)*. 2013;65:294–298.
4. Buttgereit F, Straub RH, Wehling M, Burmester GR. Glucocorticoids in the treatment of rheumatic diseases: an update on the mechanisms of action. *Arthritis & Rheumatism*. 2004;50:3408–3417.
5. Murphy PJ, Morishima Y, Chen H, et al. Visualization and mechanism of assembly of a glucocorticoid receptor.Hsp70 complex that is primed for subsequent Hsp90-dependent opening of the steroid binding cleft. *J Biol Chem*. 2003;278:34764–34773.
6. Wikstrom AC. Glucocorticoid action and novel mechanisms of steroid resistance: role of glucocorticoid receptor-interacting proteins for glucocorticoid responsiveness. *J Endocrinol*. 2003;178:331–337.
7. Pratt WB, Morishima Y, Murphy M, Harrell M. Chaperoning of glucocorticoid receptors. *Handbook of Experimental Pharmacology*. 2006:111–138.
8. Buttgereit F, Saag KG, Cutolo M, da Silva JA, Bijlsma JW. The molecular basis for the effectiveness, toxicity, and resistance to glucocorticoids: focus on the treatment of rheumatoid arthritis. *Scandinavian Journal of Rheumatology*. 2005;34:14–21.
9. Caplan A, Fett N, Rosenbach M, Werth VP, Micheletti RG. Prevention and management of glucocorticoid-induced side effects: a comprehensive review: a review of glucocorticoid pharmacology and bone health. *J Am Acad Dermatol*. 2017;76:1–9.
10. Rodriguez JM, Monsalves-Alvarez M, Henriquez S, Llanos MN, Troncoso R. Glucocorticoid resistance in chronic diseases. *Steroids*. 2016;115:182–192.
11. Croxtall JD, Choudhury Q, Flower RJ. Glucocorticoids act within minutes to inhibit recruitment of signalling factors to activated EGF receptors through a receptor-dependent, transcription-independent mechanism. *British Journal of Pharmacology*. 2000;130:289–298.
12. Hafezi-Moghadam A, Simoncini T, Yang Z, et al. Acute cardiovascular protective effects of corticosteroids are mediated by non-transcriptional activation of endothelial nitric oxide synthase. *Nature Medicine*. 2002;8:473–479.
13. Palmowski Y, Buttgereit T, Dejaco C, et al. "Official view" on glucocorticoids in rheumatoid arthritis: a systematic review of international guidelines and consensus statements. *Arthritis Care Res (Hoboken)*. 2017;69:1134–1141.
14. Badsha H, Edwards CJ. Intravenous pulses of methylprednisolone for systemic lupus erythematosus. *Seminars in Arthritis & Rheumatism*. 2003;32:370–377.
15. Vasoo S, Thumboo J, Fong KY. Refractory immune thrombocytopenia in systemic lupus erythematosus: response to mycophenolate mofetil. *Lupus*. 2003;12:630–632.
16. Bartholome B, Spies CM, Gaber T, et al. Membrane glucocorticoid receptors (mGCR) are expressed in normal human peripheral blood mononuclear cells and up-regulated after in vitro stimulation and in patients with rheumatoid arthritis. *FASEB Journal*. 2004;18:70–80.
17. Buttgereit F, Burmester GR, Straub RH, Seibel MJ, Zhou H. Exogenous and endogenous glucocorticoids in rheumatic diseases, Arthritis &. *Rheumatism*. 2011;63:1–9.
18. Hardy R, Rabbitt EH, Filer A, et al. Local and systemic glucocorticoid metabolism in inflammatory arthritis. *Annals of the Rheumatic Diseases*. 2008;67:1204–1210.
19. Buttgereit F, Zhou H, Kalak R, et al. Transgenic disruption of glucocorticoid signaling in mature osteoblasts and osteocytes attenuates K/BxN mouse serum-induced arthritis in vivo. *Arthritis & Rheumatism*. 2009;60:1998–2007.
20. Pincus T, Sokka T, Cutolo M. The past versus the present, 1980-2004: reduction of mean initial low-dose, long-term glucocorticoid therapy in rheumatoid arthritis from 10.3 to 3.6 mg/day, concomitant with early methotrexate, with long-term effectiveness and safety of less than 5 mg/day. *Neuroimmunomodulation*. 2015;22:89–103.
21. Kirwan JR, Bijlsma JW, Boers M, Shea BJ. Effects of glucocorticoids on radiological progression in rheumatoid arthritis. *Cochrane Database of Systematic Reviews*. 2007:CD006356.
22. Graudal N, Jurgens G. Similar effects of disease-modifying antirheumatic drugs, glucocorticoids, and biologic agents on radiographic progression in rheumatoid arthritis: meta-analysis of 70 randomized placebo-controlled or drug-controlled studies, including 112 comparisons. *Arthritis & Rheumatism*. 2010;62:2852–2863.
23. Gaffney K, Ledingham J, Perry JD. Intra-articular triamcinolone hexacetonide in knee osteoarthritis: factors influencing the clinical response. *Annals of the Rheumatic Diseases*. 1995;54:379–381.
24. Joseph RM, Hunter AL, Ray DW, Dixon WG. Systemic glucocorticoid therapy and adrenal insufficiency in adults: a systematic review. *Seminars in Arthritis & Rheumatism*. 2016;46:133–141.
25. van der Goes MC, Jacobs JW, Boers M, et al. Monitoring adverse events of low-dose glucocorticoid therapy: EULAR recommendations for clinical trials and daily practice. *Annals of the Rheumatic Diseases*. 2010;69:1913–1919.
26. Anwar MA, Saleh AI, Al Olabi R, Al Shehabi TS, Eid AH. Glucocorticoid-induced fetal origins of adult hypertension: association with epigenetic events. *Vascul Pharmacol*. 2016;82:41–50.
27. Caplan A, Fett N, Rosenbach M, Werth VP, Micheletti RG. Prevention and management of glucocorticoid-induced side effects: a comprehen-

sive review: gastrointestinal and endocrinologic side effects. *J Am Acad Dermatol*. 2017;76:11–16.
28. Singh JA, Hossain A, Kotb A, Wells G. Risk of serious infections with immunosuppressive drugs and glucocorticoids for lupus nephritis: a systematic review and networ k meta-analysis. *BMC Med*. 2016;14:137.
29. Black RJ, Hill CL, Lester S, Dixon WG. The association between systemic glucocorticoid use and the risk of cataract and glaucoma in patients with rheumatoid arthritis: a systematic review and meta-analysis. *PLoS ONE [Electronic Resource]*. 2016;11:e0166468.
30. Kleiman A, Tuckermann JP. Glucocorticoid receptor action in beneficial and side effects of steroid therapy: lessons from conditional knockout mice. *Molecular & Cellular Endocrinology*. 2007;275:98–108.
31. Schacke H, Zollner TM, Docke WD, et al. Characterization of ZK 245186, a novel, selective glucocorticoid receptor agonist for the topical treatment of inflammatory skin diseases. *British Journal of Pharmacology*. 2009;158:1088–1103.
32. Buttgereit F, Spies CM, Bijlsma JW. Novel glucocorticoids: where are we now and where do we want to go? *Clinical & Experimental Rheumatology*. 2015;33:S29–S33.

84

Immunomodulating Drugs

Gideon P. Smith and Edwin S.L. Chan

Excitement over biologic agents and their capacity to regulate immunological reactions that significantly impact immunologically mediated diseases, such as rheumatoid arthritis (RA) and inflammatory bowel disease (IBD), have overshadowed the older, small-molecule therapeutic agents. Nonetheless, when tested head-to-head, some small-molecule agents (most notably methotrexate) have proven to be almost as effective as biologics; moreover, combining small-molecule therapies with biological agents generally leads to significantly better outcomes than the use of either agent alone.[1] Given this synergy, and the fact that they are often much more cost-effective, it is likely that small-molecule immunomodulatory drugs will continue to be used. Here, we review the most widely accepted and commonly used immunomodulators currently available.

METHOTREXATE

Methotrexate (Fig. 84.1) was employed in the treatment of RA as early as 1951, but its popularity with regard to RA did not come until the 1980s. Over the years, extensive experience with its use in inflammatory diseases as diverse as RA (Chapter 53), psoriasis (Chapter 64), and IBD (Chapter 75) has taught us a great deal about its safety, efficacy, and toxicity, as well as its anti-inflammatory mechanisms of action. In this respect, methotrexate, much as corticosteroids, can be justly regarded as a cornerstone of immunomodulatory therapy.

Pharmacokinetics of Methotrexate

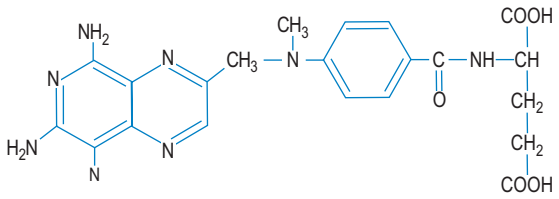
As an anti-inflammatory agent, methotrexate is administered at low doses (usually 10 to 25 mg/week) once weekly, usually orally, but it can also be given subcutaneously or intramuscularly. At these doses, oral bioavailability is high (60% to 70%), and although transporters are responsible for its absorption from the gastrointestinal (GI) tract, saturation effect does not occur. A small portion of methotrexate is metabolized by hydroxylation into 7-hydroxymethotrexate. Both compounds have a serum half-life of no more than 8 hours. The much longer anti-inflammatory action, which allows for once-weekly dosing, must therefore be mediated by other longer-lasting metabolites, such as polyglutamates. Excretion occurs principally via the urinary tract, but also via the biliary tract. Therefore renal function is an important consideration in methotrexate dosing, and any medication that impairs glomerular filtration may also potentiate methotrexate's effectiveness and toxicity.[2]

Mechanisms of Action for Methotrexate

As an analogue of folic acid, methotrexate is an inhibitor of purine and pyrimidine synthesis and thereby suppresses cellular proliferation (Table 84.1). These actions are dependent on inhibition of dihydrofolate reductase; hence toxicities arising from high-dose methotrexate therapy can be treated with folic acid derivatives, such as leucovorin. However, folic or folinic acid, which are often given in conjunction with methotrexate in inflammatory diseases to reduce the incidence of mucositis and bone marrow suppression, does little to inhibit its anti-inflammatory efficacy. Decreases in purine and pyrimidine concentrations in the serum have been observed following a single dose of methotrexate, along with a decreased proliferation of antigen-stimulated lymphocytes. However, these changes are transient and insufficient to explain the anti-inflammatory effects of once-weekly dosing. This, as well as the low doses of methotrexate required to produce an anti-inflammatory effect, suggests that the anti-inflammatory actions are mediated via different mechanisms.

One alternate mechanism is that methotrexate blocks intracellular transmethylation reactions and inhibits the production of *S*-adenosylmethionine. Since *S*-adenosylmethionine is necessary for the formation of the toxic polyamine metabolites spermine and spermidine, their accumulation at the inflammatory site is prevented. This inhibition of transmethylation is associated with an impairment of monocyte and lymphocyte function and thus potentially the synthesis of reactive oxygen species. However, diminution of transmethylation by the use of the *S*-adenosylhomocysteine hydrolase inhibitor, deaza-adenosine, has failed to produce any beneficial clinical effects in RA.

Methotrexate and its long-acting polyglutamate metabolites also exert anti-inflammatory effects by releasing the endogenous autocoid adenosine.[3] As potent inhibitors of the enzyme 5-aminoimidazole-4-carboxamide ribonucleotide (AICAR) transformylase, methotrexate polyglutamates promote the accumulation of AICAR in tissues. Since AICAR inhibits catabolizing enzymes for both adenosine and adenosine monophosphate (AMP), which can be dephosphorylated to adenosine, the net effect is intracellular and extracellular increases in adenosine levels. These metabolic pathways are pharmacologically relevant since aminoimidazole carboxamide and adenosine have been shown to be increased in urine following low-dose methotrexate treatment in patients with psoriasis.[4] Adenosine causes diminution of neutrophil accumulation, adhesion, phagocytosis, and generation of reactive oxygen species, inhibition of adhesion molecule expression, suppression of pro-inflammatory cytokines, and induction of anti-inflammatory cytokines, as well as modulation of macrophage and endothelial function.[2] Indeed, blockade of adenosine receptors reversed the anti-inflammatory effects of methotrexate in animal models. It



Methotrexate

FIG. 84.1 Methotrexate—Chemical Structure.

TABLE 84.1 Methotrexate: Mechanisms of Action

Suggested Mechanism	Rationale
Folate antagonism	Prevents purine and pyrimidine synthesis required for the proliferation of actively dividing immune cells, such as lymphocytes
Inhibition of spermine and spermidine production	Reduces formation of polyamines harmful to tissues
Alteration of cellular redox state	Reversible inhibition of lymphocyte and macrophage functions
Release of adenosine	Generation of a potent endogenous anti-inflammatory mediator through inhibition of catabolism of both adenosine and adenosine monophosphate

has also been suggested that caffeine, itself a nonselective antagonist of adenosine receptors, may both reduce the effectiveness of methotrexate in RA and protect against the development of cirrhosis of the liver, a major side effect of methotrexate.[5]

THERAPEUTIC PEARLS

Methotrexate

- Proven safety profile
- Concomitant administration of folic acid advisable
- Anti-inflammatory effects may be reduced by heavy use of caffeine
- Hepatotoxicity a rare but real concern
- Teratogenic

Risk of hepatotoxicity increased with:

- Alcohol use
- Hepatitis
- Diabetes
- Obesity
- Alpha-1-antitrypsin deficiency

Adverse Effects

Over decades, methotrexate has proven to be one of the safest disease-modifying antirheumatic drugs (DMARDs). Serious side effects such as cirrhosis are much less common than previously thought (Table 84.2). The use of folic acid has decreased the occurrence of mucosal and GI side effects, without limiting its anti-inflammatory activity, and cytopenias are managed adequately with regular blood counts. Although side effects, such as nausea and vomiting, may resolve spontaneously or respond to dose reduction or folic acid supplementation, mild

TABLE 84.2 Methotrexate: Adverse Effects

Gastrointestinal
Stomatitis
Anorexia
Nausea
Vomiting
Diarrhea
Cirrhosis
Pancreatitis

Hematological
Leukopenia
Anemia
Thrombocytopenia
Hypogammaglobulinemia
Lymphoma

Cardiovascular
PericarditisThrombosis

Pulmonary
Pulmonary fibrosis
Interstitial pneumonitis

Others
Skin rashes
Renal failure
Abortion
Impotence
Headache
Opportunistic infections

transaminasemia has rarely necessitated discontinuation of the medication. The risk of serious hepatotoxicity over 5 years of use is likely to be less than 1 in every 1000 patients with RA but may be more common in psoriasis patients. Risk factors, such as alcohol consumption, hepatitis B and C, diabetes, obesity, and alpha-1-antitrypsin deficiency, identify patients more likely to develop a methotrexate-induced hepatic injury. However, other serious side effects, such as pneumonitis, may be overlooked, since early symptoms (mild cough or shortness of breath) are often ignored and no serologic screen exists. Early identification allows for prompt discontinuation. The risk of developing solid tumors is debated, since the risk of malignancies is intrinsic to some of the conditions, such as RA, for which methotrexate is used. It is likely the risk of drug-induced malignancy in some patients is real since reports have documented tumor regression following discontinuation of methotrexate. But the risk remains extremely small.

SULFASALAZINE

Sulfasalazine (Fig. 84.2) was originally introduced in the late 1930s for the treatment of RA but is now used in a wide range of inflammatory diseases, in particular, IBD and the seronegative arthritides. It consists of a derivative of anti-inflammatory salicylic acid, 5-amino-salicylic acid, and the antimicrobial sulfapyridine. These two moieties are joined together by an azo bond. Which component is responsible for the drug's anti-inflammatory actions is unclear, but it appears to vary according to disease state. For instance, in IBD, 5-amino-salicylic acid is likely the main active component, as it is poorly absorbed following metabolism by intestinal flora. In inflammatory arthritides, sulfapyridine is likely to play a more important role, as it is relatively well absorbed and has a bioavailability in the region of 60%. Since acetylation is the principal route of the metabolism of sulfapyridine following absorption, acetylator status is a major determinant of the plasma half-life. Similarly, slow acetylators are more liable to develop side effects.

Mechanisms of Action for Sulfasalazine

Sulfasalazine has several immunomodulatory effects. Lymphocyte proliferation is suppressed *in vitro* and involves both B-cell and T-cell populations. *In vivo*, a decrease in activated

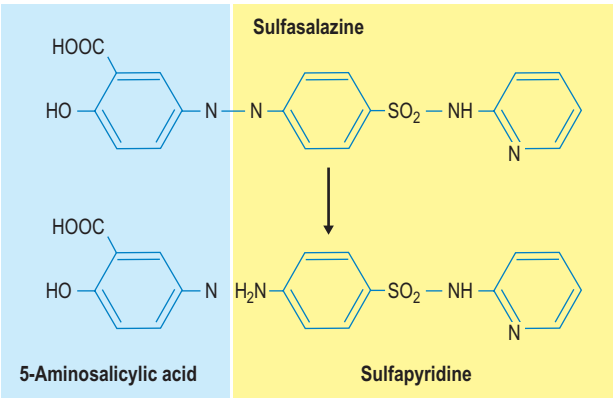


FIG. 84.2 Sulfasalazine—Chemical Structure.

lymphocytes in the peripheral blood is also seen. Tumor necrosis factor (TNF) production is suppressed, and receptor binding is inhibited. Sulfasalazine also inhibits activation of the transcription factor, nuclear factor (NF)-κB. Like methotrexate, sulfasalazine inhibits AICAR transformylase and thus promotes the accumulation of adenosine and its anti-inflammatory actions via the adenosine A_{2A} receptor. Indeed, treatment of animals with an adenosine A_{2A} receptor antagonist reversed sulfasalazine's reduction of leukocyte accumulation in an air-pouch model of inflammation.

KEY CONCEPT

Sulfasalazine: Mechanism of Action

- Suppresses the proliferation of lymphocytes
- Suppresses proinflammatory cytokine production
- Inhibits activation of nuclear factor (NF)-κB
- Promotes adenosine accumulation

Adverse Effects

In a large series, a quarter of those treated over 11 years stopped treatment because of toxicity.[6] Most toxicities occurred early and were both trivial and resolved following therapy withdrawal. Most common are nausea, vomiting, anorexia, and rash. Serious cutaneous reactions, such as toxic epidermal necrosis, are rare. Transaminasemia and drug-induced hepatitis can occur. Blood dyscrasias with megaloblastic anemia, neutropenia, aplastic anemia, and myelodysplastic syndrome may arise. Neurological adverse effects include headache and dizziness or, more seriously, peripheral neuropathy, Guillain-Barré syndrome, or transverse myelitis. Sulfasalazine should be avoided in patients with sulfa allergy, and glucose-6-phosphate dehydrogenase (G6PD). Deficiency screening should be performed before prescribing.

AZATHIOPRINE

Azathioprine, an imidazolyl derivative of 6-mercaptopurine (Fig. 84.3), has been widely used in RA and IBD as well as in solid organ transplantations. Cleavage into 6-mercaptopurine and the imidazole moiety occurs rapidly within erythrocytes, both enzymatically by glutathione transferase and nonenzymatically. Several enzymes (Table 84.3) participate in the metabolism of

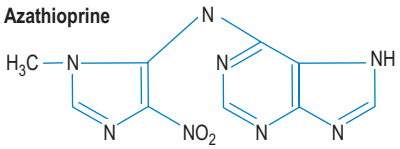


FIG. 84.3 Azathioprine—Chemical Structure.

TABLE 84.3 Principal Enzymes Involved in the Metabolism of Azathioprine

Enzyme	Action
Glutathione transferase	Cleaves azathioprine into 6-mercaptopurine and imidazole moieties
Thiopurine methyltransferase	Metabolism of 6-mercaptopurine
Xanthine oxidase	Conversion of 6-mercaptopurine to 6-thiouric acid

6-mercaptopurine into active and inactive compounds. One of these, thiopurine methyltransferase, is associated with genetic polymorphisms; inherited changes in its activity may impact patient response to azathioprine. Xanthine oxidase inactivates 6-mercaptopurine by converting it to 6-thiouric acid. Since this occurs mainly in the liver, toxicity from azathioprine therapy is a danger in enzyme deficiency states, as a result of disease or the use of drugs, such as allopurinol.

Proposed Mechanisms of Action for Azathioprine

The immunomodulatory mechanism of azathioprine remains unclear. As purine analogues, the active metabolites interfere with the salvage pathway and *de novo* synthesis of purines, and they are incorporated into RNA and DNA. The proliferation of T and B lymphocytes is inhibited, and the function of natural killer (NK) cells is suppressed without any change in cell numbers. The production of antibodies is also suppressed, although it is not known which of these effects predominate *in vivo*. Cellular responses to chemoattractants are altered, and the production of cytokines, such as interleukin-6 (IL-6), is also affected.

Adverse Effects

Azathioprine is generally well tolerated. The most common side effects are mild and affect the GI system. Pancreatitis can occur as an idiosyncratic reaction. Hepatotoxicity and cholestasis are not uncommon, and hepatic peliosis and nodular regenerative hyperplasia occur rarely. There have been reports of possible heightened risk for non-Hodgkin lymphoma, but because of the rarity of these events, no definite link has been established. Bone marrow suppression and opportunistic infections pose far greater threats.

CYCLOPHOSPHAMIDE

Alkylating agents were used in the treatment of inflammatory diseases after promising reports on using nitrogen mustard in RA. Cyclophosphamide (Fig. 84.4) is metabolized to produce the alkylating agent phosphoramide mustard as well as acrolein, which, although inactive, results in the hemorrhagic cystitis

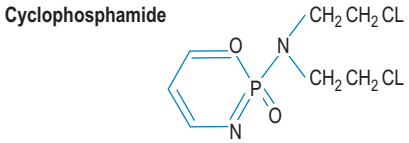


FIG. 84.4 Cyclophosphamide—Chemical Structure.

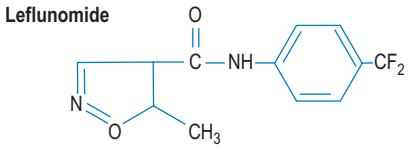


FIG. 84.5 Leflunomide—Chemical Structure.

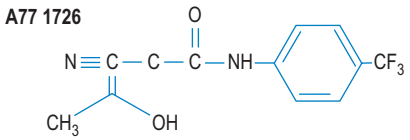


FIG. 84.6 A77 1726—Chemical Structure.

associated with cyclophosphamide. Cyclophosphamide can be given intravenously, but bioavailability following oral administration is high (>75%). Toxicity has severely limited its use in inflammatory diseases, although its contribution to the management of lupus nephritis cannot be denied. The alkylating actions occur at guanine residues (principally on DNA, but also on RNA) resulting in cross-linkable strands and disruption of transcription and translation.

Mechanisms of Action of Cyclophosphamide

This alkylating process has immunomodulating effects on resting and actively dividing cells. The numbers of circulating CD4 T lymphocytes and, to a lesser extent, CD8 T lymphocytes are reduced, thus reducing the CD4/CD8 ratio. Despite an apparent increase in immunoglobulin-secreting cells, B-cell function is suppressed, and overall immunoglobulin synthesis is reduced.

Adverse Effects

The best-known toxicity is hemorrhagic cystitis. Since this occurs more frequently following oral dosing, this route of administration is rarely used. This may relate to continuous exposure of the bladder to acrolein, so the acrolein-neutralizing agent 2-mercaptoethane sulfonate (Mesna) is used prophylactically along with copious hydration. Hemorrhagic myocarditis can also occur and may cause myocardial necrosis, hemopericardium, and congestive cardiac failure. However, survivors of acute cardiac toxicity do not show any residual electrocardiographic or echocardiographic abnormalities.

Besides bone marrow suppression, reduction of fertility, and a heightened risk of infection, cyclophosphamide therapy has been associated with secondary malignancies, which may occur years after drug cessation. Malignancies of the bladder, often of a transitional cell type, tend to occur only in those with a history of treatment-related hemorrhagic cystitis. Myeloproliferative and lymphoproliferative disorders have also been associated with cyclophosphamide use.

Other Nitrogen Mustard Derivatives

Chlorambucil, or 4-[bis(2chlor-ethyl)amino]benzenebutanoic acid, has a wide distribution in tissue and 87% oral bioavailability, but unlike cyclophosphamide, it does not require metabolism by the liver to become metabolically active.[7] The only indication approved by the US Food and Drug Administration (FDA) is for the treatment of chronic lymphocytic leukemia (CLL), but like cyclophosphamide, chlorambucil has been reported to be used in treatments for the same wide range of inflammatory conditions. The mechanism of action and side effect profile is also similar to those of cyclophosphamide, but with a higher risk of permanent aplasia.[8]

Melphalan is another phenylalanine derivative of nitrogen mustard, mainly used in treating multiple myeloma. It has been less widely adopted but has been used off-label in a variety of inflammatory conditions. Adverse effects and mechanisms are the same as those of cyclophosphamide, as described above.

LEFLUNOMIDE

Leflunomide is an inhibitor of *de novo* pyrimidine synthesis. Leflunomide (Fig. 84.5) is converted into the long-acting active compound A77 1726 (2-cyano-3-hydroxy-N-[4-trifluoromethyl]-butenamide) (Fig. 84.6), a reversible inhibitor of the enzyme dihydroorotate dehydrogenase that is involved in pyrimidine synthesis. Because of this long half-life, therapy with leflunomide is usually started with a loading dose to quickly achieve therapeutic levels. A77 1726 is highly plasma-protein bound and undergoes enterohepatic recirculation.

Mechanisms of Action of Leflunomide

By inhibiting pyrimidine synthesis, pyrimidine nucleotide availability becomes insufficient for the proliferation of immune-response cells. This deficiency is inadequately replenished by the salvage pathways, rendering cell proliferation inefficient and limiting the clonal expansion of T cells. B-cell proliferation is similarly suppressed with the reduction of Cdk2, a cyclin-dependent kinase. Leflunomide also inhibits NF-κB activation. Although the effects of moderate concentrations are reversed by uridine *in vitro*, this reversal does not occur at higher concentrations, suggesting the possible involvement of other mechanisms. Leflunomide is known to inhibit tyrosine kinase activity at higher concentrations, although the relevance of this effect to therapeutic concentrations achievable *in vivo* remains questionable.

Adverse Effects

GI symptoms are the most common side effect, and hepatic damage is the most important toxicity. Although there are similarities to the toxicity profile of methotrexate, clinical trials have shown that leflunomide and methotrexate can be safely and effectively given together to patients with RA, but transaminasemia occurs more often than with methotrexate alone.[9,10] Fulminant hepatic failure is rare, but fatal cases have occurred. Skin reactions are mostly minor; however, more serious manifestations, such as toxic epidermal necrolysis, have been reported.

MYCOPHENOLATE MOFETIL

Mycophenolate mofetil (Fig. 84.7) is widely used in solid organ transplantation and has also been increasingly employed in the treatment of autoimmune diseases. It is rapidly absorbed and hydrolyzed into the active compound, mycophenolic acid, which is a reversible inhibitor of inosine monophosphate dehydrogenase. Since inosine monophosphate dehydrogenase is a key enzyme in the *de novo* synthesis of guanine nucleotides, its inhibition is most significant in T and B lymphocytes, which are reliant on this pathway, as they lack the hypoxanthine-guanine phosphoribosyltransferase salvage pathway. The immunological effects are numerous. DNA synthesis in lymphocytes requires the incorporation of guanine nucleotides so that proliferation of lymphocytes is suppressed. Antibody production and NK-cell activity are also reduced, and *in vitro* cytokine production by activated human mononuclear cells is affected.[11] In addition, delayed-type hypersensitivity responses are suppressed. Although effective in a subset of patients with psoriasis and RA, mycophenolate mofetil has not been widely used in these conditions because other more effective medications are available. It is, however, becoming more popular in the treatment of some diseases, such as myositis, systemic contact dermatitis, severe atopic dermatitis, chronic urticaria, refractory pyoderma gangrenosum, bullous pemphigoid, pemphigus vulgaris, and pemphigus foliaceus, where it is effective with a low risk of side effects.

Adverse Effects

Absolute contraindications for mycophenolate mofetil are drug allergy and pregnancy (Category D). Relative contraindications include lactation; renal, hepatic, or cardiopulmonary disease; and peptic ulcer. It is generally well tolerated when used in autoimmune diseases, such as RA. The most common side effects are nausea, vomiting, abdominal discomfort, diarrhea, fever, headache, skin rash, back pain, and tremor, but these do not usually lead to discontinuation. Rarely reported side effects include leukopenia and other cytopenias, cutaneous and noncutaneous malignancies, and pancreatitis. Toxic doses have not been established for this medication. One patient suffered only moderate leukopenia with no significant GI side effects after ingesting 25 g of mycophenolate mofetil. Up to 4 g/day have been used in cardiac and up to 5 g/day in hepatic transplantation patients. However, increased efficacy was not observed above 2 g/day, and patients were more likely to experience GI symptoms and neutropenia at higher doses. For this reason, doses up to 2 g/day are usually employed to treat inflammatory conditions.

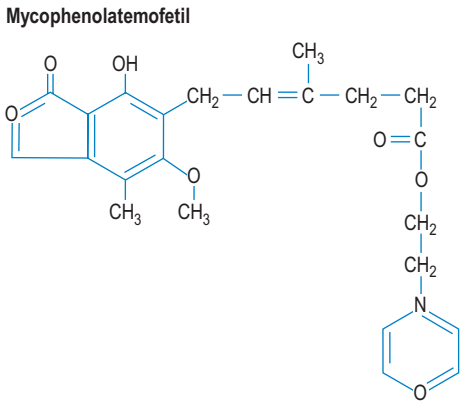


FIG. 84.7 Mycophenolate Mofetil—Chemical Structure.

HYDROXYUREA

Hydroxyurea is urea with one additional hydroxyl group. It inhibits ribonucleotide reductase, which catalyzes the reduction of ribonucleotides to deoxyribonucleotides, and is thus essential in DNA synthesis. It is effective in the treatment of psoriasis.[12] Hydroxyurea is well tolerated, with the most common side effects being hematological, usually megaloblastic anemia, but also leukopenia and thrombocytopenia. Other significant but rare adverse effects include renal and GI toxicity, a dermatomyositis-like syndrome, leg ulcers, radiation recall, and leukemias.

ORAL CYCLOSPORINE AND TACROLIMUS (FK506)

Cyclosporine (Fig. 84.8) and tacrolimus (Fig. 84.9) are structurally similar drugs that have been widely used in solid organ transplantation as well as in the treatment of immunological diseases. Cyclosporine has potent inhibitory effects in dampening the

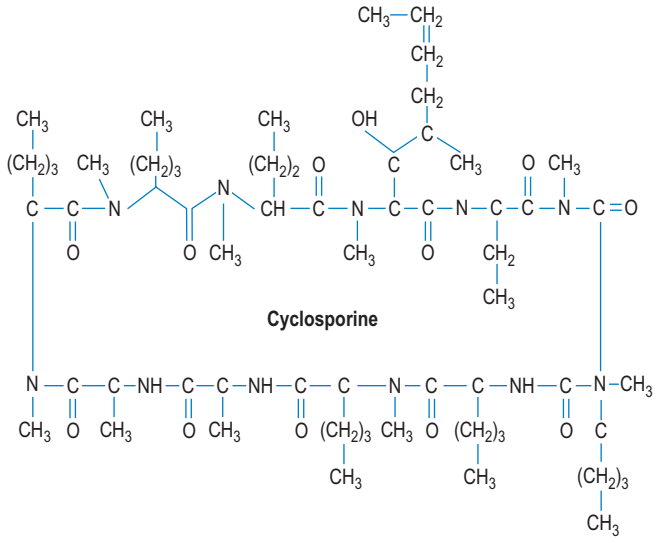


FIG. 84.8 Cyclosporine—Chemical Structure.

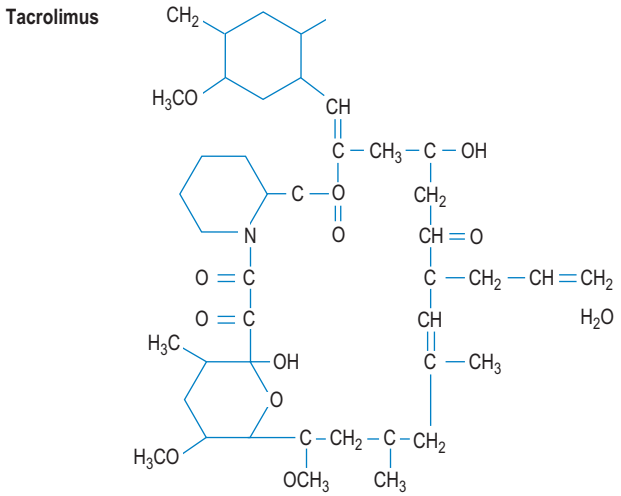


FIG. 84.9 Tacrolimus—Chemical Structure.

production of proinflammatory mediators, such as IL-2, by immunocompetent cells, most importantly T lymphocytes. It does so through binding cyclophilin, which produces a cyclosporine–cyclophilin complex. This complex binds the serine/threonine phosphatase calcineurin.[13] This disrupts the phosphorylation of regulatory proteins for which the nucleated factor of activated T cells (NF-ATs) is a critical component, preventing these proteins from translocating into the nucleus. Thus the transcription of genes, such as *IL-2*, which induces mitogenesis in activated T cells, cannot be effectively activated. Several other cytokines are affected, including IL-3, IL-6, transforming growth factor (TGF)-β, and interferon (IFN)-γ. Another T cell–specific immunophilin, FK506-binding protein (FKBP), binds tacrolimus to form a FK506-FKBP complex with similar resultant inhibitory activity on calcineurin.[14]

KEY CONCEPT

Cyclosporine: Mechanism of Action

- Association with cyclophilin
- Formation of cyclosporine–cyclophilin complex
- Binds calcineurin
- Inactivates calcineurin
- Regulatory proteins unable to translocate into nucleus
- Transcription of proinflammatory genes affected

Adverse Effects

The most common side effects of cyclosporine are hypertension, hyperkalemia, hypomagnesemia, and hyperlipidemia. More importantly, cyclosporine has well-documented short-term and long-term adverse effects on kidney function.[15] These data come from patients who had received solid organ transplants, most notably renal transplants. In such patients, initial doses of 15 to 25 mg/kg/day led to a reduction in the glomerular filtration rate (GFR) and rise in serum creatinine in a percentage of patients, as well as histologically proven nephropathy.[16] The pathogenic mechanism of kidney damage is poorly understood but is believed to consist of two phases.[17] The first phase is a period of partial ischemia secondary to vascular contraction, which is reversible with dose reduction or drug discontinuation. The later, irreversible phase results from chronic scarring of the glomeruli. Treatment recommendations for some diseases, such as psoriasis, have therefore been targeted at much lower maximum daily doses of 5 mg/kg/day, with a reduced dose if creatinine rises 30% above baseline. Otherwise healthy patients treated at these dosage levels have been successfully managed for many years on cyclosporine with no impact on the GFR.[18] Nephrotoxicity has also been a concern during tacrolimus therapy. Other adverse effects include increased rate of infections, malignancy, hepatotoxicity, GI upset, rash, tremor, headache, and insomnia.

TOPICAL PIMECROLIMUS AND TACROLIMUS (FK506)

Tacrolimus is also available in a topical formulation. Pimecrolimus is an alternative topical calcineurin inhibitor with a similar structure and identical mechanisms of action. The cyclosporine molecule is too large to penetrate the skin (1203 Da), whereas tacrolimus and pimecrolimus can, as they are much smaller molecules (804 and 80 daltons [Da], respectively). Both have been approved by the FDA for the treatment of atopic dermatitis but have found widespread use in many other conditions (psoriasis, oral and cutaneous lichen planus, vitiligo, pemphigoid, and pemphigus). They are most often used in patients or on body areas where long-term topical corticosteroids are contraindicated.

Adverse Effects

The most common side effect is local irritation at the site of application in severely inflamed skin. Therefore initial short-term treatment is often combined with topical steroids. An association with malignancy remains controversial. In 2005, 17 case reports of malignancy in patients using these topicals resulted in a black-box warning. Although further studies suggested that the rate of lymphoma from these case reports was lower than the rate observed in the general population in the United States. In 2006, the FDA revised the wording, but a black-box warning about these products remains in place.

SIROLIMUS

Sirolimus is a macrolide that binds the cytosolic protein FK-binding protein 12 (FKBP12). In contrast to the tacrolimus–FKBP12 complex, which inhibits calcineurin, the sirolimus–FKBP12 complex directly binds the mammalian target of rapamycin (mTOR) complex1 (mTORC1), thereby inhibiting the mTOR pathway. Thus it inhibits the response to IL-2, blocking the activation of T and B cells. It has shown promise in the treatment for systemic lupus erythematosus (SLE); Sjögren syndrome[19]; RA[20]; psoriasis; genetic disorders, such as tuberous sclerosis; and neoplastic disorders, such as Kaposi sarcoma.[21] The main advantage of sirolimus over tacrolimus or cyclosporine is reduced renal toxicity.

IMIQUIMOD

Imiquimod (Fig. 84.10), an imidazoquinoline drug, activates toll-like receptor (TLR)-7 and possesses both antiviral and antitumor activities.[22] As a cream preparation, it is effective in the treatment of plantar or external genital warts caused by infection with human papillomavirus (HPV). Immune amplification responses are induced through the stimulation of inflammatory cytokines.[23,24] Production of IFN-α is stimulated, and this suppresses replication of viruses in infected keratinocytes. NK-cell activity is also increased, partly through the

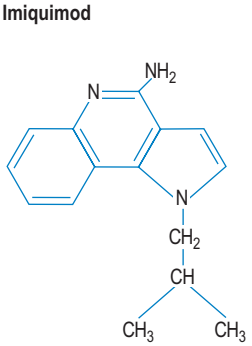


FIG. 84.10 Imiquimod—Chemical Structure.

induction of oligoadenylate synthase. The increase of dermal IFN-α transcript levels is rapid and dramatic.[25] Other cytokines modulated by imiquimod include TNF and IL-12, especially in peripheral blood monocytes.[26,27] The overall effect is a shift from a T-helper-2 (Th2)-cytokine–predominant profile toward a Th1-cytokine–predominant profile. Other conditions where imiquimod is effective include actinic keratosis, lentigo maligna, superficial basal cell carcinoma, and molluscum contagiosum.

Adverse Effects

Inflammation at the site of application represents the most common adverse reaction. However, nondermatological side effects, such as fever, fatigue, myalgia, and headache, have also been described usually when applied over large surface areas.

5-FLUOROURACIL

5-Fluorouracil is a uracil analogue that has two modes of action. First, it inhibits cell proliferation via direct incorporation into RNA causing abnormal base pairing. Second, it binds thymidylate synthetase, blocking the conversion of deoxyuridine monophosphate to deoxythymidine monophosphate, which is essential to DNA synthesis. It may also increase the expression of *p53*, a frequently mutated gene in nonmelanoma skin cancers.[28] It can be administered topically, intramuscularly, or intravenously. In topical form, its main uses are to treat actinic keratosis, superficial basal cell carcinoma, Bowen disease, keratoacanthoma, porokeratosis, and verruca vulgaris. However, intravenous treatment of recalcitrant psoriasis, mycosis fungoides, and scleroderma has also been reported.

Adverse Effects

Topical application is associated with an irritant dermatitis, but this is also seen as a marker of clinical efficacy. Parenteral administration for inflammatory conditions is not widespread; adverse effects from parenteral administration are more severe and include clinically significant bone marrow suppression, GI toxicity, and cutaneous reactions.

GLATIRAMER

Glatiramer acetate is a random polymer of glutamic acid, lysine, tyrosine, and alanine; the four amino acids found in myelin basic protein (MBP). The mechanism of action for glatiramer is unknown, but its similarity in structure to MBP may allow it to act as a decoy for immune targeting of myelin. It may also induce the expression of glatiramer acetate-specific suppressor T cells, and these are present in animal models. In contrast to imiquimod, glatiramer shifts the population of T cells from proinflammatory Th1 cells to regulatory Th2 cells that suppress the inflammatory response. It is FDA approved for the treatment of adults with relapsing-remitting multiple sclerosis (MS), even after only one event.

Adverse Effects

Absolute contraindications include allergy to glatiramer or mannitol. It is pregnancy Category B, but it is unknown if the drug is secreted in breast milk. The drug is given via subcutaneous injection, and the most common adverse effects are injection site reaction, flushing, rash, dyspnea, and transient chest pain.

FINGOLIMOD (FTY720)

Fingolimod is a relatively new immunomodulator for treating MS. It is a structural analogue of sphingosine and is phosphorylated intracellularly by sphingosine kinases.[29] Signaling via one of the sphingosine 1 phosphate receptors, S1PR1, it is believed to prevent migration of lymphocytes by halting their ability to leave lymph nodes. However, fingolimod has also been reported to act as a cPLA2 inhibitor,[30] a cannabinoid receptor antagonist,[31] and a ceramide synthase inhibitor.[32] So far, the FDA has only approved it for use in MS. However, it has also shown promise in murine models of SLE and RA, with reduced mortality from lupus nephritis[33] and improvement in arthritis, respectively.[34] It also has a potential role in the treatment of cutaneous inflammatory conditions, such as psoriasis and atopic dermatitis.[35]

Adverse Effects

The most common side effects are minor and include headache and fatigue. However, fingolimod has also been associated with serious infections, an increased rate of skin cancers, bradycardia, and one case of focal hemorrhagic encephalitis.

CONCLUSIONS

The field of small molecule immunomodulators contains both traditional and new molecules. The traditional molecules come with a long history of efficacy and side-effect data, and this allows us to safely and accurately utilize these therapies. Despite the introduction of biologics, traditional small-molecule agents, such as methotrexate and cyclosporine, are both clinically efficacious and cost-effective and thus remain the workhorses of our modern pharmaceutical armamentarium. However, as our understanding of the pathways involved in inflammation evolve, our therapeutics are also refined, allowing the production now, for example, of topically effective calcineurin inhibitors, such as tacrolimus and pimecrolimus, which avoid the adverse effects of systemic cyclosporine or topical steroids, and sirolimus, which reduces toxicity profiles of systemic FKBP2 signaling. In addition, newer agents, such as glatiramer and fingolimod, allow treatment of MS, a disease recalcitrant to more traditional therapies, and, although not currently being used in other inflammatory disorders, have shown promise in animal models. The field of small-molecule immunomodulators, therefore, remains one of both current clinical relevance and continuing, exciting new developments.

ON THE HORIZON

- Refinement of our understanding of the pathways and actions of receptors targeted by our current medications is key to producing more elegant therapies or advancing our understanding of which therapeutics are most effective to combine.
- Separating therapeutic effects from the pathways leading to side effects would enable much safer and more tolerable small-molecule medications.
- Better understanding of the dysregulation that occurs in inflammatory conditions, and the activation of the immune system, will be essential to ensuring the efficacy of such new modalities.

The challenge in the next 5 to 10 years is to continue this refinement in the targeting of small molecules. Many of the agents we currently use target early parts of pathways or indiscriminately

trigger multiple receptors. This leads to a greater range of side effects and thus limits our ability to reach adequate dosing levels. A prime example of this is methotrexate, which leads to increased adenosine levels and thus triggers all four adenosine receptors. By developing more selective molecules targeting, for instance, just the adenosine A_{2A} receptor, not only would we avoid the side effects attributable to other adenosine receptors, but we might also avoid the effects of methotrexate on bone marrow and mucous membranes via inhibition of dihydrofolate reductase. In other medications, more specific targeting of pathways may similarly reduce side effects and make for more effective therapeutics. This will require a better understanding of the immune system and its dysregulation in disease, as well as the development and application of these medications to clinically relevant models.

REFERENCES

1. Pincus T, Furer V, Sokka T. Underestimation of the efficacy, effectiveness, tolerability, and safety of weekly low-dose methotrexate in information presented to physicians and patients. *Clin Exp Rheumatol*. 2010;28:S68–S79.
2. Chan ES, Cronstein BN. Molecular action of methotrexate in inflammatory diseases. *Arthritis. Res*. 2002;4:266–273.
3. Cronstein BN, Naime D, Ostad E. The anti-inflammatory mechanism of methotrexate. Increased adenosine release at inflamed sites diminishes leukocyte accumulation in an in vivo model of inflammation. *J Clin Invest*. 1993;92:2675–2682.
4. Baggott JE, Morgan SL, Sams WM, et al. Urinary adenosine and aminoimidazole carboxamide excretion in methotrexate-treated patients with psoriasis. *Arch Dermatol*. 1999;135:813–817.
5. Chan ES, Montesinos MC, Fernandez P, et al. Adenosine A(2A) receptors play a role in the pathogenesis of hepatic cirrhosis. *Br J Pharmacol*. 2006;148:1144–1155.
6. Amos RS, Pullar T, Bax DE, et al. Sulphasalazine for rheumatoid arthritis: toxicity in 774 patients monitored for one to 11 years *Br Med J (Clin Res Ed)*. 2931986420–423.
7. Lind MJ, Ardiet C. Pharmacokinetics of alkylating agents. *Cancer Surv*. 1993;17:157–188.
8. Wolverton SE, Remlinger K. Suggested guidelines for patient monitoring: hepatic and hematologic toxicity attributable to systemic dermatologic drugs. *Dermatol Clin*. 2007;25:195–205. vi–ii.
9. Kremer JM, Genovese MC, Cannon GW, et al. Concomitant leflunomide therapy in patients with active rheumatoid arthritis despite stable doses of methotrexate. A randomized, double-blind, placebo-controlled trial. *Ann Intern Med*. 2002;137:726–733.
10. Weinblatt ME, Kremer JM, Coblyn JS, et al. Pharmacokinetics, safety, and efficacy of combination treatment with methotrexate and leflunomide in patients with active rheumatoid arthritis. *Arthritis Rheum*. 1999;42:1322–1328.
11. Nagy SE, Andersson JP, Andersson UG. Effect of mycophenolate mofetil (RS-61443) on cytokine production: inhibition of superantigen-induced cytokines. *Immunopharmacology*. 1993;26:11–20.
12. Menter A, Korman NJ, Elmets CA, et al. Guidelines of care for the management of psoriasis and psoriatic arthritis: section 4. Guidelines of care for the management and treatment of psoriasis with traditional systemic agents. *J Am Acad Dermatol*. 2009;61:451–485.
13. Liu J, Farmer Jr JD, Lane WS, et al. Calcineurin is a common target of cyclophilin-cyclosporin A and FKBP-FK506 complexes. *Cell*. 1991;66:807–815.
14. Baughman G, Wiederrecht GJ, Campbell NF, et al. FKBP51, a novel T-cell-specific immunophilin capable of calcineurin inhibition. *Mol Cell Biol*. 1995;15:4395–4402.
15. Wilkinson A, Ross EA, Hawkins R, et al. Measurement of true glomerular filtration rate in renal transplant patients receiving cyclosporine. *Transplant Proc*. 1987;19:1739–1741.
16. Mihatsch MJ, Antonovych T, Bohman SO, et al. Cyclosporin A nephropathy: standardization of the evaluation of kidney biopsies. *Clin Nephrol*. 1994;41:23–32.
17. Gaston RS. Chronic calcineurin inhibitor nephrotoxicity: reflections on an evolving paradigm. *Clin J Am Soc Nephrol*. 2009;4:2029–2034.
18. Kessel A, Toubi E. Cyclosporine-A in severe chronic urticaria: the option for long-term therapy. *Allergy*. 2010;65:1478–1482.
19. Perl A. Emerging new pathways of pathogenesis and targets for treatment in systemic lupus erythematosus and Sjogren's syndrome. *Curr Opin Rheumatol*. 2009;21:443–447.
20. Laragione T, Gulko PS. mTOR regulates the invasive properties of synovial fibroblasts in rheumatoid arthritis. *Mol Med*. 2010;16:352–358.
21. Paghdal KV, Schwartz RA. Sirolimus (rapamycin): from the soil of Easter Island to a bright future. *J Am Acad Dermatol*. 2007;57:1046–1050.
22. Hemmi H, Kaisho T, Takeuchi O, et al. Small anti-viral compounds activate immune cells via the TLR7 MyD88-dependent signaling pathway. *Nat Immunol*. 2002;3:196–200.
23. Dahl MV. Imiquimod: an immune response modifier. *J Am Acad Dermatol*. 2000;43:S1–S5.
24. Skinner Jr RB. Imiquimod. *Dermatol Clin*. 2003;21:291–300.
25. Imbertson LM, Beaurline JM, Couture AM, et al. Cytokine induction in hairless mouse and rat skin after topical application of the immune response modifiers imiquimod and S-28463. *J Invest Dermatol*. 1998;110:734–739.
26. Gibson SJ, Imbertson LM, Wagner TL, et al. Cellular requirements for cytokine production in response to the immunomodulators imiquimod and S-27609. *J Interferon Cytokine Res*. 1995;15:537–545.
27. Tyring S. Imiquimod applied topically: a novel immune response modifier. *Skin Therapy Lett*. 2001;6(6):1–4.
28. Ceilley RI. Mechanisms of action of topical 5-fluorouracil: review and implications for the treatment of dermatological disorders. *J Dermatolog Treat*. 2012;23:83–89.
29. Billich A, Bornancin F, Devay P, et al. Phosphorylation of the immunomodulatory drug FTY720 by sphingosine kinases. *J Biol Chem*. 2003;278:47408–47415.
30. Payne SG, Oskeritzian CA, Griffiths R, et al. The immunosuppressant drug FTY720 inhibits cytosolic phospholipase A2 independently of sphingosine-1-phosphate receptors. *Blood*. 2007;109:1077–1085.
31. Paugh SW, Cassidy MP, He H, et al. Sphingosine and its analog, the immunosuppressant 2-amino-2-(2-[4-octylphenyl]ethyl)-1,3-propanediol, interact with the CB1 cannabinoid receptor. *Mol Pharmacol*. 2006;70:41–50.
32. Berdyshev EV, Gorshkova I, Skobeleva A, et al. FTY720 inhibits ceramide synthases and up-regulates dihydrosphingosine 1-phosphate formation in human lung endothelial cells. *J Biol Chem*. 2009;284:5467–5477.
33. Ando S, Amano H, Amano E, et al. FTY720 exerts a survival advantage through the prevention of end-stage glomerular inflammation in lupus-prone BXSB mice. *Biochem Biophys Res Commun*. 2010;394:804–810.
34. Tsunemi S, Iwasaki T, Kitano S, et al. Effects of the novel immunosuppressant FTY720 in a murine rheumatoid arthritis model. *Clin Immunol*. 2010;136:197–204.
35. Herzinger T, Kleuser B, Schafer-Korting M, et al. Sphingosine-1-phosphate signaling and the skin. *Am J Clin Dermatol*. 2007;8:329–336.

85

Protein Kinase Antagonists

Arian Laurence, Massimo Gadina, Pamela L. Schwartzberg, and John J. O'Shea

Reversible protein phosphorylation is one of the major mechanisms controlling protein activity in all eukaryotic cells. As such, it is involved in all fundamental cellular processes, including cell cycle and cell growth, cell shape and movement, metabolism, differentiation and apoptosis. This covalent modification is an important means for transmitting information from outside the cell and between sub-cellular components within the cell and is a major component of signal transduction. Phosphorylation is a key mechanism underlying signaling in healthy cells, as exemplified by insulin and other growth factors, but in addition, the importance of protein phosphorylation is supported by evidence that mutations and dysregulation of protein kinases play causal roles in human disease. This is especially true in cancer, in which mutant protein kinases or their upstream activators function as oncogenes.

From the point of view of an immunologist, protein phosphorylation is a major mechanism by which immune receptors exert their effect. A critical first step in signaling by many cytokine receptors is the activation of phosphorylation. The receptors for classical growth factor cytokines, including stem cell factor and platelet-derived growth factor (PDGF) are receptor tyrosine kinases (RTKs), whereas the receptors for transforming growth factor family cytokines are receptor serine-threonine kinases. Type I and II cytokine receptors signal via the activation of receptor-associated Janus kinases. Other cytokines such as interleukin (IL)-1 and tumor necrosis factor (TNF) initiate proximal signaling in a kinase-independent manner, but nonetheless signal through kinase cascades to exert their effects (Chapter 14).

Other key immune receptors, including the T-cell receptor (TCR), B-cell receptor (BCR), and Fc receptors, also trigger signaling via kinases. As discussed in Chapter 4, the first step in signaling by these multi-chain immune recognition receptors (MIRRs) is tyrosine phosphorylation of the receptor itself and adapter molecules mediated by the Src family protein tyrosine kinase (PTK) (Fig. 85.1). This leads to the recruitment of the PTK members Syk (spleen tyrosine kinase) and Zap70 (zeta chain associated protein kinase 70) to invariant chains of MIRRs, followed by phosphorylation of adapters like SLP-76 and the activation of Tec (tyrosine kinase expressed in hepatocellular carcinoma) family PTKs. These initial steps lead to activation of serine-threonine kinases, including the protein kinase C (PKC) family and mitogen-activated protein kinases (MAPKs); many of these downstream effectors are used by both MIRRs and cytokine receptors. We now know a great deal of how the cascade of protein phosphorylation links events at the plasma membrane to calcium modulation, cytoskeletal rearrangement, gene transcription, and other canonical features of lymphocyte activation (Chapter 10).

The non-redundant functions of various kinases in different types of immune cells are best illustrated through studies of both gene-targeted knockout mice and humans with genetic diseases, making clear that protein phosphorylation is of major importance in immune and inflammatory mechanisms. Based on these genetic findings, targeting protein kinases was proposed to be a useful strategy in the development of novel immunosuppressant drugs and is one of the most active areas of pharmaceutical drug development. The field is now so vast that it is impractical to comprehensively review all this information in one chapter; therefore, we will focus on important historical precedents and then discuss drugs and targets most relevant for immune-mediated disease (Table 85.1). We will start by briefly reviewing some of the basics of kinase biochemistry.

STRUCTURE AND FUNCTION OF PROTEIN KINASES

Protein kinases, or phosphotransferases, catalyze the transfer of the γ-phosphate from a purine nucleotide triphosphate (*i.e.*, ATP and GTP) to the hydroxyl groups of their protein substrates. They generate phosphate monoesters using protein alcohol groups (on serine and threonine residues) and/or protein phenolic groups (on tyrosine residues) as phosphate acceptors. Thus, protein kinases can be classified by the amino acid substrate preference: serine/threonine kinases, tyrosine kinases, and dual kinases (meaning that both serine/threonine and tyrosine residues can be phosphorylated). Almost all protein kinases have catalytic domains that belong to a single eukaryotic protein kinase (ePK) superfamily. The common evolutionary ancestry of the kinase domain (also known as the catalytic domain), which consists of 250 to 300 amino acid residues, gives rise to a highly conserved three-dimensional structure.

There are 518 kinases in the human genome divided into eight major groups, which in totality represent 1.7% of the

KEY CONCEPTS

Kinase Families

518 kinases in the genome
90 protein tyrosine kinases (*e.g.*, JAKs)
400 protein serine/threonine kinases:
- AGC kinase family (*e.g.*, protein kinase B/AKT)
- CAMK kinase family (*e.g.*, Calmodulin-dependent kinase)
- CMGC kinase family (*e.g.*, MAP kinases: ERK, JNK, p38)
- STE kinase family (*e.g.*, MAPK kinases, MAPKK kinases)
- TKL kinase family (*e.g.*, IRAK)
- Others: casein kinase family, GYC kinase family, IKK family

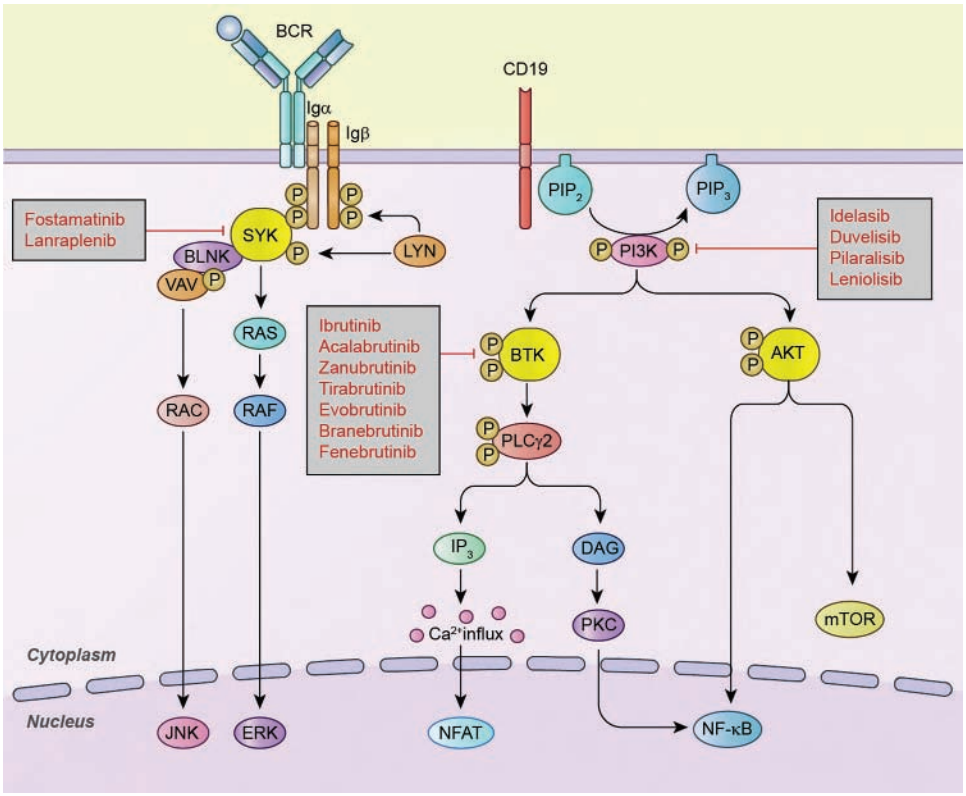


FIG. 85.1 Proximal Signaling Events in Response to B-Cell Receptor (*BCR*) Activation. Engagement of the BCR and related multichain immune recognition receptors (MIRRs) induces tyrosine phosphorylation of the receptor mediated by Src family kinases, such as Lyn. This allows recruitment of Spleen tyrosine kinase (Syk, or Zap-70 Zeta chain associated protein kinase-70) in T cells), which leads to activation of downstream serine kinases including Raf and other mitogen actived protein kinase (MAPK) family members. Signaling by MIRRs also activates the lipid kinase phosphatidylinositol 3′ kinase (PI3′K), which generates phosphatidylinositol 3,4,5, trisphosphate (PIP_3) and activates Bruton's tyrosine kinase and Akt.

Table 85.1 Kinase Inhibitors Approved for Immune-Mediated Disease

Kinase Class	Drug	Target	Indication
Abl	Imatinib	Abl	Hypereosinophilic syndrome, graft versus host disease
Janus kinase	Tofacitinib	JAK1, JAK2, JAK3	Rheumatoid Arthritis, Psoriatic Arthritis, Juvenile arthritis, Ankylosing spondyloarthritis, Ulcerative Colitis
	Ruxolitinib	JAK1, JAK2	Polycythemia vera and other Myeloproliferative neoplasms Graft versus host disease Atopic dermatitis (topical)
	Baricitinib	JAK1, JAK2	Rheumatoid Arthritis, Covid19
	Peficitinib	Pan JAK	Rheumatoid Arthritis (Japan, South Korea, Taiwan)
	Upadacitinib	JAK1 > JAK2	Rheumatoid Arthritis, Ankylosing spondyloarthritis
	Filgotinib	JAK1	Rheumatoid Arthritis (Europe, Japan)
	Abrocitinib	JAK1	Atopic Dermatitis
	Fedratinib	JAK2/FLT3	Myleloproliferative neoplasms
	Oclacitinib	Multiple JAKs	Atopic dermatitis (dogs)
	Delgocitinib	pan-JAK	Atopic dermatis - topical
Receptor Tyrosine kinase	Nintedanib	VEGFR, FGFR, PDGFR	Idiopathic pulmonary fibrosis, systemic sclerosis associated interstitial lung disease
Syk family	Fostamatinib	Syk	Idiopathic thrombocytopenic purpura
Tec family	Ibrutinib	Btk	Lymphoma
	Acalabrutinib	Btk	Lymphoma, Leukemia
	Zanubrutinib	Btk	Lymphoma, Leukemia
	Tirabrutinib	Btk	Lymphoma (Japan)

human genome. The protein tyrosine kinase (PTK) family has 90 members, one-third of which are receptor tyrosine kinases (RTK), and the remainder are cytoplasmic proteins that typically function in close proximity to, and downstream of, receptor/ligand complexes.

In terms of its catalytic role, the kinase domain has three functions: (1) the binding of the ATP (or GTP) phosphate donor as a complex with a divalent cation (usually Mg^{2+} or Mn^{2+}), (2) the binding of the protein substrate, and (3) the transfer of the γ-phosphate from ATP or GTP to the protein substrate. Despite the huge number of serine/threonine and tyrosine kinases

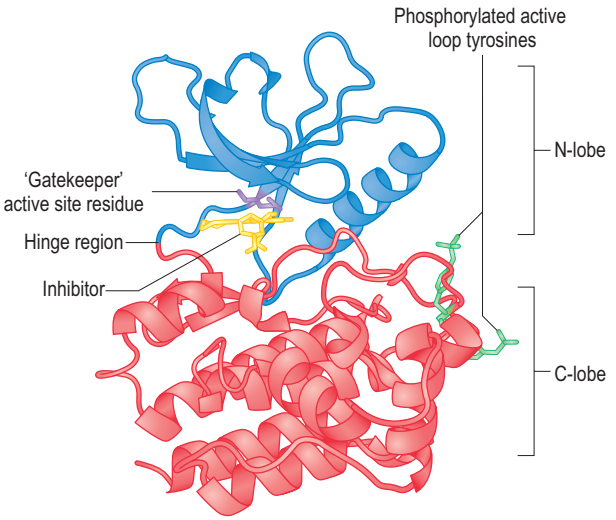


FIG. 85.2 Crystal Structure of the JAK3 Kinase Domain. This structure captures the active conformation of Janus kinase 3 (JAK3) with both active loop tyrosine residues phosphorylated *(green)*. The molecule can be described in two halves, with the N terminal lobe presented in blue and the C terminal domain in *red*. These are linked by a hinge region that forms part of the active site. Highlighted in *magenta* within the active site is the gatekeeper residue. Bound within this site is an analogue of the inhibitor staurosporine (*yellow*), and its proximity to the "gate keeper" residue highlights why this residue and this region are critical for the specificity of inhibitors for individual protein kinases.

there is evidence of a common ancestor and this is reflected in structural similarities particularly in the active (ATP bound) confirmation. The major kinase domains of all typical protein kinases consist of two lobes (N-lobe and C-lobe) that surround the nucleotide binding site (Fig. 85.2).[1] The smaller N-lobe consists of a cluster of β-pleated sheets with a single α-helix. The larger C-lobe is made up of α-helices. Within the C-lobe lies the substrate-binding site, typically a groove on the surface. A hinge region connects the two lobes. The hinge, together with two loops emerging from each lobe, form the ATP binding pocket and provide the primary target for most kinase inhibitors. In many protein kinases a loop emerging from the C-lobe, termed the activation loop, must be phosphorylated in order for the kinase to be fully active. Substrates of protein tyrosine kinases often include the activation loop of downstream kinases, creating signaling cascades of proteins that in turn phosphorylate each other; examples include the MAPKs (Fig. 85.3).

A SHORT HISTORY OF THE GENERATION OF KINASE INHIBITORS

Given that protein kinases bind ATP and that many inhibitors are competitive ATP antagonists, the notion that therapeutically useful kinase inhibitors could be generated was initially met with some skepticism. Given the hundreds of human kinases, many of which serve critical cellular functions, it was not clear *a priori* that it would be possible to attain the necessary specificity. Moreover, protein kinases are not the only kinases—there are lipid kinases and nucleotide kinases, and many other ATP-binding proteins that may share structural similarities with PTKs. For these reasons, designing a selective small molecule kinase inhibitor that sits in the ATP-binding pocket and does not target some other essential ATP-dependent process seemed daunting.[2] Fortunately, this dismal view does not reflect reality.

The first FDA-approved protein kinase inhibitor is imatinib. The mutated form of the Abelson leukemia (Abl) tyrosine kinase, BCR (breakpoint cluster region)-Abl, represents a fusion protein that is the result of a chromosomal translocation (Philadelphia chromosome) encountered in patients suffering from chronic myeloid leukemia (CML). The essential requirement for BCR-Abl kinase activity in CML made it an ideal target despite the aforementioned caveats with targeting protein kinases; in fact, imatinib has revolutionized the treatment of CML, arresting the progression of disease with minimal side effects in contrast with conventional cytotoxic chemotherapies.[3]

In addition to activity against Abl kinase, imatinib has activity against several other PTKs[3] and is effective in gastrointestinal stromal tumors and hypereosinophilic syndrome through effects on Kit and PDGFR platelet derived growth factor receptor)-FIPIL1 kinases, respectively. When considering immune-mediated disease, imatinib is used in the clinic for the treatment of fibrotic disease, including skin fibrosis associated with chronic graft *versus* host disease (GvHD) in patients with allogeneic bone marrow transplants (Chapter 92).

Receptor tyrosine kinases also emerged as targets in oncology and multiple monoclonal antibodies including bevacizumab (vascular endothelial growth factor receptor; VEGFR), ranibizumab (VEGFR), cetuximab (epidermal growth factor receptor; EGFR), pertuzumab (HER) and trastuzumab (HER2/neu) were generated, along with small molecule antagonists including erlotinib, gefitinib and others. Of note, nintedanib is a VEFGR, PDGFR, and FGFR (fibroblast growth factor receptor) signaling inhibitor approved for treatment of idiopathic pulmonary fibrosis as well as systemic sclerosis-associated interstitial lung disease.

At the time of writing, there are more than 70 FDA-approved small molecule kinase inhibitors, the majority of which are approved for oncologic indications, and more than 100 other inhibitors in clinical trial or development[4] (https://www.icoa.fr/pkidb/).

While conservation of the kinase ATP binding pocket has theoretically posed a problem for designing kinase inhibitors, in practice this has not happened for a number of reasons. While kinases may be structurally similar in an active ATP-bound confirmation, the inactive confirmation is more unique and can be used to improve selectivity.[5] In addition, the ATP binding region is made up of six polar amino acid residues that are invariant across whole families of kinases; similarly, there are a number of lipophilic residues that are highly conserved. This critical region contains an amino acid whose amide carbonyl binds to N-6 of adenine in the active confirmation. The side chain of this amino acid sticks into the reaction pocket in the inactive state and for this reason is referred to as "the gatekeeper residue." As the side chain is not involved in direct ATP binding, it varies across kinases, and variation of this gatekeeper residue is exploited by a number of inhibitors that are able to bind the inactive confirmation of specific kinases. In the case of Abl kinase, the gatekeeper residue is threonine, which binds directly to a methyl group of the phenyl ring of the Abl kinase inhibitor imatinib. Across the collective kinase superfamily, almost any amino acid can appear as the gatekeeper, although in practice it is typically a bulky nonpolar residue (methionine, tyrosine, phenylalanine, lysine). Cyclin dependent kinase 2 (CDK2) contains an additional pocket on its C-lobe next to the ATP binding pocket that can be exploited in the design of inhibitors.

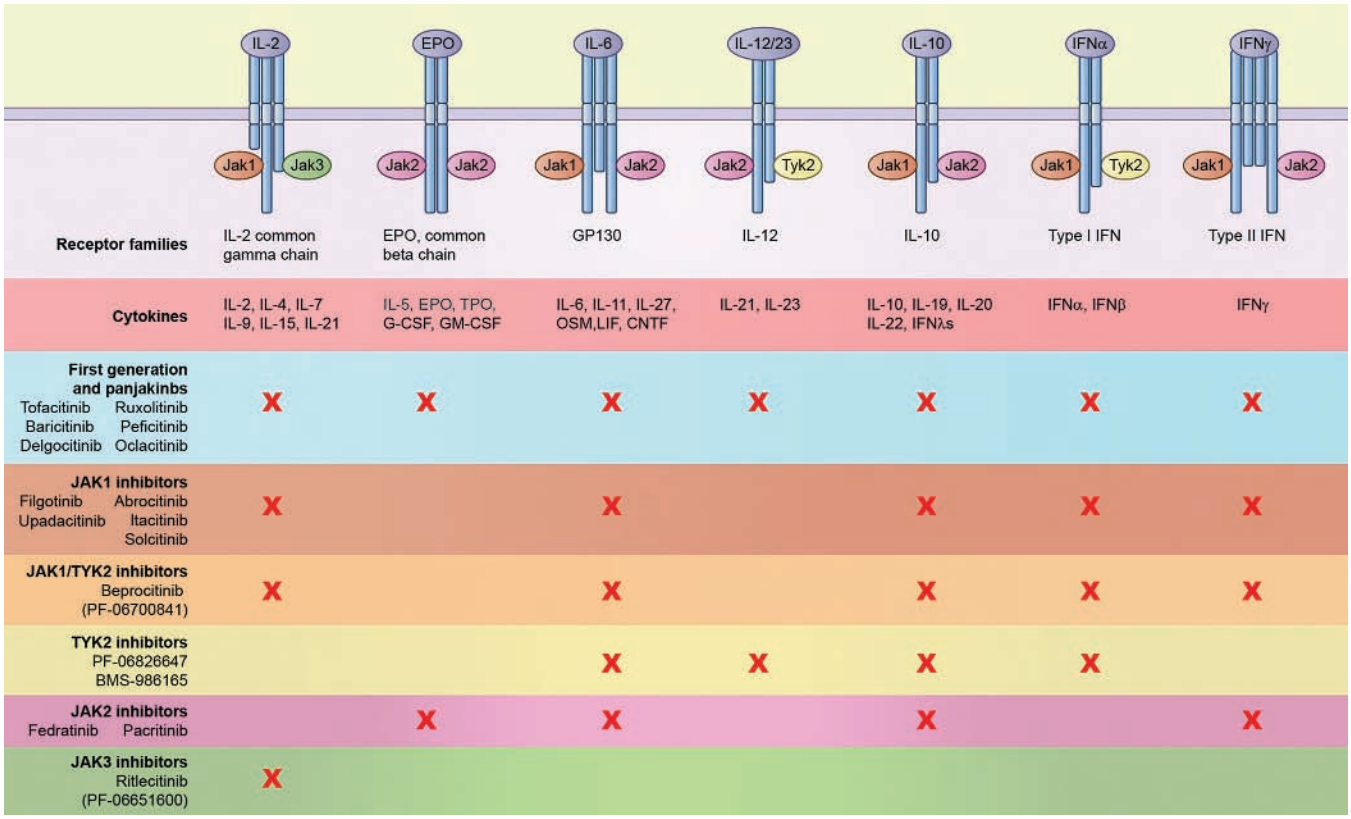


FIG. 85.3 Usage of Janus Kinases *(JAKs)* by Different Cytokine Receptors. The intracellular domains of Type I and Type II cytokine receptors are bound by JAKs, which mediate signal transduction. There are 4 JAKs that selectively bind to different receptor subunits. Some JAKs, like JAK1 and JAK2, bind to multiple classes of receptors. By contrast, JAK3 only binds to the common gamma chain. First-generation JAK inhibitors inhibit multiple JAKs, whereas newer inhibitors have greater selectivity and, therefore, inhibit a narrower range of cytokines.

Of further structural significance is the emergence of tumor drug resistance in response to the continued use of protein kinase inhibitors. Mutant forms of BCR-Abl, Kit and EGFR are associated with loss of drug activity and disease relapse, with a common site of mutation being the gatekeeper residue. This has led to a generation of "multikinase" inhibitors, including dasatinib and sunitinib, with partial inhibition of multiple kinases being less toxic than originally feared, and the lack of specificity potentially contributing to improved response rates in the treatment of CML.

TARGETING CYTOKINE SIGNALING

Cytokines regulate growth, survival, development, differentiation and activation of immune cells.

Their importance in driving inflammatory and immunological responses and their utility as targets for immune-mediated disease is clearly illustrated by the success of anti-cytokine monoclonal antibodies or "biologics." Cytokines can be divided into six large families on the basis of their receptors. We shall explore three of these families in greater detail.

JAKINIBS AND THE TYPE I/II CYTOKINE RECEPTOR FAMILY

Fifty-seven cytokines bind to receptors of the type I/II cytokine receptor family, receptors that lack intrinsic enzymatic activity and rely on recruiting intermediary Janus kinases for cell signaling (**see** Fig. 85.3). The essential in vivo function of JAKs was first established by mutations encountered in patients and shortly thereafter confirmed by knockout mice. Loss of *JAK3* results in a severe combined immunodeficiency, characterized by absence of T cells and natural killer cells and defective B cells. This phenocopies deficiency of the cognate receptor that associates with JAK3, the IL-2 receptor common γ chain, cγc (encoded by *IL2RG*), mutation. This mutation leads to X-linked severe combined immunodeficiency (X-SCID). Loss-of-function (LOF) mutations of *TYK2* (tyrosine kinase 2) are also associated with primary immunodeficiency characterized by bacterial, viral, and fungal infections.[6] *JAK1* LOF mutations cause primary immunodeficiency associated with atypical mycobacterial infection, whereas *JAK1* gain-of-function (GOF) mutations are associated with systemic immune dysregulation and hypereosinophilic syndrome. Somatic *JAK2* GOF mutations are associated with myeloproliferative neoplasms (*e.g.*, polycythemia vera). Mice lacking *Jak3* or *Tyk2* have severe but restricted immune phenotypes, whereas deficiency of *Jak2* or *Jak1* in mice is lethal, consistent with the broad range of critical cytokines that employ these respective kinases for signaling. These genetic phenotypes strongly supported the strategy for the development of a new class of immunomodulatory drugs. There are now multiple JAK inhibitors (jakinibs) approved for the treatment of immune and neoplastic disease and more in clinical trials (Tables 85.1 and 85.2).

TABLE 85.2 Selected Kinase Inhibitors in Clinical Trials

Drug	Target	Diseases
Itacitinib	JAK1	Chronic GVHD
Beprocitinib	Tyk2/JAK1	Atopic dermatitis, SLE
PF-06826647	Tyk2	Psoriasis, ulcerative colitis
Deucravacitinib	TYK2	Psoriasis, SLE
Momolotenib	JAK2	Myeloproliferative neoplasms
Gandotinib	JAK2	MPN, GvHD
Gusacitinib	JAK/Syk	Atopic dermatitis
Cerdulatinib	JAK/Syk	Vitiligo (topical)
Ritlecitinib	JAK3/Tec	Rheumatoid arthritis, alopecia areata, Crohn disease
PF-06650833	Irak4	RA, hidradenitis suppurative
GSK2982772	RIP1K	RA, Psoriasis, UC
Evobrutinib	Btk	Multiple sclerosis, rheumatoid arthritis
Branebrutinib	Btk	SLE, Rheumatoid arthritis, primary Sjogren syndrome
Fenebrutinib	Btk	Rheumatoid arthritis, urticarial
DNL747	RIPK1	Alzheimer disease, amyotrophic lateral sclerosis
Lanraplenib	Syk	SLE, lupus membranous nephropathy, Sjogren syndrome
Leniolisib	PI3K	Activated PI3Kδ syndrome
ATI-450	MK2	RA

THERAPEUTIC PRINCIPLES

- Drugs can inhibit protein kinases with a high degree of specificity
- Kinase inhibitors are useful in a wide range of immune-mediated diseases
- Multi-kinase inhibitors can be well tolerated and may be more efficacious than single kinase inhibitors (*e.g.*, second generation Abl kinase inhibitors), but high doses can be associated with more side effects.

Tofacitinib was the first jakinib to be used *in vivo* in mouse models of allogeneic organ transplantation and other models of immune-mediated disease.[7] It principally inhibits JAK1 and JAK3; to a lesser extent it inhibits JAK2 but has little effect on TYK2. Consequently, tofacitinib potently inhibits a broad range of cytokines, including those that contribute to the pathogenesis of autoimmune disorders, such as IL-6, cγc cytokines, interferons, IL-12, IL-23, and others. Functionally, tofacitinib inhibits the differentiation of T cells into cytokine-producing effector subsets but also blocks innate immune responses.[8]

Baricitinib is a JAK1/JAK2 inhibitor approved for the treatment of RA and has been successful in Phase 2 testing in systemic lupus erythematosus.[9] Of note, baricitinib also inhibits adapter associated kinase 1 (AAK1) and cyclin G-associated kinase (GAK), kinases involved in endocytosis that can impede viral entry. For this reason, baricitinib has received emergency use authorization for treatment of COVID-19 caused by the SARS-CoV2 virus, serving to inhibit both viral replication as well as the attendant cytokine storm.[10] Upadacitinib is a newer jakinib approved for treatment of rheumatoid arthritis (RA), which has a high degree of selectivity for JAK1 over JAK2. Abrocitinib has been approved for the treatment of atopic dermatitis and other jakinibs have been approved for this indication in Europe and the UK. Oclacitinib is a pan-JAK inhibitor approved for dogs with allergic dermatitis. Delgocitinib is a topically applied pan-JAK inhibitor approved in Japan for atopic dermatitis.[11,12]

The discovery that GOF mutations of *JAK2* underlie primary polycythemia and myelofibrosis provide a strong rationale for pharmacologically targeting JAK2. In fact, the JAK1/JAK2 inhibitor, ruxolitinib was the first FDA-approved JAK inhibitor for the treatment of myelofibrosis in 2011 and was subsequently approved for the treatment of primary polycythaemia in 2014. Ruxolitinib has also been approved for the treatment of acute graft-versus-host disease (GVHD)[13] and has completed Phase III trials in the treatment of chronic GVHD. Momelotinib is another JAK1/JAK2 inhibitor being evaluated for myeloproliferative neoplasms, and gandotinib is being studied for patients who have failed ruxolitinib. Fedratinib is a JAK/Flt3 (FMS-like tyrosine kinase 3) inhibitor approved for myeloproliferative neoplasms although this is principally due to its ability to inhibit the latter receptor tyrosine kinase. Jakinibs are also being tested in a variety of oncologic indications beyond myelofibrosis and related diseases.

CLINICAL PEARLS

Genetic mutations reveal key functions of kinases in patients with primary immunodeficiencies and provide clues to the impact of targeting signaling pathways

- Common-γ-chain deficiency, JAK3 deficiency: severe combined immunodeficiency
- ZAP70 deficiency: severe combined immunodeficiency
- TYK2 deficiency: rare cause of hyper IgE syndrome
- IRAK4 Autosomal recessive immunodeficiency with severe bacterial infections
- PI3KD (*PIK3CD*): autosomal dominated combined immunodeficiency with autoimmunity
- BTK: X-linked agammaglobulinemia
- ITK: autosomal recessive T cell immunodeficiency with EBV-associated lymphoproliferative disease

Among the side effects of first generation JAK inhibitors are infections, including both serious and opportunistic infection. The incidence of infection is similar to other drugs, including biologics, used to treat various autoimmune disorders, except for the increased cases of herpes zoster seen with jakinibs.[14,15] This increased risk for viral infection is likely related to the inhibition of JAK1-dependent interferon signaling. It may then come as a surprise that a number of JAK inhibitors have been used as therapies for COVID-19. A minority of patients infected with SARS-CoV2 develop an acute respiratory distress syndrome associated with the elevation of inflammatory cytokines, consistent with a virally induced macrophage activation syndrome. The IL-6 blocker tocilizumab was not effective for preventing intubation or death in patients with Covid-19.[16].

The combined use of jakinibs with other immunosuppressive drugs increases the incidence of infection. Anemia, thrombocytopenia, and leukopenia are all common complications, likely related to the fact that first generation jakinibs block JAK2 and interfere with cytokines, including erythropoietin, thrombopoietin, IL-11, oncostatin M, and colony-stimulating factors that drive bone marrow cell production. The use of tofacitinib and baricitinib in RA patients has been associated with an increased incidence of venous thromboembolism (VTE). The mechanism remains unclear as ruxoltinib is associated with a reduced rate of VTE when used in patients with primary polycythemia and myelofibrosis, diseases that are associated with an increased risk of VTE. Little reduction in $CD4^+$ T cells has been seen, but significant reductions in NK cells and $CD8^+$ T cells can occur. Despite this, patients have tolerated continued JAK inhibition for nearly a decade. Jakinibs are associated with increases in cholesterol, and compared to a tumor necrosis factor inhibitor are associated with increased cardiovascular events.[18]

In an effort to reduce some of the side effects, especially those related to JAK2 inhibition, a number of second generation jakinibs has been generated. Selective JAK1 inhibitors include filgotinib, abrocitinib and itacinib. Filgotinib has undergone multiple phase 2 and 3 studies in RA, ankylosing spondyloarthritis and ulcerative colitis and is approved in Europe. Abrocitinib has shown efficacy in clinical studies in psoriasis and is approved for atopic dermatitis.[19,20] Itacitinib failed a phase III trial when added to corticosteroid in the treatment of acute GVHD (GRAVITAS-301), although it is still being investigated in the treatment of chronic GVHD. It is possible that the ability to block JAK2 may offer some added benefit as GM-CSF, which signals through JAK2, may contribute to GVHD severity.[21,22]

Beprocitinib is a Tyk2/JAK1 inhibitor (PF-06700841) being studied in atopic dermatitis, and SLE (PF-06826647) is a TYK2 inhibitor being studied presently in Phase 2 in psoriasis and ulcerative colitis. Deucravacitinib is also a TYK2 inhibitor; however, it differs from other jakinibs in that it targets the JAK kinase-like domain rather than the ATP binding site in the catalytic domain. It has shown efficacy in psoriasis[23,24] and is being studied in SLE.

Ritlecitinib (PF-06651600) is a JAK3 inhibitor that also inhibits Tec family kinases (see below). Ritlecitinib has shown efficacy in Phase 2 trials in RA[25,26] and is also being studied in alopecia areata and Crohn disease.

IL-1 Family Receptors and Toll-Like Receptors

IL-1 is the prototypical member of an 11-member family of cytokines. The rest of the family has been renamed several times over the years and currently consists of IL-18, IL-33, four IL-36 members, IL-37, IL-38, and the IL-1 receptor antagonist (IL-1RA). The bulk of the cytokines in this family are associated with inflammation with the exception of IL-1RA, IL-36RA, and IL-37, the first two of which are natural inhibitors of IL-1 and IL-36, respectively. Deficiency of IL-1RA or IL-36RA cause the auto-inflammatory diseases DIRA and DITRA, respectively. IL-1RA is available as the biologic pharmaceutical, anakinra. It has modest benefit in RA, but has potent activity in juvenile idiopathic arthritis, Bechet disease, crystal arthopathies including gout, auto-inflammatory syndromes and, more recently, in macrophage activation syndrome.

Receptors of the IL-1 family and Toll-like family of pattern recognition receptors are found on macrophages and associate with the adapter molecule, myeloid differentiation primary response 88 (MyD88), which drives signaling that culminates in the activation of the signalosome (Fig. 85.4) A key link is the family of interleukin-1 receptor activated kinases (IRAKs), of which IRAK-1 and IRAK-4 play key roles that lead to cytokine release in activated macrophages together with the closely related TGFβ activated kinase (TAK1 a.k.a. MAP3K7). Deficiency of *IRAK4* in humans is associated with susceptibility to pyogenic infections. Selective IRAK4 inhibitors have been generated and are currently in clinical trials. A phase 2 trial using PF-06650833 in RA has shown efficacy, with infections being a common side effect (Phase 2 RA).[26] It is also being studied in hidradenitis suprativa.

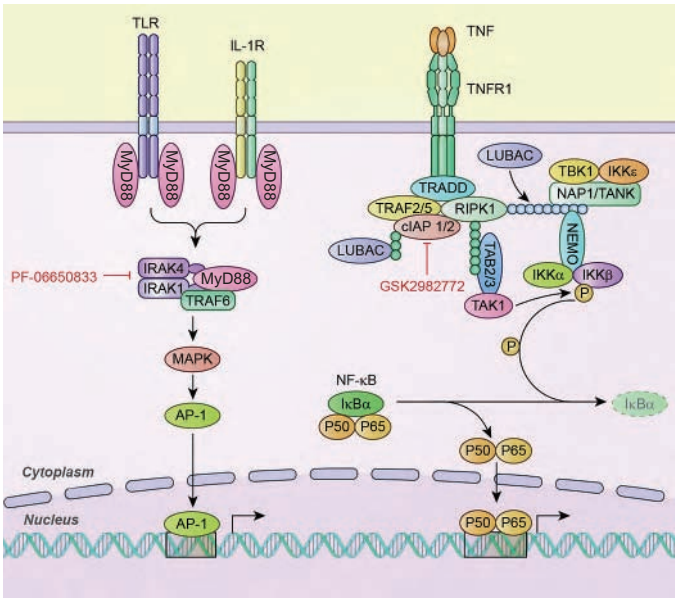


FIG. 85.4 Interleukin *(IL)*-1 and Tumor Necrosis Factor *(TNF)* Signaling. IL-1 family cytokines associate with the adapter molecule, myeloid differentiation primary response 88 (*MyD88*) that couples the receptor to activation of interleukin-1 receptor activated kinases (*IRAKs*), IRAK-1 and IRAK-4. TGFβ _activated kinase (*TAK1* a.k.a. *MAP3K7*). Receptor-interacting protein kinase 1 (*RIPK1*), a target of TAK1, that promotes proinflammatory cytokine production, apoptosis and necroptosis in response to TNF receptor engagement via the formation of a multi protein complex containing: TANK binding kinase 1 (TBK1), Tumor necrosis factor receptor type 1-associated DEATH domain protein (TRADD), cellular inhibitor of apoptosis proteins 1 and 2 (cIAP1/2), NF-kB activating kinase associated protein 1 (NAP1), TRAF family member-associated NFKB activator (TANK), linear ubiquitin chain assembly complex (LUBAC), and I kappaB kinase (IKK). Downstream of TAK1, which is also associated to TAK1)-binding protein 2/3 (TAB2/3 and RIPK1 lies the p38 mitogen activated protein kinase (MAPK) signaling pathway which leads to translocation of activator protein-1 (AP-1) in the nucleus.

TARGETING KINASES INVOLVED IN TNFR SIGNALING

TNF is an inflammatory hormone ubiquitously expressed in both lymphoid and myeloid cells. Its central role in disease pathophysiology was identified with the success of TNF antibodies in the treatment of RA and inflammatory bowel disease.[27,28] TAK1 lies downstream of both the IL-1 and TNF receptor families and thus contributes to inflammatory responses (see Fig. 85.4). A selective TAK1 inhibitor (takinib) has been generated and is efficacious in preclinical arthritis models.[29,30]

Receptor-interacting protein kinase 1 (RIPK1), a target of TAK1, is another regulator of proinflammatory cytokine production, apoptosis and necroptosis in response to TNF receptor engagement. GSK2982772 is a RIPK1 inhibitor studied in psoriasis. Subjects showed improvement with no severe drug-related AEs reported. The drug is also being studied in ulcerative colitis, which often responds to TNFα blockade. DNL747 is another RIPK1 inhibitor and is being studied in Alzheimer disease and amyotrophic lateral sclerosis. Downstream of TAK1 and RIPK1 lies the p38 MAPK signaling pathway that will be discussed later in this chapter.

TARGETING ANTIGEN- AND RELATED RECEPTOR SIGNALING

Antigen receptors play a key role in the initial activation of cells of the adaptive immune system; these include the T-cell receptor (TCR) and B-cell receptor (BCR) in lymphocytes, and the receptors for the Fc portion of immunoglobulin molecules (FcR) in macrophages and other cells. In all these cases the receptor complex lacks intrinsic receptor kinase activity and relies instead on the recruitment of members of the Src family of tyrosine kinases. The importance of Src family kinases in antigen receptor signaling (**see** Fig. 85.1) made members of this family attractive therapeutic candidates. Deficiency of Lck in mice leads to a progressive lymphopenia, and several Lck inhibitors have been developed; however, suitable agents for use as an immunosuppressive agents, remain elusive. For multichain immune recognition receptors, activation of Src kinases leads to the recruitment of a second set of tyrosine kinases that associate with receptor complexes, namely Zap70 or spleen tyrosine kinase (Syk). Deficiency of Zap70 causes severe combined immunodeficiency and preferential loss of $CD8^+$ T cells, but a successful Zap70 inhibitor has yet to be obtained. In contrast, Syk inhibitors have been generated; fostamatinib is approved for the treatment of immune thrombocytopenia purpura (ITP). Lanraspenib is in clinical trials for multiple indications including SLE, lupus membranous nephropathy, and Sjogren syndrome. As indicated above, combined JAK/Syk inhibitors have been generated, and one agent, cerdulatinib, is being tested in lymphoid malignancies. A topical formulation of cerdulatinib is being tested in vitiligo. Gusacitinib is a Syk/JAK inhibitor being studied in psoriasis and atopic dermatitis.

Downstream kinase activation of (phospholipase C) γ1 leads to production of intracellular calcium, which in turn activates the phosphatase calcineurin. Calcineurin dephosphorylates and activates nuclear factor of activated T cells (NFAT), which translocate to the nucleus, and, in cooperation with AP-1 transcription factors, activates expression of IL-2 and other key lymphocyte genes. Drugs that inhibit calcineurin, including cyclosporine and tacrolimus, have revolutionized organ transplantation, but despite their success, these drugs are associated with renal toxicity that can limit their long-term use. Voclosporin is a calcineurin inhibitor approved for lupus nephritis. In addition to the release of intracellular calcium, PLCγ1 activity leads to the generation of free diacylglycerol (DAG), which is critical for activation of members of the protein kinase C (PKC) family that in turn activate the transcription factor complex NF-κB. DAG also leads to activation of the Ras guanine nucleotide exchange factor (RasGEF). PKCs play important roles in oncology, but despite many attempts at generating useful inhibitors, no candidates have emerged to be useful agents for the treatment of immune-mediated disorders.

Lipid Kinases and Downstream Signaling

Kinases can phosphorylate lipids in addition to proteins, and these modifications are relevant in signal transduction mediated by both antigen and cytokine receptors. In addition to the production of inositol triphosphate and DAG by the action of PLCγ1, there is a second pathway of inositol lipid metabolism regulated by diverse receptors, including costimulatory molecules (*e.g.*, CD28), cytokines, and chemokines. This response is mediated by the Class I group of phosphatidylinositol 3 kinases (PI3Ks), which is composed of four isoforms (PI3Kα, β, γ, and δ), which phosphorylate the 3′-OH position of the inositol ring of phosphatidylinositol[4,5] bisphosphate ($PI(4,5)P_2$) to produce $PI(3,4,5)P_3$ (**see** Fig. 85.1). This lipid and its metabolite $PI(3,4)P_2$ bind to the pleckstrin homology (PH) domains of proteins and either induce localization of the protein to defined areas of the plasma membrane where activation can occur or induce conformational changes that allow for allosteric modifications of activity. Targets for D3 phosphoinositides in T cells include a number of downstream protein serine/thronine kinases, the Tec family of tyrosine kinases and the Rac-1 and RhoA guanine nucleotide exchange proteins.

The small molecule PI3K inhibitors, Wortmannin and LY294002, are both potent inhibitors of lymphocyte activation, although toxicity prevents either from being clinically useful. In contrast to the more widely expressed PI3Kα and β isoforms, PI3Kγ and δ are primarily expressed in hematopoietic cells. Deletion of PI3Kγ in mice results in defective migration of neutrophils and macrophages to sites of inflammation. The limited expression makes PI3Kγ a potentially useful target, and selective PI3Kγ inhibitors have shown to be effective in mouse models of collagen-induced arthritis. However, the recent description of combined immunodeficiency in patients with bi-allelic inactivating mutations affecting PI3Kγ suggests that such inhibitors may have more wide-ranging effects.

PI3Ks comprise two subunits: a catalytic subunit (p110α, β, γ, and δ) and a regulatory subunit (p85) (**see** Fig. 85.1). Both gain- and loss-of-function mutations affecting *PIK3CD* (encoding p110δ) and *PIK3R1* (encoding p85α) cause a combined immunodeficiency syndrome, referred to as activated PI3Kδ syndrome (APDS; or p110δ-activating mutation causing senescent T cells, lymphadenopathy and immunodeficiency [PASLI]).[31] Idelalisib, duvelisib, and copanlisib are PI3Kδ inhibitors approved for treatment of lymphomas and chronic lymphocytic leukemia. Leniolisib is a PI3Kδ inhibitor that is currently being tried for treatment of APDS.[32]

The Tec family kinases, which include Tec, Bmx bone marrow tyrosine kinase on chromosome X), Rlk (resting lymphocyte kinase), Itk (inducible tyrosine kinase) and Btk (Bruton's tyrosine kinase, contain PH domains (with the exception of Rlk) and are recruited to the plasma membrane by the presence of $PI(3,4,5)P_3$, where they are part of antigen receptor signaling complexes and are required for full PLCγ activity (**see** Fig. 85.1).[33] However, Btk is also involved in other signaling pathways, including those downstream from TLRs. The importance of this class of kinases is exemplified by the fact that mutations affecting *BTK* underlie Bruton agammaglobulinemia, a condition characterized by the absence of all B cells. Ibrutinib was the first FDA approved BTK inhibitor approved for the treatment of lymphoma, leukemias, and GvHD. Ibrutinib also has activity against the Tec kinase Itk, which is expressed in T cells and is important for activation of PLCγ1 and T-cell activation. Mutations affecting *ITK* also lead to a primary immunodeficiency. Zanubrutinib, tirabrutinib, and acalabrutinib (which is more specific for Btk), have also been approved. Fenebrutinib has shown efficacy in RA.[33] Other Btk inhibitors have been generated, evobrutinib and branebrutinib, and being studied for the treatment of malignancies and autoimmune diseases. Preliminary studies suggest that Btk inhibitors may be useful in the treatment of cytokine storm associated with COVID-19 due to their effects on innate cell cytokine production, and are being tested for this indication.[34]

$PI(3,4,5)P_3$-regulated kinases activated both by the TCR and IL-2 family cytokines include protein kinase B (PKB/AKT) and mechanistic target of rapamycin (mTOR). Both have been seen as targets for novel anti-cancer and anti-inflammatory agents

with mixed success. PKB regulates the activity of many proteins that are critical for cell survival and metabolism including the mTOR complex 1 (mTORC1), which regulates protein synthesis in response to cellular nutrient and energy availability. Many signaling pathways link growth factor receptors with activation of mTOR including the AMP dependent kinase (AMPK) and PI3K. mTORC1 promotes cell growth by activation of p70 S6K1 and inactivation of 4E-BP1,[35] which are critical for translation of new proteins.

As its name suggests, mTOR is inhibited by the macrolide rapamycin, now licensed for the treatment of graft rejection as the drug sirolimus. Sirolimus does not inhibit mTOR by direct binding to the ATP binding pocket but acts indirectly, associating with FK506 binding protein 12 (FKBP12). This in turn inhibits the kinase complex made up of mTOR, mLST8 and raptor (mTORC1). Sirolimus has been successfully used as an immunosuppressant, typically as part of a combination regimen for allograft rejection prophylaxis. Temsirolimus and everolimus are approved for treatment of malignancies, organ rejection, and tuberous sclerosis complex. In view of the ubiquitous expression of mTOR and its role in protein translation, it is not surprising that side effects include myelosuppression, hyperlipidemia, hypertriglyceridemia, myelosuppression, and delayed wound healing. There is some evidence of renal toxicity, but this is minor compared to the calcineurin inhibitors cyclosporine A and tacrolimus.

MAP Kinase Pathways

Another key set of kinases activated downstream of cytokine, antigen, and other adaptive and innate immune receptors is the Mitogen Activated Protein Kinase (MAPK) family. These kinases constitute a complex phospho-relay system of signal transduction, composed of sequentially activated kinases that are themselves modulated by phosphorylation. Three main MAPK cascades have been identified in mammalian cells—the ERK (Extracellular signal Regulated Kinase) cascade, the JNK (c-Jun N-terminal Kinase) cascade, and the p38 MAPK cascade. All start with a membrane-localized activator followed by three MAPKs (MAPKKK, MAPKK, and MAPK) that sequentially phosphorylate each other (see Fig. 85.3). Substrates of MAPK pathways include transcription factors, phospholipases, cytoskeletal proteins, and other protein kinases.

ERK1 and 2 were identified as kinases that were activated in response to growth factor stimulation, which are mimicked by expression of constitutively active Ras, an upstream activator of the ERK pathway that is frequently mutated in cancers. Fast-track designation was granted for the farnesyl inhibitor, tipifarnib, for *HRAS*-mutant head and neck squamous cell carcinomas. Ras is linked to ERKs by MEK1 and RAF1, a MAPK kinase (M3K). Sorafenib inhibits numerous kinases, including RAF, as well as receptor tyrosine kinases including PDGFR, VEGFR, Kit and FLT-3; however, its role as an immunosuppressant has yet to be explored. There are three RAF isoforms: A-RAF; B-RAF and C-RAF. A number of cancers are associated with B-RAF mutations resulting in a constitutively active kinase; the best known of these is the V600E mutant. These observations have led to a number of successful FDA-approved inhibitors of MAPK pathways. Binimetinib and cobmetinib are MEK1/2 inhibitor approved for treatment of melanoma.

Another limb of the MAPK pathways involves the c-Jun N-terminal kinase (JNK pathway). Many inflammatory agents including LPS, TNF and IL-1 are able to activate the JNK pathway. Many small molecule inhibitors of Jun kinases have been identified, but few have advanced to clinical trials, in part due to problems with selectivity. Brimapitide is being studied for the treatment of inflammatory eye disease and acute hearing loss.

The p38 MAPK cascade was originally identified as part of a drug screen looking for inhibitors of TNFα-mediated inflammatory responses.[36,37] TLR-dependent production of IL-1 and TNFα is p38 MAPK-dependent. The success of TNFα blocking antibodies in the treatment of RA has led to much interest in the development of p38 MAPK inhibitors. However, as with JNK, many p38 inhibitors have been generated, but their development into therapeutic drugs has been frustrated by either unacceptable toxicity or poor efficacy. A topically activating inhaled p38 MAPK inhibitor JNJ 49095397 (RV568) has had some success in phase II trials of chronic obstructive pulmonary disease. ASK1 (Apoptosis signal-regulating kinase 1) is a kinase that lies upstream of both p38 and JNK and has been investigated as an alternative target: multiple ASK1 inhibitors have been generated.[38] Selonsertib was investigated for the treatment of liver fibrosis, and while found to inhibit p38 *in vivo*, unfortunately was not efficacious.[39] MAPK-activated kinase 2 (MK2) is activated downstream by p38 MAPK. ATI-450 is a selective inhibitor of MK2 that inhibits proinflammatory cytokine production, spares other downstream kinases,[40] and is being studied in RA.

CONCLUSIONS

ON THE HORIZON

- New kinase inhibitors with improved selectivity have been approved and are in development.
- Topical and inhaled kinase inhibitors will hopefully be efficacious and safer.
- Further understanding of the pathophysiology of autoimmune disease will lead to the increased use and creation of kinase inhibitors.
- Increasing use of targeted therapy—specific kinase inhibitors—may improve our understanding of the detailed pathophysiology of autoimmune disease.

The scientific advances in the 1990s led to the elucidation of multiple intracellular signaling pathways that link membrane-bound receptor and cytokine signaling with changes in gene expression and cellular activation necessary to trigger an immune cell response. Many of these pathways are interconnected, leading to a complex array of networks made up of enzymes, adaptor proteins, and transcription factors; many of then became targets for drug discovery in the quest for therapies that effectively treat allergic and autoimmune diseases with an acceptable degree of immunosuppression. The success of the anti-cancer BCR-ABL inhibitor, imatinib and the immunosuppressive mTOR inhibitor, sirolimus placed the protein kinases center stage as targets of future drug discovery. More than two decades since the identification of many of these new targets, the numerous drugs designed to interfere with specific immune cell-signaling pathways have now been brought to the clinic. Development and testing of many new agents for an array of indications is well underway. Precisely how best to use these drugs separately and in combination, in different phases of immune pathology, and what biomarkers we need to use to optimize effectiveness remain important challenges.

REFERENCES

1. Hanks SK, Quinn AM, Hunter T. The protein kinase family: conserved features and deduced phylogeny of the catalytic domains. *Science*. 1988;241(4861):42–52.
2. Noble ME, Endicott JA, Johnson LN. Protein kinase inhibitors: insights into drug design from structure. *Science*. 2004;303(5665):1800–1805.
3. Druker BJ, Lydon NB. Lessons learned from the development of an Abl tyrosine kinase inhibitor for chronic myelogenous leukemia. *J Clin Invest*. 2000;105(1):3–7.
4. Roskoski R Jr. Properties of FDA-approved small molecule protein kinase inhibitors: a 2020 update. *Pharmacol Res*. 2020;152:104609.
5. Schindler T, Bornmann W, Pellicena P, et al. Structural mechanism for STI-571 inhibition of Abelson tyrosine kinase. *Science*. 2000;289(5486):1938–1942.
6. Lyons JJ, Milner JD. The clinical and mechanistic intersection of primary atopic disorders and inborn errors of growth and metabolism. *Immunol Rev*. 2019;287(1):135–144.
7. Changelian PS, Flanagan ME, Ball DJ, et al. Prevention of organ allograft rejection by a specific Janus kinase 3 inhibitor. *Science*. 2003;302(5646):875–878.
8. Ghoreschi K, Jesson MI, Li X, et al. Modulation of innate and adaptive immune responses by tofacitinib (CP-690,550). *J Immunol*. 2011;186(7):4234–4243.
9. Wallace DJ, Furie RA, Tanaka Y, et al. Baricitinib for systemic lupus erythematosus: a double-blind, randomised, placebo-controlled, phase 2 trial. *Lancet*. 2018;392(10143):222–231.
10. Stebbing J, Phelan A, Griffin I, et al. COVID-19: combining antiviral and anti-inflammatory treatments. *Lancet Infect Dis*. 2020;20(4):400–402.
11. Nakagawa H, Nemoto O, Igarashi A, et al. Delgocitinib ointment, a topical Janus kinase inhibitor, in adult patients with moderate to severe atopic dermatitis: a phase 3, randomized, double-blind, vehicle-controlled study and an open-label, long-term extension study. *J Am Acad Dermatol*. 2020;82(4):823–831.
12. Worm M, Bauer A, Elsner P, et al. Efficacy and safety of topical delgocitinib in patients with chronic hand eczema: data from a randomized, double-blind, vehicle-controlled phase IIa study. *Br J Dermatol*. 2020;182(5):1103–1110.
13. Zeiser R, von Bubnoff N, Butler J, et al. Ruxolitinib for Glucocorticoid-Refractory Acute Graft-versus-Host Disease. *N Engl J Med*. 2020;382(19):1800–1810.
14. Winthrop KL. The emerging safety profile of JAK inhibitors in rheumatic disease. *Nat Rev Rheumatol*. 2017;13(4):234–243.
15. Bechman K, Subesinghe S, Norton S, et al. A systematic review and meta-analysis of infection risk with small molecule JAK inhibitors in rheumatoid arthritis. *Rheumatology (Oxford)*. 2019;58(10):1755–1766.
16. Stone JH, Frigaula MJ, Serling-Boyd NJ, et al. Efficacy of tocilizumab in patients hospitalized with Covid-19. *N Eng J Med*. 2020;383:2333–2344.
17. Cao Y, Wei J, Zou L, et al. Ruxolitinib in treatment of severe coronavirus disease 2019 (COVID-19): a multicenter, single-blind, randomized controlled trial. *J Allergy Clin Immunol*. 2020;146(1):137–146. e3.
18. Gladman DD, Charles-Schoeman C, McInnes IB, et al. Changes in lipid levels and incidence of cardiovascular events following tofacitinib treatment in patients with psoriatic arthritis: a pooled analysis across phase III and long-term extension studies. *Arthritis Care Res (Hoboken)*. 2019;71(10):1387–1395.
19. Gooderham MJ, Forman SB, Bissonnette R, et al. Efficacy and safety of oral Janus kinase 1 inhibitor abrocitinib for patients with atopic dermatitis: a phase 2 randomized clinical trial. *JAMA Dermatol*. 2019;155(12):1371–1379.
20. Schmieder GJ, Draelos ZD, Pariser DM, et al. Efficacy and safety of the Janus kinase 1 inhibitor PF-04965842 in patients with moderate-to-severe psoriasis: phase II, randomized, double-blind, placebo-controlled study. *Br J Dermatol*. 2018;179(1):54–62.
21. Piper C, Zhou V, Komorowski R, et al. Pathogenic Bhlhe40+ GM-CSF+ CD4+ T cells promote indirect alloantigen presentation in the GI tract during GVHD. *Blood*. 2020;135(8):568–581.
22. Forman SB, Pariser DM, Poulin Y, et al. TYK2/JAK1 inhibitor PF-06700841 in patients with plaque psoriasis: phase IIa, randomized, double-blind, placebo-controlled trial. *J Invest Dermatol*. 2020;140(12):2359–2370. e5.
23. Page KM, Suarez-Farinas M, Suprun M, et al. Molecular and cellular responses to the TYK2/JAK1 inhibitor PF-06700841 reveal reduction of skin inflammation in plaque psoriasis. *J Invest Dermatol*. 2020;140(8):1546–1555. e4.
24. Burke JR, Cheng L, Gillooly KM, et al. Autoimmune pathways in mice and humans are blocked by pharmacological stabilization of the TYK2 pseudokinase domain. *Sci Transl Med*. 2019;11(502).
25. Papp K, Gordon K, Thaci D, et al. Phase 2 trial of selective tyrosine kinase 2 inhibition in psoriasis. *N Engl J Med*. 2018;379(14):1313–1321.
26. Robinson MF, Damjanov N, Stamenkovic B, et al. Efficacy and safety of PF-06651600 (Ritlecitinib), a novel JAK3/TEC inhibitor in patients with moderate to severe rheumatoid arthritis and an inadequate response to methotrexate. *Arthritis Rheumatol*. 2020;72(10):1621–1631.
27. Danto SI, Shojaee N, Singh RSP, et al. Safety, tolerability, pharmacokinetics, and pharmacodynamics of PF-06650833, a selective interleukin-1 receptor-associated kinase 4 (IRAK4) inhibitor, in single and multiple ascending dose randomized phase 1 studies in healthy subjects. *Arthritis Res Ther*. 2019;21(1):269.
28. Monaco C, Nanchahal J, Taylor P, et al. Anti-TNF therapy: past, present and future. *Int Immunol*. 2015;27(1):55–62.
29. Scarneo SA, Eibschutz LS, Bendele PJ, et al. Pharmacological inhibition of TAK1, with the selective inhibitor takinib, alleviates clinical manifestation of arthritis in CIA mice. *Arthritis Res Ther*. 2019;21(1):292.
30. Scarneo SA, Hughes PF, Yang KW, et al. A highly selective inhibitor of interleukin-1 receptor-associated kinases 1/4 (IRAK-1/4) delineates the distinct signaling roles of IRAK-1/4 and the TAK1 kinase. *J Biol Chem*. 2020;295(6):1565–1574.
31. Lucas CL, Chandra A, Nejentsev S, et al. PI3Kdelta and primary immunodeficiencies. *Nat Rev Immunol*. 2016;16(11):702–714.
32. Rao VK, Webster S, Dalm V, et al. Effective "activated PI3Kdelta syndrome"-targeted therapy with the PI3Kdelta inhibitor leniolisib. *Blood*. 2017;130(21):2307–2316.
33. Blank CU, Haining WN, Held W, et al. Defining 'T cell exhaustion'. *Nat Rev Immunol*. 2019;19(11):665–674.
34. Cohen S, Tuckwell K, Katsumoto TR, et al. Fenebrutinib versus placebo or adalimumab in rheumatoid arthritis: a randomized, double-blind, phase II trial (ANDES study). *Arthritis Rheumatol*. 2020;72(9):1435–1446.
35. Roschewski M, Lionakis MS, Sharman JP, et al. Inhibition of Bruton tyrosine kinase in patients with severe COVID-19. *Sci Immunol*. 2020;5(48).
36. McManus EJ, Alessi DR. TSC1-TSC2: a complex tale of PKB-mediated S6K regulation. *Nat Cell Biol*. 2002;4(9):E214–E216.
37. Lee JC, Laydon JT, McDonnell PC, et al. A protein kinase involved in the regulation of inflammatory cytokine biosynthesis. *Nature*. 1994;372(6508):739–746.
38. Ogier JM, Nayagam BA, Lockhart PJ. ASK1 inhibition: a therapeutic strategy with multi-system benefits. *J Mol Med (Berl)*. 2020;98(3):335–348.
39. Harrison SA, Wong VW, Okanoue T, et al. Selonsertib for patients with bridging fibrosis or compensated cirrhosis due to NASH: results from randomized phase III STELLAR trials. *J Hepatol*. 2020;73(1):26–39.
40. Wang C, Hockerman S, Jacobsen EJ, et al. Selective inhibition of the p38alpha MAPK-MK2 axis inhibits inflammatory cues including inflammasome priming signals. *J Exp Med*. 2018;215(5):1315–1325.

Biologic Response Modifiers

John M. Bridges and Walter Winn Chatham

In recent decades, the elucidation of specific inflammatory mediators as well as lymphocyte lineages that contribute to the pathology underlying chronic inflammatory disorders has identified immunologic targets amenable to modification by use of recombinant DNA technologies. The resulting biologic modifiers of inflammatory diseases have revolutionized the treatment of inflammatory, autoimmune, and allergic disorders. These targeted therapies, while costly, offer patients affected with these disorders options for better disease control with less of the morbidity associated with exposure to broader immune suppression and/or corticosteroid use.

IMMUNOMODULATORY CYTOKINES

IL-2 (rIL-2, Aldesleukin)

First identified as a T-cell growth factor, recombinant interleukin-2 (rIL-2) has potent immunomodulatory and antitumor activity, promoting the proliferation, differentiation, and recruitment of T and B cells, and natural killer (NK) cells as well as proinflammatory T-cell cytokines, such as IL-1, tumor necrosis factor-alpha (TNF-α), and interferon-gamma (IFN-γ) (Table 86.1). These all likely play a major role in dose-related IL-2 toxicity that may include hypotension, cardiac arrhythmias, increased capillary permeability with pulmonary edema, fever, and rarely death. Therapeutic use of rIL-2 was traditionally in patients with advanced melanoma and renal cell carcinoma, but the use of aldesleukin in cancer patients has been increasingly supplanted by more targeted immune-enhancing "check-point" therapies targeting T-cell death receptors (PD-1) and cytotoxic T-lymphocyte antigen-4 (CTLA4) that have comparable antitumor effects with much less acute toxicity.

More recently, the administration of low-dose rIL-2 has been shown to promote preferential growth and proliferation of regulatory T cells (Treg) without significantly affecting release of inflammatory mediators by $CD4^+$ T cells. Such low-dose regimens of rIL-2 have been used with success in suppressing manifestations of graft-versus-host disease (GvHD) and may also favorably impact the course of type 1 diabetes. The use of rIL-2 and/or other cytokine therapies promoting enhanced Treg differentiation and proliferation may, therefore, provide novel interventions for managing a variety of inflammatory disorders associated with autoimmunity, including systemic lupus erythematosus (SLE).[1]

Interferons

Interferons (IFNs), released primarily by T lymphocytes as well as dendritic cells in the context of stimulation of Toll-like receptors (TLRs), have a wide array of immunomodulatory effects that include upregulation of genes governing angiogenesis, cell differentiation, expression of human leukocyte antigens (HLAs), and production of inflammatory cytokines. IFN-α and IFN-β bind to the same cell surface receptor (IFN-1R) and are designated type 1 IFNs, while IFN-γ binds to a different receptor (IFN-2R) and is designated as a type 2 IFN. Recombinant preparations of all three interferons have been used for management of inflammatory disorders associated with chronic viral infection, primary immunodeficiency, or select autoimmune disorders (most notably multiple sclerosis [MS]). While there are immunomodulatory attributes that render interferons useful in certain circumstances, constitutional symptoms and upregulation of a wide variety of genes that promote inflammation frequently limit the use of both type 1 and type 2 interferons.

Interferon-α

Recombinant IFN-α2b (rIFN-α2b) has most commonly been used in combination with ribavirin for treatment of hepatitis C, including hepatitis C virus (HCV)-associated cryoglobulin syndromes. Other disorders linked to viral infections, including lymphomatoid granulomatosis caused by Epstein-Barr virus (EBV), and polyarteritis nodosa caused by hepatitis B virus (HBV), have been successfully treated with regimens incorporating IFN-α2b as part of the treatment regimen.[2] Refractory retinal vasculitis in the context of Behçet disease, severe flares of familial Mediterranean fever not responding to treatment with colchicine, and eosinophilic granulomatosis with polyangiitis (EGPA) have also been reported to respond favorably to treatment with rIFN-α2b.[2] Common side effects of rIFN-α2b include flu-like syndrome, fatigue, anorexia and nausea, weight loss, hair loss, emotional lability and depression, cytopenias, and induction of autoantibodies with enhancement of autoimmune disease.

Interferon-β

Recombinant interferon-β1a (rIFN-β1a) or rIFN-β1b has been shown to decrease relapse rates, disease severity, and central nervous system magnetic resonance imaging (MRI) lesions in patients with relapsing multiple sclerosis (MS). However, there are conflicting data with regard to the efficacy of IFN-β1a or IFN-β1b preparations for patients with secondary progressive variants of MS.[3] Immunomodulatory effects attributed to the beneficial impact of IFN-β1b on MS include enhancement of suppressor T-cell activity, reduction of proinflammatory cytokines, downregulation of antigen presentation, and reduced trafficking of lymphocytes into the central nervous system.[3] Side effects observed during treatment with both rIFN-β

TABLE 86.1 Recombinant Immunomodulatory Cytokines

Molecule	Construct	Half-Life	Dosing
Aldesleukin	rIL-2	1.5 h	0.3–3×10^6 IU/m^2 sc qd
IFN-α2b	rIFN-α2b	1.7 h	3–10×10^6 IU 3×/wk
IFN-α2b pegylate	rIFN-α2b pegylate	17 h	50–150 mcg/wk
IFN-β1a	rIFN-β1a	69 h	8.8–44 mcg sc 3×/wk
IFN-β1b	rIFN-β1b	0.2–4.3 h	2–8×10^6 IU q.o.d.
IFN-γ	rIFN-γ	6 h	50 mcg (1×10^6 IU)/m^2 sc 3×/wk

preparations include injection-site reactions as well as most of the effects observed during treatment with rIFN-α2b. Reported autoimmune complications include features often associated with lupus, such as immune complex glomerulonephritis, cutaneous vasculitis, and panniculitis.[4]

Interferon-γ

Recombinant interferon-γ (rIFN-γ) is effective in decreasing the frequency and severity of infections in patients with chronic granulomatous disease (CGD), an inherited disorder associated with any number of genetic defects that impair assembly of the nicotinamide adenine dinucleotide phosphate (NADPH) oxidase and production of superoxide anion metabolites in neutrophils.[5] Fever, myalgias, skin rash, fatigue, and diarrhea are the most common adverse events associated with IFN-γ treatment. During intercurrent infections, treatment with rIFN-γ often enhances infection-related constitutional symptoms, potentially obscuring responses to antimicrobial therapy. With greater adherence to antimicrobial prophylactic regimens incorporating combination therapy with itraconazole and trimethoprim-sulfa, the added benefit of treatment with rIFN-γ in CGD patients may be marginal.

KEY CONCEPTS

- Interferon-alpha (IFN-α) has been used with success to enhance resolution of chronic viruses, notably hepatitis C virus (HCV); newer antivirals are supplanting the need for IFN therapy in HCV.
- IFN-β has immunomodulatory effects that are of benefit in the treatment of certain subsets of patients with multiple sclerosis.
- Type-1 IFNs (including IFN-α and IFN-β) may unmask or trigger flares of autoimmune disease, most notably systemic lupus erythematosus (SLE).
- Interferon-gamma may be of some benefit in the management of patients with chronic granulomatous disease (CGD) not responding to anti-fungal and and-bacterial prophylactic regimens.

INHIBITORS OF INFLAMMATORY CYTOKINES

Tumor Necrosis Factor-α Inhibitors

TNF-α is a significant mediator of inflammation associated with psoriasis, rheumatoid arthritis (RA), spondyloarthopathies, and chronic inflammatory bowel diseases (IBDs: Crohn disease and ulcerative colitis [UC]). TNF-α promotes the ingress of immunocompetent cells at sites of inflammation through activation of the vascular endothelium and upregulation of adhesion molecules. TNF-α also stimulates synthesis of other proinflammatory cytokines (IL-1β, IL-6, granulocyte–macrophage colony-stimulating factor [GM-CSF]), chemokines (IL-8), and inflammatory eicosanoids (prostaglandin E_1 [PGE_1], leukotriene B_4 [LTB_4]). In RA, TNF-α stimulates macrophages and monocyte-derived osteoclasts to release mediators destructive to bone and cartilage, including matrix metalloproteinases such as collagenase and stromelysin.[6] These noted effects of TNF-α have, therefore, prompted the development of both soluble receptor and monoclonal reagent strategies to inhibit effects of TNF-α in disorders in which it appears to have a significant role in promoting inflammation and tissue injury (Table 86.2).

Etanercept is a recombinant soluble p75 TNF receptor (TNFR–CD120b)-IgG Fc fusion protein that binds soluble TNF-α as well as lymphotoxin-β. The avidity of the p75 TNFR for soluble TNF-α is comparable to that of the p55 TNFR and therefore effectively decreases net TNF signaling through either of these receptors. The clinical impact of binding lymphotoxin-β is not entirely known. Recent work highlighting the role of lymphotoxin-β in maintaining the physiologic regulatory function of macrophages around lymphoid germinal centers suggests that inhibition of this cytokine may enhance the delivery of apoptotic body-derived self-antigens to germinal centers—a consideration of potential clinical relevance in patients at risk for SLE or related autoimmune syndromes (see below).

Infliximab, adalimumab, and *golimumab* are monoclonal antibody reagents with comparable avidity for both soluble as well as cell-bound TNF-α, but differ with regard to the derivation of the antigen-binding sequence and half-life. *Certolizumab pegol* is a construct consisting of recombinant human sequence-derived F(ab)' anti-TNF-α covalently linked to 40 kDa polyethylene glycol. The pegolated construct enhances the half-life of the reagent, and the absence of Fc and complement-binding domains renders certolizumab less likely to engender local injection site reactions and also precludes placental crossing of certolizumab into the fetal circulation when administered to pregnant women.

In cytokine profile studies of joint tissues and synovial fluids derived from patients with RA, TNF-α is consistently among the highest expressed inflammatory cytokines.[6] All five TNF-α inhibitors have been shown to decrease the signs and symptoms of disease activity as well as inhibit the progression of structural damage in rheumatoid arthritis, with greatest benefits and probability of disease remission seen among patients with early RA.[7]

TNF-α levels are also notably increased in biopsy samples of skin and synovial tissues from patients with psoriasis and psoriatic arthritis, and all of the TNF-α inhibitors have been shown to be highly effective in suppressing manifestations of plaque psoriasis as well as articular disease in psoriatic arthritis.[8] Improvements in joint symptoms (axial and peripheral) and uveitis manifestations are observed in patients with other spondyloarthropathies, including ankylosing spondylitis during treatment with TNF-α inhibitors. However, use of these agents has not been shown to delay fusion of axial joints in patients with spondylitis. The known inhibitory effect of TNF-α on Smad pathway-mediated bone formation by osteoblasts suggest that TNF inhibition may potentially promote ossification at sights of inflammation and increased bone turnover.

Monoclonal antibody inhibitors of TNF-α as well as certolizumab (but not etanercept) have been shown to be beneficial in the management of intestinal inflammation in patients with Crohn disease or UC and are approved for treatment of these disorders.[9]

Although very well tolerated as a class, TNF-α inhibitors may impair innate host defenses, resulting in delay of resolution of

TABLE 86.2 Recombinant Inhibitors of Proinflammatory Cytokines

Molecule	Construct	Half-Life	Dosing (maintenance)
Etanercept	sTNFR:Fc	3–4 days	50 mg sc q wk or 25 mg sc 2×/wk
Infliximab	aTNFα (chimeric IgG1κ)	7–12 days	3–10 mg/kg iv q4–8 wk
Adalimumab	aTNFα (human IgG1κ)	10–20 days	10–40 mg sc q1–2 wk
Certolizumab	aTNFα (humanized Fab')-PEG	14 days	200–400 mg sc q2–4 wks
Golimumab	aTNFα (human IgG1κ)	14 days	50–100 mg sc q4 wk 2 mg/kg iv q8 wk
Anakinra	sIL-1ra	4–6 h	1–8 mg/kg sc daily (max 200 mg/d)
Rilonacept	IL-1R1:Fc(IgG1):IL-1RAcP	8–9 days	2.2 mg/kg (max 160 mg) sc q wk
Canakinumab	aIL-1β (human IgG1κ)	26 days	2–8 mg/kg sc (max 600 mg) sc q8 wk
Tocilizumab	aIL-6R (humanized IgG1κ)	13 days	4–12 mg/kg (max 800 mg) iv q2–4 wk 160 mg sc q1–2 wk
Satralizumab	aIL-6R (humanized IgG2)	22–37 days	120 mg sc every 4 weeks
Sarilumab	aIL-6R (human IgG1)	8–10 days	150–200 mg sc every 2 weeks
Siltuximab	aIL-6 (chimeric IgG1k)	14–30 days	11 mg/kg iv every 3 weeks
Ustekinumab	aIL-12/23 p40 (human IgG1κ)	15–32 days	0.75 mg/kg (ped) or 45–90 mg sc q12 wk
Gesulkumab	aIL-23 p19 (humanized IgG1κ)	15–18 days	100 mg sc every 8 weeks
Tildrakizumab	aIL-23 p19 (humanized IgG1κ)	20–28 days	100 mg sc every 12 weeks
Risankizumab	aIL-23 p40 (humanized IgG1κ)	28–29 days	150 mg sc every 12 weeks
Secukinumab	aIL-17A (human IgG1κ)	22–31 days	150–300 mg sc q4 wk
Ixekizumab	aIL-17A (humanized IgG4κ)	13 days	80 mg sc q2–4 wk
Brodalumab	aIL-17A (human IgG2)	11 days	210 mg sc every 2 weeks

intercurrent infections. TNF-α inhibitors are associated with reactivation of tuberculosis and fungal infections, including histoplasmosis, coccidioidomycosis, and viral infections, such as hepatitis B. In all circumstances of intercurrent infection, the standard recommendation is for treatment with TNF-α inhibitors to be withheld until the infection has resolved. No complications have been reported regarding their use in patients with intercurrent hepatitis C infection or in patients with HIV infection that is well controlled with highly active antiretroviral therapy.

In the context of its role in promoting cell apoptosis, TNF-α constitutes a component of the host defense against tumor cell survival and growth. As such, the incidence and prevalence of cancer in populations treated with TNF-α inhibitors have been emphasized in pharmacovigilance programs. Lymphoma has been reported in patients taking TNF-α inhibitors, with most reported cases occurring in the context of treatment for RA. However, the relative risk for lymphoma in patients with RA in the pre-biologic era approximates 3.0 and the prevalence of lymphoma in RA patients exposed to TNF-α inhibitors has not been shown to exceed the expected prevalence among patients with RA. Among patients with RA, spondyloarthropathies—or psoriasis—the risk of cancer has not been shown to be increased in the context of treatment with TNF-α inhibitors.[10] While it is often recommended that TNF-α inhibitor therapy be discontinued in the setting of a newly diagnosed cancer—and that initiation of such therapy be avoided in patients with known malignancy—there are no clinical data to confirm continued treatment with TNF-α inhibitors impairs the therapeutic response to cancer therapies.

The role of TNF-α inhibition in facilitating apoptosis is also of relevance with regard to the development or potentiation of autoimmunity. Although rare, the development of psoriaform skin eruptions, demyelinating syndromes, and drug-induced lupus erythematosus are well documented occurrences in patients treated with TNF-α inhibitors.[11] Infliximab, adalimumab, and golimumab may induce release of apoptotic products in the context of antibody-dependent cellular cytotoxicity (ADCC) of cells bearing surface TNF-α, potentially driving production of antibodies to nucleoproteins.[12] Binding of lymphotoxin-β by etanercept may alter the clearance of apoptotic products by germinal center adjacent follicular dendritic cells and macrophages.[13] Given these collective considerations, it is generally recommended that TNF-α inhibitors be discontinued in patients who develop autoimmune complications during treatment with this class of biologic therapy. Moreover, anti-TNF-α therapy is not recommended for patients with established SLE or related overlap syndromes associated with autoantibody production.

KEY CONCEPTS

- TNF-α inhibitor treatment significantly improves patient symptoms and function and inhibits structural damage in patients with RA.
- TNF-α inhibitors are very effective in suppressing skin lesions due to psoriasis as well as enthesal inflammation in patients with seronegative spondyloarthopathies; they do not appear to have any inhibitory effects on the development of bony ankyloses.
- Vigilance for underlying fungal disease or mycobacterial disease is prudent prior to commencing with and during treatment courses with TNF-α inhibitors.
- TNF-α inhibitors may exacerbate certain autoimmune syndromes and should be used with caution or avoided in patients with autoantibody-associated immune disorders such as systemic lupus erythematosus (SLE).

IL-1β Inhibitors

IL-1β is synthesized as an inactive precursor, with activation of IL-1β occurring following engagement of nucleotide-binding oligomerization domain (NOD)-like receptors by a variety of exogenous or endogenous danger signals, which then trigger formation of molecular platforms (NALP-inflammasomes) that facilitate cleavage of the IL-1β precursor by IL-1 converting enzyme (ICE). IL-1β stimulates proliferation of lymphocytes, upregulates the expression of adhesion molecules, and triggers the release of a variety of inflammatory mediators from leukocytes, including chemotactic factors, prostaglandins, proteases, and procoagulants. In patients with rheumatoid arthritis, IL-1β triggers the release of proteases from phagocytic cells and macrophages that are destructive to bone and cartilage. Inflammasone activation has also been shown to mediate acute flares of arthritis in patients with gout or calcium pyrophosphate

deposition. Dysregulation of NLRP3 due to gain-of-function mutations in inflammasome-related genes has been implicated in the pathogenesis of familial cryopyrin-associated periodic syndromes (CAPS; cryopyrinopathies), familial Mediterranean fever (FMF), and the pyogenic arthritis, pyoderma gangrenosum, and acne syndrome (PAPA). Increased expression and levels of IL-1β are also noted during flares of inflammation in patients with adult-onset Still disease (AoSD), systemic-onset juvenile idiopathic arthritis (soJIA), and macrophage activation syndromes (MAS) associated with these and other autoimmune or infectious disorders. The naturally occurring IL-1 receptor antagonist (IL-1ra) prevents the binding of IL-1β and IL-1α to the IL-1 receptor (IL-1R), but tissue levels of IL-1ra in the above disorders may be insufficient to counteract the effects of IL-1β.

Anakinra is a recombinant non-glycosylated IL-1 receptor antagonist (rIL-1ra) that differs from the endogenous IL-1Ra by a single amino acid addition at the amino terminus. Administered subcutaneously at daily (or for some indications more frequent) intervals due to its very short serum half-life, anakinra functions as a competitive inhibitor of IL-1α and IL-1β binding to IL-1 receptors.

Anakinra is presently approved for the treatment of rheumatoid arthritis and inhibits joint erosions. However, compared to the clinical responses observed with use of TNF inhibitors and IL-6 inhibitors with regard to tender and swollen joints, such responses with anakinra are much more modest. Given these observations with biologics that require much less frequent dosing, *anakinra* is used infrequently in the treatment of RA. With the increased understanding of the role of IL-1 inflammasome activation in crystal-induced arthropathies—such as gout and systemic inflammatory disorders such as soJIA and AoSD—there has been increased interest in anakinra and other IL-1 inhibitors as therapeutic options in these disorders. Case reports reinforce the use of *anakinra* in managing acute flares of gout or flares of calcium pyrophosphate associated pseudogout when the use of corticosteroids or nonsteroidal antiinflammatory drugs (NSAIDs) is not favored due to comorbid conditions, such as uncompensated heart failure or diabetes with significant renal impairment.[14]

Prompt resolution of inflammatory markers and clinical manifestations have been reported in patients with severe flares of AoSD and soJIA treated with *anakinra*, including those associated with MAS.[15,16] In a study re-examining outcomes from a previous sepsis trial, survival was improved in patients with clinical features of MAS who were randomized to *anakinra* treatment.[17] The relatively short half-life of *anakinra* may render its use feasible when short-term blockade of IL-1 may be of benefit in managing severe flares of IL-1–driven inflammation when there is a concern for emerging intercurrent infection.

Rilonacept (also referred to as IL-1-Trap) is a recombinant fusion protein composed of the extracellular domain of the IL-1 accessory protein and IL-1 receptor type 1 attached to the Fc portion of IgG1. *Rilonacept* binds to IL-1a and IL-1b with high affinity and is approved for the treatment of cryopyrin-associated periodic syndromes (CAPS). Recent studies have confirmed efficacy of *rilanacept* in the management of recurrent idiopathic pericarditis.[18] Similar to anakinra, *rilonacept* is generally well tolerated with injection-site reactions being the most common adverse events.

Canakinumab is a human genome sequence-derived monoclonal antibody with specificity for IL-1β. Approved for the treatment of CAPS as well as soJIA, *canakinumab* has the longest half-life of the currently approved therapies that target IL-1. In a randomized clinical trial comparing the use of *canakinumab* to parenteral *triamcinolone* administration for acute gout flares, *canakinumab* was shown to have greater reductions in joint pain/swelling, decreased needs for rescue medications, and a greater time to subsequent flare.[14] However, given observed rates of serious infections were twice as high in the *canakinumab* arm, a half-life that well exceeds the duration of typical flares, and a cost that greatly exceeds the cost of current therapies effective for managing acute gout, the use of *canakinumab* has not found favor in treatment paradigms for managing gout.

IL-6 Inhibitors

IL-6 mediates activation of macrophages and osteoclasts, terminal proliferation and differentiation of B cells, differentiation of Th17 cells, and production of liver acute phase proteins.[19] However, IL-6 also governs homeostatic processes, including granulopoiesis, enteric epithelial proliferation, and antiinflammatory responses involved in resolution of inflammation, such as production of the soluble TNF receptor p55 and the IL-1 receptor antagonist (IL-1ra).[19]

The biology of IL-6 signaling is complex in that signaling may occur directly through the cell membrane–bound IL-6 receptor (IL-6R)/gp130 protein complex (classic signaling) or through binding of IL-6 to soluble IL-6R (sIL-6R) with the resulting heterodimer then ligating a variety of other gp130-containing membrane receptors (trans-signaling) that mediate signaling from cytokines other than IL-6. Whereas IL-6R is expressed primarily on leukocytes, hepatocytes, and megakaryocytes, gp130-containing receptor complexes are expressed in almost all organs, including heart, kidney, spleen, liver, lung, placenta, and brain. In murine models of inflammation employing sgp130:Fc constructs that selectively bind and neutralize IL-6/sIL-6R, trans-signaling appears to mediate many of the observed inflammatory consequences of the upregulation of IL-6, whereas classic signaling through membrane IL-6R primarily mediates homeostatic processes such as granulopoiesis, thrombopoiesis, and epithelial cell proliferation.[19] As greater understanding of the different IL-6 signaling pathways in humans evolves, implementation of therapeutic strategies targeting IL-6 will require consideration of the impact of blocking classic membrane-bound IL-6R signaling and IL-6/sIL-6R trans-signaling.

Tocilizumab is a humanized mAb with specificity for the human IL-6 receptor that blocks the binding of IL-6 to IL-6R, inhibiting both classic signaling by IL-6 through membrane IL-6R and formation of IL-6/sIL-6R heterodimer ligands that engender trans-signaling. Treatment with tocilizumab improves tender and swollen joint counts and slows the development of joint erosions in patients with RA. It is also effective as a steroid-sparing therapy in patients with giant cell arteritis.[20] Treatment with tocilizumab is also reported to be of benefit in patients with soJIA, AoSD, multicentric Castleman disease, and cytokine release syndrome in treated cancer patients.[20,21]

Satralizumab is a humanized monoclonal antibody targeting the IL-6 receptor that has been studied primarily in patients with neuromyelitis optica spectrum disorder. When added to standard immunosuppressive regimens, *satralizumab* was shown to decrease occurrence of relapse compared to the immunosuppressive regimen alone.[22]

Sarilumab is a monoclonal antibody also with specificity for IL6R, binding both membrane-bound IL6R and soluble IL6/sIL6R heterodimers. Safety and efficacy comparable to that of

tocilizumab and superior to the anti-TNF inhibitor *adalimumab* has been shown in studies using *sarilumab* for treatment of RA.[23]

Siltuximab is a monoclonal antibody with specificity for soluble IL-6 that has demonstrated efficacy and has been approved for use in the treatment of idiopathic (non-HIV [human immunodeficiency virus] and non-HHV8 [human herpesvirus 8]-associated) multicentric Castleman disease.[24]

Presumably related to the role of IL-6 in promoting granulopoiesis, neutropenia has been observed in some patients following treatment with antibodies targeting IL-6R, but this is uncommon. Thrombocytopenia and elevated serum aminotransferase levels have also been uncommonly observed. A predictable mild increase in serum lipid levels is observed following the initiation of treatment with anti-IL-6R therapies; the clinical significance of this remains indeterminate but monitoring of lipid levels is recommended. Reactivation of tuberculosis and invasive fungal infections can occur, and rare instances of gastric and intestinal rupture have been reported in patients treated with tocilizumab.

ON THE HORIZON

Other Anti-IL-6 and sgp130:Fc Reagents

Sirukumab is a humanized anti-IL-6 monoclonal construct that has also been shown in phase II trials to have efficacy in RA. *Clazakizumab* (humanized anti-IL6) has shown promising results for the treatment of RA as well as psoriatic arthritis; by virtue of engineered aglycosylation, *clazakizumab* has a longer half-life than other monoclonal reagents targeting IL-6 or IL-6R, thereby requiring less frequent dosing.

Of potentially greater interest have been studies with the development of *olokizumab*—a humanized monoclonal reagent that targets the gp130-binding domain of IL-6—and soluble gp130 covalently linked to immunoglobulin Fc (sgp130:Fc), both of which primarily block the interaction of the IL-6/sIL-6R complex with membrane gp130, thereby selectively inhibiting IL-6/sIL-6R trans-signaling.[25] Selective inhibition of the trans-signaling pathway linked to inflammatory responses may be superior to complete IL-6 blockade, because important physiologic functions of IL-6 (hematopoiesis, maintenance of intestinal integrity) mediated through membrane-bound IL-6R are left intact.

KEY CONCEPTS

- IL-β inhibitors are very effective in the treatment of disorders associated with constitutional or secondary activation of the IL-1 inflammasome, including cryopyrin syndromes, familial Mediterranean fever flares, recurrent pericarditis, and flares of crystal-induced arthropathy (gout and CPPD/pseudogout).
- Use of IL-1β inhibitors may be of benefit in managing severe flares of inflammation associated with macrophage activation, such as soJIA, AoSD, and severe sepsis syndromes.
- IL-6 signaling is complex in the context of homeostatic signaling through the IL-6R and proinflammatory trans-signaling through many other receptors engendered by IL-6/sIL-6R complexes.
- Monoclonal antibody or soluble receptor constructs that inhibit IL-6 binding to IL-6R or IL-6/sIL-6R complexes to membrane-bound gp130 are very effective in suppressing clinical manifestations of rheumatoid arthritis.

IL-23 and IL-17 Inhibitors

IL-23 is a heterodimeric molecule composed of a p19 subunit and a common p40 subunit shared with IL-12. IL-23 is secreted by B cells as well as innate immune cells including activated dendritic cells, monocytes, macrophages, and innate lymphoid cells. A key function of IL-23 is to promote the survival and expansion of Th17-lineage T cells, primarily by stabilizing RORγt transcription factor.[26] Although sharing a common subunit, IL-12 and IL-23 have different immunologic effects on T-cell-lineage development, with IL-12 primarily promoting development and maturation of Th1 T cells. Strategies targeting the shared p40 subunit of IL-12/IL-23 in murine models of autoimmune inflammation yielded promising results and led to the development and approval of a biologic reagent targeting p40 for treatment of psoriasis, psoriatic arthritis, and inflammatory bowel disease. However, comparative IL12/23p40, IL-12p35, and IL-23p19 knock-out studies in murine models of inflammation suggested the observed ameliorative effects of anti-p40 on inflammation are likely primarily due to IL-23 inhibition, and selectively targeting IL-23 as well as IL-17 (the primary effector cytokine of Th17 cells) evolved as strategies to target autoimmune inflammation.[26]

Ustekinumab is a monoclonal antibody with specificity for the common p40 subunit of IL-12 and IL-23 and is approved for the treatment of psoriasis, psoriatic arthritis, Crohn disease and ulcerative colitis. *Ustekinumab* significantly improves psoriatic skin lesions and decreases both peripheral and axial joint/enthesis manifestations in psoriatic arthritis.[27] Significant improvements in Crohn disease were observed among patients who failed to respond adequately to anti-TNF-α therapies.[28] Higher than expected rates of infection or malignancy have not been observed in controlled studies of *ustekinumab* in psoriatic disease, spondylitis, or inflammatory bowel disease. Nonetheless, vigilance for mycobacterial, fungal, and *Salmonella* infections is recommended given the role of IL-12 and IL-23 in host defense against these pathogens.

Guselkumab is a monoclonal antibody that selectively binds the p19 subunit component of IL-23, blocking its interaction with the IL-23 receptor and thereby limiting the release of IL-23 triggered proinflammatory cytokines. It is approved for the treatment of moderate to severe plaque psoriasis and was shown to be more effective at reducing skin involvement as measured by Psoriasis Area and Severity Index (PASI) than a*dalimumab* or *secukinumab*.[29] In reported clinical trials, treatment with *gesulkumab* was not associated with a significant increase in rates of infection or other serious adverse events compared to placebo. Several phase III trials for the use of *guselkumab* in psoriatic arthritis (NCT03158285, NCT03162796) and in lupus nephritis (NTC04376827) are ongoing.

Tildrakizumab is a humanized monoclonal antibody that targets the p19 component of IL-23 approved for the treatment of moderate-to-severe plaque psoriasis in adults. In two phase III clinical trials, *tildrakizumab* had no significant increased risk of infections or serious adverse events compared to placebo, although diarrhea was more likely in those treated with tildrakizumab compared to those treated with placebo.[29] *Tildrakizumab* is currently in phase II studies for psoriatic arthritis, ankylosing spondylitis, and non-radiographic axial spondyloarthritis (NCT04314544, NCT03552276).

Risankinzumab is a humanized monoclonal antibody targeting the p19 component of IL-23 approved for the treatment of moderate-to-severe plaque psoriasis in adults. Compared to placebo, *risankizumab* increased the risk of infection (mostly upper respiratory infections and tinea infections), though there was no appreciable difference in rates of serious infections. In comparison to *adalimumab, risankizumab* showed significantly

greater efficacy in improving skin involvement in patients with moderate-to-severe plaque psoriasis.[29] Phase III trials for its use in inflammatory bowel disease (NCT03398148, NCT 03104413, NCT03105102, NCT04524611, NCT03105128), hidradenitis suppurativa (NCT03926169), and psoriatic arthritis (NCT03675308) are ongoing.

ON THE HORIZON

The anti-IL23(p19) mAb *mirikizumab* has recently been studied in a completed phase III comparative trial with the IL-17 inhibitor secukinumab for treatment of plaque psoriasis. *Brazikumab*, a mAb that also targets IL23(p19), has shown promising benefit in phase II studies for treatment of inflammatory bowel disease.

IL-17 Inhibitors

IL-17 is elaborated primarily by the Th17 lineage of effector T cells, which differentiate from $Foxp3^{-}CD4^{+}$ thymocytes under the influence of IL-6 and TGFβ and proliferate/survive under the influence of IL-23. IL-17 is also produced by other innate immune cells, including neutrophils, mast cells, keratinocytes, macrophages, NK cells, NKT cells, and innate lymphoid (LLC/LTi) cells. Th17 cells are prevalent in the inflammatory lesions of a variety of inflammatory disorders including RA, inflammatory bowel disease, psoriasis, and spondyloarthropathies. IL-17 released by Th17 cells and other innate cells triggers the induction and release of IL-6, TNF-α, CCL2, CCL3, and MMPs from macrophages, induces activation of osteoclasts at sites of bone resorption, and promotes the proliferation, maturation, and chemotaxis of neutrophils.[30] As such, IL-17 is important both in the initial innate host defense to infection and also inflammation driven by acquired immune responses. Given the multitude of effects mediated by IL-17 in perpetuating inflammation in a number of autoimmune disorders, monoclonal reagent–based strategies have been developed targeting IL-17 and its receptor.

Secukinumab is a fully human mAb that selectively binds to and neutralizes IL-17A. In clinical trials, *secukinumab* has been shown to significantly decrease the activity of skin lesions in patients with psoriasis, decrease tender and swollen joints in patients with psoriatic arthritis, and decrease axial pain and limitation in patients with ankylosing spondylitis.[31,32] Rates of infection were increased over that observed in patients randomized to placebo, but not in excess of the infection rates noted with anti-TNF therapy. Clinical responses to *secukinumab* in patients with RA have been less robust. Despite the apparent role of Th17 cells in patients with inflammatory bowel disease, no significant improvement was observed in patients with Crohn disease, with some patients developing worsening of disease during treatment with *secukinumab*.[33] Given the pivotal role of Th17 in host responses to bacteria, parasites, and fungi, it is possible that alterations in the containment of microbial pathogens in the enteric mucosa may account for the failure of targeting IL-17 in inflammatory bowel disease. *Secukinumab* is currently being studies as an adjunctive treatment for lupus nephritis (NCT04181762).

Ixekizumab is a monoclonal antibody targeting IL-17A that has been shown to be effective and is approved for the treatment of plaque psoriasis.[31] Similar to what has been observed in clinical trials with *secukinumab*, rates of infection are slightly increased in patients treated with *ixekizumab* versus placebo, but the rate of serious infections has not been observed to be increased relative to placebo or a comparative arm randomized to anti-TNF treatment with *etanercept*.

Brodalumab is a monoclonal antibody targeting the IL-17A receptor. Data from two identically designed phase III randomized trials demonstrated clearing of skin lesions to be significantly greater than that observed in a comparator group receiving *ustekinumab*, but Candida infections and neutropenia were more frequent in the *brodalumab* groups than in the *ustekinumab* and placebo groups.[29,31]

ON THE HORIZON

Recently completed phase II trials have demonstrated efficacy of the dual IL17A/IL17F inhibitor *bimikizumab* in treating manifestations of psoriatic skin and joint disease as well as ankylosing spondylitis.[34]

KEY CONCEPTS

- Selective inhibition of pathways promoting differentiation and activation products of Th17 helper T cells is of significant benefit in the treatment of psoriasis and the seronegative spondyloarthropathies.
- Selective inhibition of IL-17 appears to be of marginal benefit in the treatment of inflammatory bowel disease, whereas selective inhibition of IL-23 by targeting p19 or both IL-12 and IL-23 by targeting their shared p40 subunit appears to be of benefit in both psoriasis and inflammatory bowel disease.

Inhibitors of Interferons

Emapalumab. The observed increases in IFNγ and related gene signatures prevalent in patients with cytokine storm syndromes (CSS) associated with familial variants of hemophagocytic lymphohistiocytosis (fHLH) and severe flares of soJIA or AoSD have prompted efforts to develop biologic reagents targeting this cytokine. In studies of children and adults with fHLH, the humanized anti-IFN-gamma monoclonal antibody emapalumab was shown to be efficacious in ameliorating clinical features of CSS and is now approved for use in this setting.[35] There is evolving experience in published case reports suggesting emapalumab may also be of benefit in CSS associated with severe flares of soJIA/AoSD.[36]

ON THE HORIZON

Targeting Type 1 Interferons

With recognition of the role of type-1 interferons in the immunopathogenesis of systemic lupus erythematosus (SLE), strategies to target α-IFNs for treatment of SLE have emerged.

- Initial trials with monoclonal reagents, such as *rontalizumab* or *sifalimumab* with specificity to one or more of the known α-interferon subtypes (14 known presently), have yielded modest results, emphasizing the complexity of the interferon system in SLE as autoantibodies to α-interferon subtypes are often prevalent in SLE sera.
- The alternative approach of targeting the type-1 IFN receptor in phase II and phase III trials with the monoclonal antibody **anifrolumab** has yielded more encouraging results.[37]
- Other approaches targeting the production of type-1 interferons currently under investigation include the use of monoclonal reagents to decrease activation of or deplete plasmacytoid dendritic cells (pDC), perceived to be the major source of IFN-1 in SLE.
- In clinical trials conducted thus far with antibodies targeting α-IFNs or IFN-1R, higher than expected occurrences of herpes zoster have occurred.
- Whether targeting type-1 interferons will prove efficacious, yet not at the expense of significant susceptibility to viral diseases such as varicella zoster virus and Covid-19, remains to be determined.

Complement and Kinin Pathway Inhibitors

In addition to opsonization of microbial pathogens and products of apoptosis, products of complement activation play a significant role in recruitment and activation of phagocytic cells (C3a, C5a) and inducing cell-membrane damage (C5b-9). Since both C5a and C5b-9 have potent procoagulant activity at sites of inflammation, complement activation is also hypothesized to be critical to the generation of intravascular thrombi in thrombotic microangiopathy syndromes including atypical hemolytic-uremic syndrome (aHUS), catastrophic anti-phospholipid syndrome (CAPS), paroxysmal nocturnal hemoglobinuria (PNH), and systemic lupus erythematosus.

Eculizumab is a humanized monoclonal antibody with specificity for C5 that blocks the generation of C5a and C5b by the C5 convertase (Table 86.3). *Eculizumab* is currently approved for use in patients with aHUS, severe myasthenia gravis, aquaporin-4 antibody-positive neuromyelitis optica spectrum disorder, and paroxysmal nocturnal hemoglobinuria. It has also been reported to be of benefit in patients with catastrophic antiphospholipid syndrome and SLE associated with thrombotic microngiopathy.[38,39] Because *eculizumab* inhibits terminal complement activation and life-threatening or fatal meningococcal infections have occurred in patients who received *eculizumab*, vaccination targeting both A and B serotypes of meningococcus is therefore recommended at least two weeks prior to receiving *eculizumab*. In cases wherein the risks of delaying *eculizumab* therapy outweigh the risk of developing a meningococcal infection, meningococcal vaccines should be administered as soon as possible with at least two weeks of prophylactic penicillin or macrolide antibiotics.

Ravulizumab is a humanized monoclonal antibody with a longer therapeutic half-life than *eculizumab* that also binds to C5, blocking activation of C5. It is approved in the United States and EU for use in the management of aHUS as well as the management of hemolysis in patients with paroxysmal hemoglobinuria. Clinical trials in patients with PNH have demonstrated efficacy noninferior to eculizumab and continued efficacy in patients transitioned from every two-week dosing with *eculizumab* to every 8-week dosing with *ravulizumab*.[40] Appropriate pre-treatment meningococcal vaccination or concurrent immunization with prophylactic antibiotics should be administered.

Excess generation of bradykinin, the vasodilator responsible for symptoms in hereditary angioedema (HAE), may occur in the context of a deficiency in the quantity or functional activity of C1 inhibitor (C1-INH), a regulatory inhibitor of kallikrein. Frequency of attacks of HAE may be decreased by infusions of pooled plasma derived C1-INH, but breakthrough attacks are not uncommon.

Lanadelumab is a fully human monoclonal antibody that binds to plasma kallikrein, decreasing kallikrein-mediated generation of bradykinin from high molecular weight kininogen (Table 86.3). Treatment with *lanadelumab* significantly reduces the rate of attacks in HAE patients and is approved for use in adults and children older than the age of 12 years with HAE.[41]

TABLE 86.3 Recombinant Inhibitors of Complement and Kinin Pathway Activation and Molecules Mediating Cell Migration

Molecule	Construct	Half-Life	Dosing (maintenance)
Eculizumab	Anti-C5 (humanized IgG2/4κ)	8–15 days	300–1200 mg iv q1–2 wk
Ravulizumab	Anti-C5 (humanized IgG2/4κ)	50–52 days	2100–3600 mg iv q8 wk
Lanadelumab	Anti-kallikrein (human IgG1κ)	14 days	150–300 mg sc q2–4 wk
Natalizumab	Anti-α4β1 (humanized IgG4κ)	7–15 days	300 mg iv q4 wk
Vedolizumab	Anti-α4β7 (humanized IgG1κ)	25 days	300 mg iv q8 wk

Adhesion Molecule Inhibitors

Trafficking of phagocytic cells and lymphocytes across vascular endothelium is critical to the development of inflammatory lesions and associated tissue injury. Strategies to decrease the influx of monocytes—and/or T and B lymphocytes—using biologics that target adhesion molecules have, therefore, been deployed as treatments for several inflammatory disorders, including multiple sclerosis (MS) and inflammatory bowel disease.

Natalizumab is a humanized monoclonal antibody with specificity for the α-4 subunit of the very-late activation antigen (VLA4)-integrin molecule expressed on lymphocytes and monocytes (Table 86.3). *Natalizumab* blocks association of the VLA4 α4β1 and α4β7 integrins with their respective vascular receptors, thereby limiting cell transmigration into tissue sites of inflammation in the CNS as well as intestinal mucosa. Administration of *natalizumab* blocks the interaction of T-cell α4β1 with its addressin ligand on venules in the CNS and has been shown to be effective in decreasing the frequency of relapses of multiple sclerosis (MS).[42] *Natalizumab* also inhibits the interaction of the α4β7 integrin with mucosal addressin-cell adhesion molecule-1 (MAdCAM-1) expressed on venules in the enteric mucosa and has been shown to induce remissions and prevent flares of Crohn disease in patients who have failed to have an adequate response to anti-TNF-α therapy.[42] *Natalizumab* primarily affects transmigration of Th1 cells with much less of an inhibitory effect on the transmigration of Th17 cells; as such, there may be differential efficacy of natalizumab in the treatment of Th1 versus Th17 driven subtypes of MS and/or inflammatory bowel disease.[43]

The major concern with regard to use of *natalizumab* is reactivation of JC polyoma virus in patients who are carriers, resulting in progressive multifocal leukoencephalopathy (PML). While quite debilitating and often fatal, the occurrence of PML is uncommon and the risk can be mitigated by limiting the duration of natalizumab treatment to one year in known JC virus carriers (as defined by seroconversion) and not administering natalizumab concurrent with other immunosuppressive therapy.

Vedolizumab is a humanized monoclonal antibody that specifically binds to $\alpha_4\beta_7$ integrin, blocking its interaction with mucosal addressin-cell adhesion molecule-1 (MAdCAM-1) on intestinal endothelial cells. Since *vedolizumab* does not bind to or block the interaction of the $\alpha_4\beta_1$ integrin to its addressin ligand in the CNS, the risk of developing PML during treatment is substantially lower than that associated with the use of *natalizumab*. In clinical trials, *vedolizmab* has been shown to significantly improve disease activity and reduce flares in patients with Crohn disease or ulcerative colitis.[44] This strategy may prove to be particularly useful in subsets of patients with inflammatory bowel disease who also have clinical features of SLE, in whom use of anti-TNF-α agents may carry an increased risk of SLE exacerbation.

KEY CONCEPTS

- Inhibition of C5 cleavage to its active products and resulting assembly of the membrane attack complex with eculizumab or ravulizumab is of significant benefit to patients who have complement mediated-hemolytic or thrombotic microangiopathy syndromes associated with paroxysmal nocturnal hemoglobinuria, atypical hemolytic-uremic syndrome, catastrophic anti-phospholipid syndrome, neuromyelitis optica spectrum disorder, or severe flares of SLE (systemic lupus erythematosus).
- Immunization against *Neisseria* species is recommended for patients who require dosing with complement pathway inhibitors.
- The loss of appropriate regulatory control of kallikrein by C1-INH that occurs in patients with hereditary angioedema can be effectively treated with lanadelumab, which inhibits the generation of bradykinin by kallikrein.
- Antibodies targeting the lymphocyte integrin $\alpha_4\beta_1$ (natalizumab) have been shown to be of benefit in managing multiple sclerosis and inflammatory bowel disease, but are associated with a risk of JC virus activation as PML (progressive multifocal leukoencephalopathy).
- Antibodies to the $\alpha_4\beta_7$ integrin (vedolizumab) selectively inhibit trafficking of lymphocytes into the intestinal lamina propria and are of benefit in the management of inflammatory bowel disease.

Inhibitors of B-Cell Activation

Identified targets for regulation of B-cell activation include growth and survival factors such as BLyS(BAFF) and its respective receptors (BAFF-R, TACI, and BCMA), co-stimulatory receptors and their ligands such as CD40/CD40-ligand, and cell surface receptors such as CD22 or FcγRIIb that engender inhibitory signaling when ligated. Given the prominent role of multiple autoantibodies and generalized B-cell activation in SLE, strategies to target B-cell activation have been undertaken in patients with this disorder.

Belimumab is a recombinant human genome–derived monoclonal antibody with specificity for soluble (non-membrane bound) BLyS (B-lymphocyte stimulator, BAFF) (Table 86.4). Through ligation of the BAFF-R and TACI receptors on B lymphocytes, BLyS promotes the maturation of B cells into antibody-secreting plasmablasts. Following encouraging results in murine models validating the hypothesis that autoreactive B lymphocytes may have greater dependency upon BLyS to survive and proliferate, human studies with *belimumab* demonstrated improvement in SLE disease activity using validated disease-activity measures (SLEDAI and BILAG) as well as additional secondary endpoints related to disease flares and sparing of corticosteroid use.[45] Noted clinical improvements typically do not become manifest until after 6 months of treatment with belimumab, with the majority of clinical improvements observed in musculoskeletal, mucocutaneous, and serologic domains of disease activity.[45] Subsequent studies have shown addition of *belimumab* to standard lupus nephritis induction regimens, and subsequent maintenance therapy increases the likelihood of achieving and maintaining renal remission at 2 years.[46] The potential effects on more severe neurologic domains of disease activity have not yet been assessed in controlled trials.

Significant decreases in the measured numbers of circulating activated B cells and plasmacytoid B lymphocytes are observed following 6 months treatment with *belimumab*.[47] Although receptors for BLyS (BAFF-R) and B-cell maturation antigen (BCMA) have also been identified on murine T-follicular helper (Tfh) cells—with ligation of BAFF-R promoting activation of Tfh and ligation of BCMA appearing to play a role in downregulation of Tfh responses—the significance of these findings in human SLE and what impact treatment with *belimumab* has on Tfh responses remain to be determined. In human trials there were no observed decreases in $CD4^+$ and $CD8^+$ T lymphocytes following 1 year of treatment with *belimumab*.[47] Levels of autoantibodies, including anti-dsDNA, anti-Smith, anti-SSA, and anti-cardiolipin, are decreased 40% to 50% after the first year of treatment and continue to decrease over time with long-term treatment.[47] In contrast, total antibody levels decrease on average only 15% with no significant decreases in measured pre-existing antibody titers to influenza, tetanus toxoid, or pneumococcal serotypes. Furthermore, treatment with *belimumab* does not appear to have any significant effects on the primary immune responses to pneumococcal bacterial antigens.[45] In the two major randomized phase III trials of 52 and 76 months duration, the frequency of infections was not shown to be increased in the *belimumab* treatment arms.[45]

TABLE 86.4 Recombinant Inhibitors of Lymphocyte Proliferation, Survival, and Activation

Molecule	Construct	Half-Life	Dosing (maintenance)
Belimumab	aBLyS/BAFF (human IgG1λ)	19 days	10 mg/kg iv q4 wk
Abatacept	CTLA4:IgG1Fc	13 days	10 mg/kg; 500–1000 mg sc or iv q4 wk
Belatacept	CTLA4:IgG1Fc	10 days	10 mg/kg iv q4 wk
Basiliximab	aCD25/IL-2ra (chimeric IgG1κ)	7–9 days	20 mg iv (repeat ×1 at 4 days)
Dupilumab	aIL-4Rα (human IgG4)	Unknown[a]	300 mg sc q2 weeks

[a]See text.

ON THE HORIZON

Other Inhibitors of B-Cell Activation

Atacicept is a recombinant human fusion protein containing the extracellular, ligand-binding portion of the receptor TACI (transmembrane activator and CAML interactor) and a modified Fc portion of human IgG. *Atacicept* binds both BLyS (BAFF) and a proliferation-inducing ligand (APRIL), thereby functioning as an antagonist to the ability of these two ligands to stimulate B lymphocytes. However, the therapeutic window with this approach may be relatively narrow as membrane-bound TACI is the major receptor mediating immunoglobulin class switch during B-cell maturation, and excess of the soluble receptor administered over time may result in significant humoral immune deficiency due to decreases in IgG-secreting B cells and plasma cells. Human SLE studies using TACI-Ig as a therapeutic intervention remain underway to determine its potential efficacy and safety in lupus and IgA nephropathy (NCT02808429).

Dapirolizumab is a pegylated anti-CD40-ligand F(ab)′ monoclonal construct that interferes with the interaction of T-cell CD40L with B-cell CD40, thereby blocking co-stimulatory signals required for cognate T-cell help that promotes antigen-specific B-cell proliferative responses. Following favorable results targeting CD40L in murine lupus models, human trials with monoclonal anti-CD40L were undertaken but halted in the context of observed thrombotic complications. Subsequent studies demonstrated upregulation of CD40L on platelets, with platelet aggregation occurring in the

context of complement fixation to platelet membrane-bound anti-CD40L. *Dapirolizumab* appears to be free of the effect on platelet activation/aggregation and is being evaluated in human SLE trials (NCT04294667).

Rozibafusp alfa is a bispecific peptide–antibody conjugate that targets BAFF and ICOSL, thereby inhibiting both BAFF-induced survival and proliferation of B cells, as well as T-cell co-stimulation mediated by T-cell surface-bound iCOS and B-cell/dendritic cell expressed ICOSL. Early-phase clinical trials are underway to determine efficacy of the construct in SLE (NCT04058028).

Inhibitors of T-Cell Activation

Given the role of T-cell lymphocytes in orchestrating adaptive immune responses, selective inhibition of T-cell activation is an attractive target for modulating inflammatory disorders associated with immune responses to autoantigens or allografts. Blocking cell receptors for T-cell growth factors such as IL-2 has been employed to prevent allograft rejection and may be potentially of use in the management of autoimmune disorders. Since productive immune responses are not generated in the absence of effective co-stimulatory signals, blocking T-cell co-stimulation has also been an attractive target for treatment of inflammatory diseases driven by autoreactive T cells.

Basiliximab is a chimeric (murine/human) monoclonal antibody that blocks the alpha-chain of the IL-2 receptor complex expressed on activated T lymphocytes, inhibiting the binding of IL-2 to IL-2R (CD25) (Table 86.4). *Basiliximab* is approved for and used primarily in induction regimens for prevention of transplanted allograft rejection. Uncontrolled small case series report manifestations of systemic sclerosis and pulmonary complications associated with amyopathic dermatomyositis responding well to (off-label) treatment with *basiliximab*.[48]

Abatacept (CTLA4-Ig) is a recombinant human protein consisting of the extracellular domain of CTLA-4 linked to the Fc portion of IgG1 (Table 86.4). Prominent among the T-cell co-stimulatory molecules is CD28, which binds CD80/CD86 on antigen-presenting cells. In the context of T-cell activation, CTLA4 expression is upregulated on the surface of T cells, and due to greater binding, avidity to CD80/86 relative to CD28 is preferentially ligated, disrupting further co-stimulation through CD28 as well as engendering inhibitory signaling in the T cell. Although abatacept does not engender inhibitory signaling, by virtue of the higher avidity of the construct for CD80/86, T-cell co-stimulation through CD28 is inhibited.

Although the time to maximum clinical response is somewhat longer than that observed with TNF-α inhibitors, *abatacept* is equally effective in reducing disease activity and inhibiting progression of structural damage in rheumatoid arthritis that is unresponsive to treatment with *methotrexate*.[49] *Abatacept* has no direct impact on phagocytic cell responses, and its use in patients with RA may be associated with fewer bacterial infection complications relative to patients on anti-TNF therapy. It is nonetheless recommended that treatment with *abatacept* be withheld in the context of serious intercurrent microbial infections and should not be used in conjunction with other biologic therapies targeting inflammation. *Abatacept* may be associated with an increased risk of lung cancer. However, the occurrence of lymphoma in patients with RA treated with *abatacept* has not been shown to exceed the expected occurrence in patients with RA. In contrast to TNF-α inhibitors, use of *abatacept* does not promote autoimmune complications and may, therefore, be a preferred option for treatment of patients with RA needing biologic therapy who have overlap features of SLE or other autoimmune disorders.

Belatacept is a second-generation CTLA4-Ig that, compared with *abatacept*, has higher binding to both CD80 and CD86 (Table 86.4). Currently used primarily in organ transplantation, *belatacept* is associated with improved patient and renal allograft survival compared to cyclosporine. Although perceived to be of potential use in autoimmune disorders, studies of *belatacept* use in disorders such as SLE or RA have not yet been reported.

Dupilumab is a monoclonal antibody with specificity for the alpha subunit of the IL-4 receptor, blocking activation of the receptor by IL-4 as well as blocking IL-13/IL-13Rα1 complex binding to IL-4Rα, thereby preventing IL-4Rα-mediated signaling induced by both IL-4 and IL-13 (Table 86.4). In so doing, activation of IL-4–driven Th2 responses in T cells is significantly impaired, with noted decreases in serum levels of Th2 biomarkers including IL-13, antigen-specific IgE, CCL17 and CCL18, as well as intra-lesional Th2-associated gene expression in patients with atopic dermatitis.[50,51] Administration of dupilumab has been shown to be effective in the management of eczema as well as steroid-dependent asthma, improving FEV_1 (forced expiratory volume in one second), reducing steroid requirement, and reducing asthma flares.[50,51] The most commonly reported adverse effects of dupilumab include injection-site reactions, conjunctivitis, and headache. Although the half-life of *dupilumab* in humans is not known, peak serum concentrations of *dupilumab* are noted 7 days following the initial subcutaneous injection with steady-state concentrations reached at 16 weeks with repeat every-other-week dosing; the antibody is undetectable an average of 10 to 13 weeks after the last steady-state dose (Food and Drug Administraon [FDA] package insert).

ON THE HORIZON

IL-13 Inhibitors

Tralokinumab is a fully human monoclonal antibody with specificity for IL-13. Although clinical trials have failed to confirm efficacy in the treatment of asthma, *tralokinumab* has demonstrated efficacy in the management of atopic dermatitis for which an FDA approval application is pending. The monoclonal antibody *lebrikizumab* also targets IL-13, but in so doing, has also been shown to inhibit formation of the IL-13Rα/IL-4Rα signaling complex. Clinical trials with *lebrikizumab* confirm efficacy in the management of atopic dermatitis, but have thus far failed to demonstrate significant clinical impact on the course of steroid-dependent asthma.

KEY CONCEPTS

- Inhibition of BAF/BLyS over a period of 6–12 months decreases the survival and maturation of autoreactive B cells, decreases autoantibody titers, and decreases disease activity in SLE, with minimal impact on pre-existing antibody titers to microbial pathogens and pneumococcal vaccine response.
- Inhibition of T-cell co-stimulation targeting CD80/86 and CD28 via exogenous CTLA4:IgFc constructs is effective in suppressing disease activity in rheumatoid arthritis and rejection of transplanted allografts.
- Antibodies targeting the IL-4α receptor or IL-13 attenuate Th2 responses and are effective in the treatment of severe atopic dermatitis and eczema.

ON THE HORIZON

Recombinant Promoters of Treg Function

In addition to attenuating activation of Th1, Th2, and Th17 responses, alternative strategies are evolving to manage autoimmune disorders by upregulating Treg function. Low doses of recombinant IL-2 (*aldesleukin*) induce proliferation and enhance the function of Treg cells, with one small randomized trial demonstrating clinical efficacy in SLE.[1] *Efavaleukin alfa* is an Fc:IL-2 fusion protein that appears to have increased Treg selectivity compared to rIL-2; early phase studies are underway to assess its potential efficacy in systemic lupus (EudraCT 2020-003509-72). *Recombinant IL-27* may also modulate autoimmune inflammation via promotion of Treg lineage and function by antagonizing IL-6–STAT-3–mediated T-cell commitment to Th17 lineage. Treg function may also be enhanced by *recombinant IL-35*.

INHIBITORS OF MAST CELL ACTIVATION

Omalizumab is a recombinant humanized monoclonal antibody that binds to the Cε3 domain of immunoglobulin E (IgE) (Table 86.5). The binding domain is the same site at which IgE normally binds to both high- and low-affinity FcεRI on mast cells and basophils. As a consequence, free IgE is prevented from binding to the mast cell FcεRI receptor. *Omalizumab* is specific to IgE and does not bind to IgG or IgA. *Omalizumab* also cannot bind to FcεRI or to IgE already attached to FcεRI and, therefore, does not interact with cell-bound IgE or activate mast cells or basophils. *Omalizumab* is most useful and presently approved for use in the treatment of poorly-controlled asthma despite inhaled corticosteroid use and in the setting of documented sensitization to a perennial allergen in the setting of serum IgE levels 30 IU/mL or greater.[51] It is also approved for use in adults and adolescents with chronic idiopathic urticaria who remain symptomatic despite H1 antihistamine treatment. Omalizumab may also decrease the severity of asthma in patients with non-atopic (intrinsic) asthma, occupational asthma, viral-induced asthma, and eosinophilic granulomatosis with polyangiitis, but its use has not been fully studied in these populations to merit approved labeling. Controlled and long-term use studies of *omalizumab* have shown the incidence of adverse events is not significantly increased.[51]

Observed clinical-pathologic responses to treatment with *omalizumab* include a marked downregulation of the surface expression of FcεRI on basophils and mast cells and reductions in FcεR1-mediated production of Th2 cytokines by basophils.

TABLE 86.5 Recombinant Inhibitors of Mast Cell and Eosinophil Activation

Molecule	Construct	Half-Life	Dosing (maintenance)
Omalizumab	aIgE (humanized IgG1κ)	24–26 days	150–300 mg sc q2–4 wk
Mepolizumab	aIL-5 (humanized IgG1κ)	16–22 days	100 mg sc q4 wk
Reslizumab	aIL-5 (humanized IgG4κ)	24 days	3 mg/kg iv q4 wk
Benralizumab	aIL-5Ra (humanized IgG1)	15–18 days	30 mg sc q4-8 wk

Markers of airway inflammation are significantly reduced, with demonstrated reductions in sputum eosinophil counts and reduced numbers of eosinophils, $CD3^+$, $CD4^+$, and $CD8^+$ T lymphocytes, B lymphocytes, interleukin-4 (IL-4)-positive cells, and IgE-positive cells in the bronchial mucosa.[51]

INHIBITORS OF EOSINOPHIL ACTIVATION

Eosinophils mediate airway inflammation in patients with asthma, contribute to vascular and organ inflammation in patients with eosinophilic granulomatosis with polyangiitis (EGPA), and may mediate organ injury in other hypereosinophilic syndromes. Interleukin (IL)-5 is a potent cytokine mediator of eosinophil hematopoiesis and has been shown to mediate eosinophilic inflammation in the airways. Although corticosteroids are potent suppressors of eosinophil survival, proliferation, and function, biologics targeting IL-5 or the IL-5 receptor may prove useful as steroid-sparing therapy in managing chronic disorders mediated predominantly by eosinophils.

Mepolizumab is a monoclonal antibody administered subcutaneously that binds to IL-5, thereby inhibiting its binding to the alpha chain of the IL-5 receptor complex expressed on the surface of eosinophils (Table 86.5). Mepolizumab has demonstrated efficacy in patients with severe asthma with blood eosinophil counts of 150/μL or greater and has been approved by the FDA for maintenance treatment of severe asthma in patients who are age 12 or older with demonstrated hypereosinophilia.[52] *Mepolizumab* is also efficacious and approved for use in the management of EGPA.[53] *Reslizumab* is a monoclonal antibody targeting the IL-5R-binding domain of IL-5 (Table 86.5). Administered intravenously, *reslizumab* has been shown effective in decreasing the frequency and severity of asthma exacerbations and is approved by the FDA for use in adult patients with demonstrated eosinophilia and otherwise treatment-resistant asthma.[52] Adverse events associated with use of *mepolizumab* and *reslizumab* include occasional episodes of severe hypersensitivity reactions during or following administration and occurrence of herpes zoster.

Benralizumab is a monoclonal antibody targeting the IL-5 receptor (Table 86.5). *Benralizumab* blocks activation of eosinophils by IL-5, but also depletes IL-5 receptor-bearing eosinophils and basophils via antibody-dependent cellular cytotoxicity.[52] The recombinant antibody is afucosylated, rendering a high affinity for FcγRIII and enhanced ADCC of IL-5R-bearing cells by NK cells. Benralizumab is approved by the FDA for managing asthma, with additional studies underway to confirm its clinical efficacy and utility in managing other disorders mediated by tissue infiltration of eosinophils.[51,52]

KEY CONCEPTS

- Antibodies targeting the Fc-binding domain of IgE (omalizumab) are effective in the treatment of severe asthma and recurrent mast-cell-mediated chronic urticaria.
- Antibodies targeting IL-5 or the IL-5R-binding domains of IL-5 are effective in the treatment of resistant asthma associated with eosinophilia as well as syndromes associated with eosinophil-mediated tissue injury.

B-CELL AND T-CELL DEPLETING AGENTS

Since B cells have a demonstrated role in the generation of autoantibodies as well as antigen presentation to and co-stimulatory support for autoreactive T cells, strategies to deplete B lymphocytes have been successfully employed in the treatment of autoimmune and inflammatory diseases. Case reports and limited case series have also documented the successful use of combined and T-cell and B-cell depleting strategies in the management of patients with severe, refractory flares of SLE, multiple sclerosis, and GvHD.

Rituximab is a human–murine chimeric monoclonal antibody with specificity for the B-lymphocyte surface antigen CD20, a cell-surface molecule expressed on the surface of pre-B cells through activated mature B cells (Table 86.6). Rituximab induces lysis of $CD20^+$ B cells by several mechanisms, including complement activation, antibody-dependent cell-mediated cytotoxicity (ADCC), and induction of apoptosis. Induction regimens using four weekly doses or using two larger doses administered 2 weeks apart appear to be equally effective in depleting circulating $CD20^+$ B cells, which can last up to 9 months or longer after a single course of therapy. Treatment with rituximab has also been shown to effect significant transient decreases in inflammatory $CD4^+$ T cells that express IL-17; the extent to which this is due to an indirect consequence of $CD20^+$ B-cell depletion or depletion of identified subsets of $CD4^+CD20^+$ T cells is uncertain.[54]

Rituximab improves the signs and symptoms of disease, functional status, quality of life, and slows radiographic progression of disease in patients with RA, with greater clinical responses observed among RA patients who are seropositive for rheumatoid factor (RF) and/or anti-citrullinated peptide (CCP) autoantibody compared to RF- or CCP-seronegative patients.[55] *Rituximab* has also been used with considerable success and is approved for use in the management of ANCA-associated vasculitis syndromes, with efficacy equivalent to or greater than *cyclophosphamide* in the treatment of granulomatosis with polyangiitis (GPA) or microscopic polyangiitis (MPA).[56] In cryoglobulin syndromes, *rituximab* has been shown to decrease cryoglobulin and constituent immunoglobulin titers as well as hasten resolution of cryoglobulin-associated vasculitic skin ulcers and neuropathy, glomerulonephritis, arthritis, and/or hyperviscosity complications.[56] *Rituximab* has been used with variable reported success in the management of other autoantibody-mediated disorders including SLE, primary Sjögren syndrome, inflammatory myopathy, chronic inflammatory demyelinating polyneuropathy (CIDP), multiple sclerosis, and pemphigus.[56]

Minimal increases in serious or opportunistic infections have been reported in patients with RA or ANCA (antineutrophil cytoplasmic autoantibodies) vasculitis treated with repeated cycles of rituximab. Use of rituximab is associated with increased risk of viral infections, including cytomegalovirus, herpes simplex virus, varicella-zoster virus, and hepatitis B virus (HBV), with assessments for latent HBV with HBsAg and IgM-HBcAb recommended prior to dosing. Mild decreases in the overall levels of serum immunoglobulins may be observed during treatment, but immunoglobulin levels are rarely depleted—likely due to the preservation of more mature B cells and plasma cells that have lost surface expression of CD20. However, when used concomitantly with other immunosuppressive agents impacting lymphocyte proliferation, significant hypogammaglobulinemia may ensue over time due to inability to replenish the plasma cell compartment. Reactivation of JC virus with development of progressive multifocal leukoencephalopathy (PML)—a debilitating and often fatal demyelinating disease of the central nervous system—has been reported among patients with hematologic malignancies, SLE, and RA treated with *rituximab*, but usually in the context of concomitant treatment with other therapies impacting lymphocyte survival and proliferation.[57]

Ofatumumab is a fully human mAb that binds to an epitope, encompassing both small and large loops of the extracellular domain of the CD20 cell surface antigen on B lymphocytes (Table 86.6). The binding epitope of *ofatumumab* is distinct from that of *rituximab*, residing more proximate to the cell membrane. In comparative studies with *rituximab* using chronic lymphocytic leukemia (CLL) cells, *ofatumumab* elicits similar ADCC but elicits greater complement-dependent cytotoxicity (CDC), presumably due to the greater proximity of the binding site to the cell membrane and/or binding affinity to CD20 epitopes.[58] *Ofatumumab* is currently approved for use in managing multiple sclerosis.

Ocrelizumab is a humanized anti-CD20 monoclonal antibody that depletes CD20-positive B cells through both complement-dependent cytotoxicity and antibody-dependent cellular cytotoxicity (Table 86.6). It is currently FDA approved for the treatment of both relapsing-remitting and primary/secondary progressive variants of multiple sclerosis.[58]

Obinutuzumab is a humanized, glycoengineered type II anti-CD20 monoclonal antibody that is approved for treatment of B-cell lineage lymphomas (Table 86.6). Glycoengineering the Fc portion of obinutuzumab results in a higher affinity of the antibody for FcγRIII receptors on immune effector cells such as natural killer (NK) cells and phagocytic cells, effecting greater induction of direct cell death and ADCC and cellular phagocytosis relative to *rituximab* and *ofatumumab*. However, complement-dependent cytotoxicity (CDC) is significantly reduced with *obinutuzumab*.[58] Following encouraging phase II trial results, a phase III trial is currently underway to determine whether *obitnutuzumab* improves remission rates when added to standard induction regimens for lupus nephritis (NCT04221477).

Inebilizumab is a humanized afucosylated IgG1 monoclonal reagent with specifity for CD19, effecting depletion of $CD19^+$ B

TABLE 86.6 Recombinant T- and B-Cell Depleting Agents

Molecule	Construct	Half-Life	Dosing (maintenance)
Rituximab	aCD20 (chimeric IgG1)	18–23 days	375 mg/m² iv q wk ×4 (q4–6 mo) 1000 mg iv q2 wk ×2 (q4–6 mo)
Ofatumumab	aCD20 (human IgG1)	17 days	1000 mg iv q4–8 wk
Ocrelizumab	aCD20 (humanized IgG1)	26–28 days	600 mg iv every 6 mo
Obinutuzumab	aCD20 (humanized IgG1)	24–36 days	1000 mg iv q2 wk×2
Inebilizumab	aCD19 (humanized IgG1)	16–18 days	300 mg iv every 6 m
Alemtuzumab	aCD52 (humanized IgG1)	1–14 days	12 mg iv daily× 5 days 10–30 mg sc 3×/wk

cells primarily via ADCC (Table 86.6). As CD19 is also found on plasmablasts and most plasma cells as well as pro-B cells, targeting CD19-positive cells could be a more potent strategy for controlling B-cell-mediated diseases than anti-CD20 therapy.[59] Currently approved for just treatment of aquaporin-4 antibody-positive neuromyelitis optica spectrum disorder, inebilizumab is being studied in other disorders mediated by known autoantibodies in which responses to antibodies specific for CD20 have been suboptimal.

Alemtuzumab is a monoclonal antibody with specificity for CD52, an antigen present on the surface of B and T lymphocytes as well as the majority of monocytes, macrophages, NK cells, and a subpopulation of neutrophils (Table 86.6). Approved for use in the treatment of B-cell CLL and relapsing-remitting multiple sclerosis, *alemtuzumab* has also been used with success for the treatment of T-cell prolymphocytic leukemia, prevention and for treatment of acute GvHD, and prevention of allograft rejection. *Alemtuzumab* has also been used off-label with reported success in the treatment of patients with severe SLE and Behçet disease refractory to other treatments. Despite the depletion of T lymphocyte, B lymphocyte, NK cell, and monocyte populations following treatment, reported rates of serious infections following treatment with *alemtuzuma*b are not significantly increased compared to other immunosuppressive regimens employed to manage the disorders for which it is used. However, there is a significant occurrence of secondary autoimmunity following use of *alemtuzumab*, possibly due to homeostatic proliferation of self-reactive memory T cells in the absence of an effective Treg response during immune reconstitution.[60]

KEY CONCEPTS

- Monoclonal antibody-mediated depletion of CD20⁺ B lymphocytes is effective in the treatment of ANCA-associated vasculitis, cryoglobulin syndromes, and rheumatoid arthritis; the best responses in RA patients are those with significant elevations in antibodies specific for rheumatoid factor and CCP.
- There are subsets of T cells that also express CD20, and part of the clinical efficacy of anti-CD20 monoclonal reagents may be due in part to depletion of these subsets of T cells.
- Monoclonal reagents with greater binding affinity and binding sites more proximate to the cell membrane appear to have greater efficacy to deplete CD20⁺ lymphocytes.
- Depletion of CD20⁺ lymphocytes has been associated with reactivation of hepatitis B virus, and vigilance for latent HBV reactivation is prudent when using anti-CD20 monoclonal reagents.
- Monoclonal reagents targeting CD52 target both T-helper as well as Treg T-cell subsets; imbalances in relative homeostatic T-cell reconstitution following treatment may result in autoimmune complications.

REFERENCES

1. He J, Zhang R, Shao M, et al. Efficacy and safety of low-dose IL-2 in the treatment of systemic lupus erythematosus: a randomised, double-blind, placebo-controlled trial. *Ann Rheum Dis*. 2020;79(1):141–149.
2. Kötter I, Hamuryudan V, Oztürk ZE, Yazici H. Interferon therapy in rheumatic diseases: state-of-the-art 2010. *Curr Opin Rheumatol*. 2010;22(3):278–283.
3. Kasper LH, Reder AT. Immunomodulatory activity of interferon-beta. *Ann Clin Transl Neurol*. 2014;1(8):622–631.
4. Bonaci-Nikolic B, Jeremic I, Andrejevic S, et al. Anti-double stranded DNA and lupus syndrome induced by interferon-beta therapy in a patient with multiple sclerosis. *Lupus*. 2009;18(1):78–80.
5. The International Chronic Granulomatous Disease Cooperative Study Group. A controlled trial of interferon gamma to prevent infection in chronic granulomatous disease. *N Engl J Med*. 1991;324(8):509.
6. Ridderstad A, Abedi-Valugerdi M, Möller E. Cytokines in rheumatoid arthritis. *Ann Med*. 1991;23(3):219–223.
7. Aaltonen KJ, Virkki LM, Malmivaara A, et al. Systematic review and meta-analysis of the efficacy and safety of existing TNF blocking agents in treatment of rheumatoid arthritis. *PLoS One*. 2012;7(1):e30275.
8. Boehncke WH, Prinz J, Gottlieb AB. Biologic therapies for psoriasis. A systematic review. *J Rheumatol*. 2006;33(7):1447.
9. Peyrin-Biroulet L, Deltenre P, de Suray N, et al. Efficacy and safety of tumor necrosis factor antagonists in Crohn's disease: meta-analysis of placebo-controlled trials. *Clin Gastroenterol Hepatol*. 2008;6(6):644.
10. Bonovas S, Minozzi S, Lytras T, et al. Risk of malignancies using anti-TNF agents in rheumatoid arthritis, psoriatic arthritis, and ankylosing spondylitis: a systematic review and meta-analysis. *Expert Opin Drug Saf*. 2016;15(sup1):35–54.
11. Pérez-De-Lis M, Retamozo S, Flores-Chávez A, et al. Autoimmune diseases induced by biological agents. A review of 12,731 cases (BIOGEAS Registry). *Expert Opin Drug Saf*. 2017;16(11):1255–1271.
12. Cantaert T, De Rycke L, Mavragani CP, et al. Exposure to nuclear antigens contributes to the induction of humoral autoimmunity during tumour necrosis factor alpha blockade. *Ann Rheum Dis*. 2009 Jun;68(6):1022–1029.
13. Kranich J, Krautler NJ, Heinen E, et al. Follicular dendritic cells control engulfment of apoptotic bodies by secreting Mfge8. *J Exp Med*. 2008;205(6):1293–1302.
14. So A, Dumusc A, Nasi S. The role of IL-1 in gout: from bench to bedside. *Rheumatology* (Oxford). 2018;57(suppl_1):i12–i19.
15. Bruck N, Suttorp M, Kabus M, et al. Rapid and sustained remission of systemic juvenile idiopathic arthritis-associated macrophage activation syndrome through treatment with anakinra and corticosteroids. *J Clin Rheumatol*. 2011;17(1):23–27.
16. Ortiz-Sanjuán F, Blanco R, Riancho-Zarrabeitia L, et al. Efficacy of anakinra in refractory adult-onset Still's disease: multicenter study of 41 patients and literature review. *Medicine (Baltimore)*. 2015;94(39):e1554.
17. Shakoory B, Carcillo JA, Chatham WW, et al. Interleukin-1 receptor blockade is associated with reduced mortality in sepsis patients with features of macrophage activation syndrome: reanalysis of a prior phase III trial. *Crit Care Med*. 2016;44(2):275–281.
18. Klein AL, Lin D, Cremer PC, et al. Efficacy and safety of rilonacept for recurrent pericarditis: results from a phase II clinical trial. *Heart*. 2020;0:1–9. https://doi.org/10.1136/heartjnl-2020-317928.
19. Hunter CA, Jones SA. IL-6 as a keystone cytokine in health and disease. *Nat Immunol*. 2015;16(5):448–457.
20. Rubbert-Roth A, Furst DE, Nebesky JM, Jin A, Berber E. A review of recent advances using tocilizumab in the treatment of rheumatic diseases. *Rheumatol Ther*. 2018;5(1):21–42.
21. Kotch C, Barrett D, Teachey DT. Tocilizumab for the treatment of chimeric antigen receptor T cell-induced cytokine release syndrome. *Expert Rev Clin Immunol*. 2019;15(8):813–822.
22. Duchow A, Bellmann-Strobl J. Satralizumab in the treatment of neuromyelitis optica spectrum disorder. *Neurodegener Dis Manag*. 2021;11(1):49–59. https://doi.org/10.2217/nmt-2020-0046.
23. Burmester GR, Lin Y, Patel R, et al. Efficacy and safety of sarilumab monotherapy versus adalimumab monotherapy for the treatment of patients with active rheumatoid arthritis (MONARCH): a randomised, double-blind, parallel-group phase III trial. *Ann Rheum Dis*. 2017;76(5):840–847.
24. van Rhee F, Wong RS, Munshi N. Siltuximab for multicentric Castleman's disease: a randomised, double-blind, placebo-controlled trial. *Lancet Oncol*. 2014;15:966–974.
25. Shaw S, Bourne T, Meier C, et al. Discovery and characterization of olokizumab: a humanized antibody targeting interleukin-6 and neutralizing gp130-signaling. *MAbs*. 2014;6(3):774–782.
26. Tang C, Chen S, Qian H, et al. IL-23 as a drug target for autoimmune inflammatory disease. *Immunology*. 2012;135(2):112–124.
27. Ritchlin C, Rahman P, Kavanaugh A, et al.; PSUMMIT 2 Study Group. Efficacy and safety of the anti-IL-12/23 p40 monoclonal antibody, ustekinumab, in patients with active psoriatic arthritis despite conventional

non-biological and biological anti-tumour necrosis factor therapy: 6-month and 1-year results of the phase 3, multicentre, double-blind, placebo-controlled, randomised PSUMMIT 2 trial. *Ann Rheum Dis*. 2014;73(6):990–999.
28. Wils P, Bouhnik Y, Michetti P, et al. Groupe d'Etude Thérapeutique des Affections Inflammatoires du Tube Digestif (GETAID). Subcutaneous *ustekinumab* provides clinical benefit for two-thirds of patients with Crohn's disease refractory to anti-tumor necrosis factor agents. *Clin Gastroenterol Hepatol*. 2016;14(2):242–250.
29. Bai F, Li GG, Liu Q, et al. Short-term efficacy and safety of IL-17, IL-12/23, and IL-23 inhibitors brodalumab, secukinumab, ixekizumab, ustekinumab, guselkumab, tildrakizumab, and risankizumab for the treatment of moderate to severe plaque psoriasis: a systematic review and network meta-analysis of randomized controlled trials. *J Immunol Res*. 2019;2019:2546161.
30. Isailovica N, Daigob K, Mantovanib A, et al. Interleukin-17 and innate immunity in infections and chronic inflammation. *J Autoimmunity*. 2015;60:1–11.
31. Ly K, Smith MP, Thibodeaux Q, et al. Anti IL-17 in psoriasis. *Expert Rev Clin Immunol*. 2019;15(11):1185–1194.
32. Baeten D, Sieper J, Braun J, et al. MEASURE 1 Study Group; MEASURE 2 Study Group. Secukinumab, an interleukin-17A inhibitor, in ankylosing spondylitis. *N Engl J Med*. 2015;373(26):2534–2548.
33. Hueber W, Sands B, Lewitzky S, et al. Secukinumab, a human anti-IL-17A monoclonal antibody, for moderate to severe Crohn's disease: unexpected results of a randomised, double-blind placebo-controlled trial. *Gut*. 2012;61:1693–1700.
34. van der Heijde D, Gensler LS, Deodhar A, et al. Dual neutralisation of interleukin-17A and interleukin-17F with bimekizumab in patients with active ankylosing spondylitis: results from a 48-week phase IIb, randomised, double-blind, placebo-controlled, dose-ranging study. *Ann Rheum Dis*. 2020;79(5):595–604.
35. Vallurupalli M, Berliner N. Emapalumab for the treatment of relapsed/refractory hemophagocytic lymphohistiocytosis. *Blood*. 2019;134(21):1783–1786.
36. Gabr JB, Liu E, Mian S, et al. Successful treatment of secondary macrophage activation syndrome with emapalumab in a patient with newly diagnosed adult-onset Still's disease: case report and review of the literature. *Ann Transl Med*. 2020;8(14):887.
37. Morand EF, Furie R, Tanaka Y, et al. Trial of anifrolumab in active systemic lupus erythematosus. *N Engl J Med*. 2020;382(3):211–221.
38. Cofiell R, Kukreja A, Bedard K, et al. Eculizumab reduces complement activation, inflammation, endothelial damage, thrombosis, and renal injury markers in aHUS. *Blood*. 2015;125(21):3253–3262.
39. Kello N, Khoury LE, Marder G, et al. Secondary thrombotic microangiopathy in systemic lupus erythematosus and antiphospholipid syndrome, the role of complement and use of eculizumab: case series and review of literature. *Semin Arthritis Rheum*. 2019;49(1):74–83.
40. Kulasekararaj AG, Hill A, Rottinghaus ST, et al. Ravulizumab (ALXN1210) vs eculizumab in C5-inhibitor-experienced adult patients with PNH: the 302 study. *Blood*. 2019;133(6):540–549.
41. Banerji A, Riedl MA, Bernstein JA, et al. Effect of lanadelumab compared with placebo on prevention of hereditary angioedema attacks: a randomized clinical trial. *JAMA*. 2018;320(20):2108–2121.
42. Keeley KA, Rivey MP, Allington DR. Natalizumab for the treatment of multiple sclerosis and Crohn's disease. *Ann Pharmacother*. 2005;39(11):1833–1843.
43. Glatigny S, Duhen R, Oukka M, et al. Cutting edge: loss of α4 integrin expression differentially affects the homing of Th1 and Th17 cells. *J Immunol*. 2011;187(12):6176–6179.
44. Sands BE, Feagan BG, Rutgeerts P, et al. Effects of vedolizumab induction therapy for patients with Crohn's disease in whom tumor necrosis factor antagonist treatment failed. *Gastroenterology*. 2014;147(3):618–627.
45. Blair HA, Duggan ST. Belimumab. A review in systemic lupus erythematosus. *Drugs*. 2018;78(3):355–366.
46. Furie R, Rovin BH, Houssiau F, et al. Two-year, randomized, controlled trial of belimumab in lupus nephritis. *N Engl J Med*. 2020;383(12):1117–1128.
47. Stohl W, Hiepe F, Latinis KM, et al. Belimumab reduces autoantibodies, normalizes low complement, and reduces select B-cell populations in patients with systemic lupus erythematosus. *Arthritis Rheum*. 2012;64(7):2328–2337.
48. Becker MO, Brückner C, Scherer HU, et al. The monoclonal anti-CD25 antibody basiliximab for the treatment of progressive systemic sclerosis: an open-label study. *Ann Rheum Dis*. 2011;70(7):1340–1341.
49. Maxwell L, Singh JA. Abatacept for rheumatoid arthritis. *Cochrane Database Syst Rev*. 2009;2009(4):CD007277.
50. Guttman-Yassky E, Bissonnette R, Ungar B, et al. Dupilumab progressively improves systemic and cutaneous abnormalities in patients with atopic dermatitis. *J Allergy Clin Immunol*. 2019;143(1):155–172.
51. Agache I, Rocha C, Beltran J, et al. Efficacy and safety of treatment with biologicals (benralizumab, dupilumab and omalizumab) for severe allergic asthma: A systematic review for the EAACI Guidelines—recommendations on the use of biologicals in severe asthma. *Allergy*. 2020;75(5):1043–1057.
52. Farne HA, Wilson A, Powell C, et al. Anti-IL5 therapies for asthma. *Cochrane Database Syst Rev*. 2017;9(9):CD010834.
53. Wechsler ME, Akuthota P, Jayne D, et al.; EGPA Mepolizumab Study Team. Mepolizumab or placebo for eosinophilic granulomatosis with polyangiitis. *N Engl J Med*. 2017;376(20):1921–1932.
54. Alunno A, Carubbi F, Bistoni O, et al. Interleukin (IL)-17-producing pathogenic T lymphocytes co-express CD20 and are depleted by rituximab in primary Sjögren's syndrome: a pilot study. *Clin Exp Immunol*. 2016;184(3):284–292.
55. Lopez-Olivo MA, Amezaga Urruela M, McGahan L, et al. Rituximab for rheumatoid arthritis. *Cochrane Database Syst Rev*. 2015;1:CD007356.
56. Kaegi C, Wuest B, Schreiner J, et al. Systematic review of safety and efficacy of rituximab in treating immune-mediated disorders. *Front Immunol*. 2019;10:1990.
57. Focosi D, Tuccori M, Maggi F. Progressive multifocal leukoencephalopathy and anti-CD20 monoclonal antibodies: What do we know after 20 years of rituximab. *Rev Med Virol*. 2019;29(6):e2077.
58. Du FH, Mills EA, Mao-Draayer Y. Next-generation anti-CD20 monoclonal antibodies in autoimmune disease treatment. *Auto Immun Highlights*. 2017;8(1):12.
59. Cree BAC, Bennett JL, Kim HJ, et al. Inebilizumab for the treatment of neuromyelitis optica spectrum disorder (N-MOmentum): a double-blind, randomised placebo-controlled phase 2/3 trial. *Lancet*. 2019;394:1352–1363.
60. Jones JL, Coles AJ. Mode of action and clinical studies with alemtuzumab. *Exp Neurol*. 2014;262(Pt A):37–43.

87

Vaccines

Katherine V. Houser, Myra Happe, Rachel Bean, and Emily E. Coates

INTRODUCTION

Vaccines are clinically simple but immunologically complex interventions that can dramatically reduce morbidity and mortality due to diseases across all age groups. Vaccines offer an elegant solution to infectious diseases as they provide a societal benefit that reaches beyond individual protection. Vulnerable community members whose immune systems are less able to adequately respond to vaccines (newborns, immunocompromised persons, the elderly) or who are unable to receive vaccines (due to allergy or a medical contraindication) depend on immunization of surrounding community members for protection against vaccine-preventable diseases. In pregnant women, vaccinations may offer the double benefit of protecting both mother and infant against the targeted pathogen.[1]

While older adults (>65 years old) experience high proportions of the total morbidity and mortality for several vaccine-preventable diseases (*e.g.*, seasonal influenza, pneumococcal disease, herpes zoster) due to immunosenescence, they are less able to mount their own protective immune responses after vaccination. Vaccination of children, who are the primary spreaders of many vaccine-preventable infectious diseases, and younger adults can provide dramatic reductions in disease incidence in older adults through community protection. Despite suboptimal vaccine responses with aging, several vaccines are specifically recommended for older adults. In addition, some vaccines for older adults now employ novel strategies to enhance immunogenicity, including higher antigen doses[2] and the addition of an adjuvant.[3]

Vaccines against microbes are increasingly appreciated for their potential role in the heightening battle against antimicrobial-resistant pathogens. Preventing an illness through vaccination obviates the need to treat a bacterial infection with antibiotics, thereby avoiding potential induction of antibiotic resistance in either the targeted pathogenic bacterium or the patient's healthy microbiota. The sparing of antimicrobial use to prevent the emergence of resistance can be considered another form of public health intervention provided by vaccines.

Over the last 300 years, vaccinology has made impressive advances in combating human suffering and death caused by infectious diseases. These advances have accelerated rapidly in the past century with the explosion of knowledge in microbiology, immunology, and genetics. Current scientific understanding has answered many questions about immunity and how to provide it through vaccination, yet significant challenges remain and are joined by emerging epidemic and pandemic infectious diseases with alarming regularity. The next-generation tools of rational vaccine design are anticipated to yield important and life-saving innovations.

This chapter first reviews selected events in the history of vaccination. The remarkable accomplishments that have resulted from programs of vaccination to date are then highlighted. We review important recent milestones in vaccine development strategies that have the potential to revolutionize the field and offer great hope for unmet vaccine needs. Vaccine development in response to recent epidemics and pandemics is reviewed. Current vaccination recommendations in the United States and around the world are then summarized. Finally, we discuss present and future challenges for the field of vaccinology.

CLINICAL RELEVANCE

- Vaccines are highly effective interventions for preventing infectious diseases with public health importance.
- Both individual protection and community (herd) immunity result from vaccination programs.
- The reductions in disease burden (morbidity and mortality) achieved through implementation of childhood vaccination programs are extraordinary.
- Vaccination is not just for children: in recent years, new adolescent and adult vaccines have become available and are now recommended.
- Clinicians of all specialties should take vaccine histories and provide access to vaccines relevant to their patients' ages and medical conditions. Access can be provided through referral or by stocking and administering the indicated vaccines.

HISTORY OF VACCINATION

The earliest known vaccines were against smallpox and were used in Asia in the second millennium. The practice was called variolation and involved exposing, usually through the intranasal route, a susceptible person to material from the dried scabs of a smallpox victim. If the recipient survived, she/he was protected against future smallpox disease. Since natural smallpox had a 30% mortality rate, and variolation had a lower (~1%) mortality rate, this ancient practice was an early example of weighing the risk-to-benefit ratio for a human health intervention. By the 1700s, variolation was employed in societies in Africa, India, the Ottoman Empire, England, and France (https://www.nlm.nih.gov/exhibition/smallpox/sp_variolation.html). The practice of variolation involved inherent risks, including occasional outbreaks of a mild form of the disease.

An English physician was searching for a safer alternative to variolation and would become known as the father of vaccinology. Dr. Edward Jenner performed a smallpox vaccination experiment on James Phipps on May 14, 1796, using cowpox pus from lesions on the hands of a milkmaid.[4] Dr. Jenner then collected lesion material from a smallpox patient to use as a viral challenge. Phipps survived both the vaccination and the challenge.

Dr. Jenner's work was disliked by some because of the introduction of a cow virus into humans. Other opponents of vaccination were those with financial interests in lucrative variolation practices. When vaccination in England was made compulsory by the Vaccination Act of 1853, an organized anti-vaccine movement quickly arose. Incredibly, even in the present day and despite the evidence supporting the safety and effectiveness of licensed vaccines, organized anti-vaccine movements continue to challenge contemporary clinicians and public health officials. The internet and social media have facilitated self-publication with rapid wide dissemination of misinformation, anti-vaccine propaganda, and pseudoscience that circumvent traditional scientific peer review to feed on the general public's fears and misunderstandings.

KEY CONCEPTS

Jenner's Work on Smallpox Vaccination Highlights Many Dimensions Relevant for Translational Vaccinology Today

- **Disease burden, surveillance, epidemiology** A significant and unacceptable burden of smallpox disease drove development of a safer intervention to improve public health.
- **Innovation** Jenner's innovation resulted from the need for an improved biomedical intervention to address the significant risk of harm associated with the centuries-old variolation practice.
- **Clinical insight** An observation that dairymaids who had recovered from an occupational illness (cowpox) were seldom affected by smallpox led to Jenner's promotion of smallpox vaccination. The observation of the protected state (immunity) in dairymaids led to a concept that was tested and promoted by Jenner.
- **Post-vaccination challenge** After the vaccination procedure, Jenner's subjects were subsequently intentionally exposed to (challenged with) wild-type smallpox and observed for safety and disease outcomes. Human challenge with smallpox would not be considered ethical today, although human challenge experiments are performed when developing vaccines for certain self-limited or treatable infectious diseases.
- **Presentation of experimental results** To disseminate his scientific findings and advocate for wider vaccination deployment, Jenner presented his work to the Royal Society and then self-published his manuscript after it was rejected for publication.
- **Branding** The name "vaccination" was applied to the intervention. Vacca is the Latin word for cow.
- **Anti-vaccination movement and conflicts of interest** Jenner experienced significant opposition to his vaccine from groups opposed to the new technique and from individuals with variolation practices who faced financial losses as public acceptance of vaccination grew.

While Dr. Jenner's smallpox challenge experiment presented a high risk to the participant that may be questioned by today's standards, certain human challenge studies remain safe, acceptable, and valuable today. Human challenge studies are performed for self-limited and/or treatable infections in order to study vaccine and therapeutic efficacy or to characterize the host response to the infection in detail, for example, influenza,[5] primary dengue,[6] norovirus,[7] and malaria.[8] A human challenge experiment can rapidly provide feedback to vaccine developers and public health officials to help prioritize resource-intensive field trial evaluations of promising candidate vaccines. If an encouraging preliminary efficacy signal is observed in a post-vaccination human challenge trial, it may support vaccine approval by regulatory agencies. In 2020 the US Food and Drug Administration (FDA) approved a single-dose live oral cholera vaccine, Vaxchora, targeting *Vibrio cholerae* serogroup O1, representing the first time human challenge data have supported vaccine approval by the FDA. For the pivotal efficacy trial, participants received vaccination with Vaxchora, followed by controlled human infection with *Vibrio cholerae*.[9] Vaccine efficacy was found to be 90% at 10 days and 80% at 3 months post-vaccination.

A second phase of vaccination's history ensued over the nineteenth century with the emergence of the germ theory, in which infectious diseases were caused by microorganisms too small to be seen without magnification. Robert Koch (1843–1910) and Louis Pasteur (1822–1895) contributed many key observations and experiments regarding both infectious diseases and vaccines. Koch's four postulates laid out the requirements for establishing causality of infectious diseases by microbes and proved that *Bacillus anthracis* was the cause of anthrax, providing the first proof of a microbial etiology of a specific disease. Through attenuation or inactivation of the wild-type microbes, Pasteur produced vaccines that induced protection against a number of diseases. He performed a number of classical vaccination and challenge experiments to show that vaccines would protect susceptible farm animals from devastating veterinary pathogens like chicken cholera and anthrax, or human pathogens like rabies.[10]

In the early twentieth century, passive immunization was developed as a therapy for infectious diseases. While active immunization involves administering a vaccine to trigger a protected state (immunity), passive immunization involves transferring the protective proteins (antibodies) from an immune donor to a susceptible patient without administering a vaccine. Emil von Bering administered sera from immune horses to humans to cure and prevent diphtheria and was awarded the Nobel Prize in 1901 for his work (http://www.nobelprize.org/nobel_prizes/medicine/laureates/1901/behring-facts.html).

Laboratory growth of poliovirus permitted the development of both the inactivated polio vaccine (IPV; Salk, licensed in 1955) and the live-attenuated oral polio vaccine (OPV; Sabin, monovalent licensed in 1961, trivalent in 1963). As a result of those vaccines, poliovirus type 2 was eradicated in 1999, and no wild-type poliovirus type 3 has been detected since 2012. A historic milestone occurred in August 2020, when after four years of no new reported cases, Africa was officially declared poliovirus type 1 free (https://www.cdc.gov/polio/why-it-matters/africa-kicks-out-wild-polio.htm). However, it remains endemic in other regions. According to CDC, there were 176 cases of poliovirus type 1 reported in two countries in 2019: 29 (16%) in Afghanistan and 147 (84%) in Pakistan. In February 2022, after five years of no cases reported in Africa, a case of wild-type poliovirus type I was detected in a young child in Lilongwe, Malawi. However, due to the isolate being genetically linked to a sequence in Pakistan's Sindh Province, the continent's wild poliovirus-free certification remains unaffected (https://www.who.int/emergencies/disease-outbreak-news/item/wild-poliovirus-type-1-(WPV1)-malawi). While the progress toward global polio eradication is impressive, final eradication will require internationally coordinated efforts, persistence in vaccinating endemic countries' populations, and sustained attention to surveillance.

Several other live attenuated viral vaccines developed in the late twentieth century, such as measles, mumps, and rubella, have become staples of childhood vaccination programs in the US and globally. The development of the Oka strain of the varicella zoster virus led to live attenuated vaccines for both chicken pox in children and herpes zoster in older adults. To produce these vaccines, the serial passage of wild-type viruses promotes viral adaptation for growth in cell cultures and diminishes virulence in humans. Importantly, these attenuated vaccine-strain

viruses are not only well-tolerated and safe in humans but retain the ability to provoke protective immune responses.

Recognition and subsequent exploitation of key antigenic substructures rather than whole microbes were important technical advances. The studies of the polysaccharide capsules of *Streptococcus pneumonia*[11] and M proteins of *Streptococcus* species,[12] respectively, led to characterization, isolation, and serotyping of these bacterial structures and their recognition as key antigens in immunity to Streptococcal diseases. Such observations led eventually to safer vaccination with components (subunits) of pathogens, as opposed to entire microbes. When delivered as vaccines, these isolated microbial components produce protective antibody and cellular immune responses but do not cause the disease induced by the complete wild-type organisms.

The polysaccharide vaccines developed for the prevention of bacterial diseases caused by *Haemophilus influenzae*, *Neisseria meningitidis*, and *Streptococcus pneumoniae* were also welcome advances. These bacterial polysaccharides were covalently coupled, or conjugated, to a protein carrier such as tetanus or diphtheria toxoids. This maneuver converted the T-cell-independent polysaccharide vaccines into T-cell-dependent protein-polysaccharide conjugate vaccines and resulted in B cell memory, improved immunity, utility in newborns, and herd immunity.[13]

The molecular biology revolution of the Twenty-First Century, in particular recombinant DNA technology, along with fundamental dissections of the innate and adaptive immune responses, have generated novel approaches to vaccination. Some of these next-generation vaccine platforms, including nucleic acid vaccines and viral-vectored vaccines, are discussed further in the sections below.

ACCOMPLISHMENTS OF VACCINATION

It is generally believed that elimination of an infectious disease from human circulation through vaccination can be achieved only when the following conditions are met: (1) the pathogen has no animal reservoir, and (2) the vaccine induces long-lasting immunity (Table 87.1). Smallpox eradication was achieved after a worldwide vaccination campaign and is the signature accomplishment of vaccination. The fields of medicine and public health celebrate this remarkable success as it showcases the power of vaccines to improve human health. Smallpox was a scourge of humanity for millennia, disfiguring and blinding survivors and killing 30% of those infected. The world's last known naturally occurring smallpox case occurred in Somalia in 1977. After the disease was eliminated, routine vaccination against smallpox in the general public was discontinued since it was no longer necessary for prevention. In 1980, the World Health Organization (WHO) certified that smallpox had been eradicated.[14]

The US Centers for Disease Control and Prevention (CDC) designated vaccination as first on the list of the ten greatest public health achievements of the twentieth century,[15] and the WHO named "Vaccine Hesitancy" as one of the "Ten threats to global health in 2019" (https://www.who.int/news-room/spotlight/ten-threats-to-global-health-in-2019). In addition to smallpox eradication, the control of many common childhood infections and attendant reductions in morbidity and mortality are great achievements. Implementation of routine US childhood immunization programs led to major reductions from mid-twentieth century disease peaks to record low levels for several infectious diseases today (Table 87.2). For example, in the US, the incidence of polio, measles, rubella, and mumps declined by 100%, 99.9%, 99.9%, and 95.9%, respectively.[16] It is estimated that for each annual birth cohort of approximately four million US children, vaccines in the US childhood immunization schedule prevent an estimated 20 million cases of disease and 42,000 deaths.[17] Furthermore, while it is true that a considerable investment of resources is required to complete the annual programs of childhood vaccination, vaccines result in very significant cost savings, hence are highly cost-effective interventions. For each annual US birth cohort, vaccines result in nearly $14 billion in annual net direct cost savings and $69 billion in annual net societal cost savings, including reductions in parental missed work to care for an ill child.[17]

TABLE 87.1 Stages of Reduction of Infectious Disease Incidence by Vaccination and Other Prevention Interventions

- **Control**. The reduction of disease incidence and prevalence to a locally acceptable level due to vaccination and/or other interventions; continued interventions are needed to maintain the reduction. Example: diarrheal diseases.
- **Elimination of disease**. Reduction to zero of the incidence of a specified disease in a defined geographical area as a result of vaccination and/or other interventions; continued measures are required. Example: neonatal tetanus.
- **Elimination of infection**. Reduction to zero of the incidence of infection caused by a specific agent in a defined geographical area as a result of vaccination and/or other interventions; continued measures to prevent re-establishment of transmission are required. Example: poliomyelitis elimination from North America.
- **Eradication**. Permanent reduction to zero of the worldwide incidence of infection due to a specific agent as a result of vaccination and/or other prevention efforts; interventions are no longer needed. Example: smallpox.
- **Extinction**. An infectious agent no longer exists in either nature or the laboratory. Example: none.

Adapted from Dowdle WR. The principles of disease elimination and eradication. *Bull World Health Organ.* 1998;76(suppl 2):22–25.

TABLE 87.2 Historical Comparisons of Morbidity and Mortality for Vaccine-Preventable Diseases in the United States

Disease	Pre-Vaccination: Estimated Annual Average Number of Cases	Post Vaccination: Annual Cases (Reported or Estimated) in Year 2006	% Reduction
Diphtheria	21,053	0	100
Measles	530,217	55	99.9
Mumps	162,344	6,584	95.9
Pertussis	200,752	15,632	92.2
Paralytic Poliomyelitis	16,316	0	100
Rubella	47,745	11	99.9
Smallpox	29,005	0	100
Tetanus	580	41	92.9
Hepatitis A	117,333	15,296	87
Acute hepatitis B	66,232	13,169	80.1
Invasive Hib	20,000	<50	99.8
Invasive pneumococcal disease	63,067	41,550	34.1
Varicella	4,085,120	48,445	85

Adapted from Roush SW, Murphy TV, Vaccine-Preventable Disease Table Working Group. Historical comparisons of morbidity and mortality for vaccine-preventable diseases in the United States. *JAMA.* 2007;298(18):2155–2163.

Vaccines protect the recipient against disease and reduce transmission of disease-causing microbes to unvaccinated persons. The term for this protection is herd immunity or community immunity. A disease that has been studied closely with regard to community immunity is measles. Measles is highly contagious and easily recognizable in epidemic form. Clustering of poor vaccination coverage often occurs in communities, as demonstrated in recent measles outbreaks in the United States. In 2019, CDC reported 1282 individual cases reported in 31 states, the largest number of cases reported in the US since 1992. The majority of these cases (~89%) were among people who were not vaccinated or whose vaccination status was unknown.[18] Such outbreaks point out the importance of community immunity to protect vulnerable (unvaccinated) members of our communities. Given that many of the recent measles outbreaks in the United States have been linked to imported cases, another important lesson is that as long as a vaccine-preventable, highly transmissible infectious disease exists anywhere, it remains a potential threat everywhere—and global vaccination programs will continue to be important to ensure the health of all community members.

Another powerful example of vaccine-induced community immunity comes from pneumococcal vaccines. There are many unique challenges relating to pneumococcal vaccines: a large number of circulating serotypes, suboptimal immunogenicity of polysaccharide-only vaccines, and noninvasive carriage of the organism. In spite of these, introduction of the pneumococcal conjugate vaccines in infants in 2000 not only led to decreased invasive disease among vaccinated children, but also produced a significant decrease in adults, particularly among older adults in whom this bacterium frequently causes pneumonia.[19] The impact of this vaccine highlights the effectiveness of community immunity produced by vaccines.

Other recently introduced vaccines have made significant impact in relatively brief periods. Before the 2006 implementation of routine rotavirus vaccination, rotavirus infections were a significant cause of severe gastroenteritis in young children and accounted for an estimated 410,000 physician visits, 205,000 to 272,000 emergency department visits, and 55,000 to 70,000 hospitalizations annually with total costs of up to $1 billion in the US alone. The licensure and approval of a rotavirus vaccine reduced hospitalizations by 70% to 80%.[20] Another example is the human papilloma virus (HPV) vaccine, a recombinant virus-like particle (VLP) vaccine for primary prevention of cancer. The US CDC Advisory Committee on Immunization Practices (ACIP) recommended routine HPV vaccination for young females in 2006 and for young males in 2011. Since the introduction of the HPV vaccine, there has been a significant reduction in HPV infections and cervical precancers. A comprehensive meta-analysis of more than 60 million HPV vaccinated individuals in 14 countries demonstrated that the rate of HPV 16 to 18 infections decreased by 83% among females 13 to 19 years of age and by 66% among those 20 to 24 years of age, whereas the prevalence of precancerous lesions decreased by 51% and 31%, respectively.[21]

Recent Changes in Vaccine Development Strategies

Early vaccines were live attenuated or inactivated versions of whole wild-type human pathogens, for example, rabies, yellow fever virus, and influenza. In a few cases, attenuated zoonotic organisms closely related to human pathogens were employed to produce cross-reactive protective responses in humans, for example, vaccinia, an animal poxvirus utilized as a vaccine against human smallpox, and bacille Calmette-Guerin (BCG), an agent of bovine tuberculosis developed as a human tuberculosis vaccine. Later, split virus vaccines utilized partially purified protein antigens derived from whole inactivated viruses, for example, split virus influenza vaccines. The polysaccharide capsules of bacteria were purified from cultures of multiple serotypes of a single bacterial species leading to polyvalent polysaccharide vaccines; for example, the 23-valent pneumococcal polysaccharide and the quadrivalent meningococcal polysaccharide vaccines. Bacterial toxins were purified from cultures and made harmless by heat or chemical treatment to produce toxoid vaccines, for example, tetanus and diphtheria vaccines.

Recent decades have featured explosive discoveries in genetics, molecular biology, immunology, and microbiology, leading to new theory-based (so-called rational) approaches to vaccine design. These advances have led to structure-based vaccine design, generations of recombinant vaccines (based on combining two or more sources of DNA), recombinant viral-vectored vaccines, and nucleic acid-based vaccines (Table 87.3).

Advances in the development of a vaccine for respiratory syncytial virus (RSV) exemplify the impact of structural biology and molecular engineering on vaccine design. RSV is the leading cause of viral acute lower respiratory tract infections globally, with the highest burden of disease occurring in infants under six months of age.[22] Despite nearly 60 years of research and development efforts, no licensed vaccine for RSV exists. In the 1960s, one clinical trial administering a formalin-inactivated RSV vaccine candidate (FI-RSV) in infants and young children resulted in the hospitalization of 80 percent of vaccine recipients, with two fatalities due to disease enhancement following natural RSV infection.[23] Twenty years after the trial, it was determined that while FI-RSV elicited antibodies in nearly all recipients, the majority were directed against nonprotective epitopes.[24] Structural biology became a prominent tool in demonstrating why FI-RSV preferentially produced non-neutralizing antibodies. The fusion (F) glycoprotein of RSV, required for viral entry into host cells, exists in two conformational states: pre-fusion (pre-F) and post-fusion (post-F). While the pre-F conformation is used for viral entry, it is metastable and irreversibly rearranges to a nonfunctional post-F state.[25] Due to the unstable structure of F, formalin

TABLE 87.3 Vaccine Platforms: Classical and Next-Generation

Platform Type	Subtype	Examples
Whole pathogen	Live attenuated	Measles, mumps, rubella, varicella zoster, yellow fever vaccines
	Inactivated	Rabies vaccine
Subunit	Polysaccharide	23-valent *Streptococcus pneumoniae* vaccine
	Polysaccharide conjugated to protein	13-valent *Streptococcus pneumoniae, Haemophilus influenzae, Neisseria meningitidis* vaccines
	Protein	Influenza vaccine
	Virus-like particle	Human papillomavirus vaccine
Next-Generation	Viral vectored	Dengue, Ebola vaccines
	Nucleic acid based	Zika (in development) and SARS-CoV-2 vaccines
	Nanoparticle based	Influenza (in development)

inactivation in FI-RSV resulted in an almost entirely post-F antigen. Recent advances in structural biology and molecular engineering were prominent features in the design of a productive RSV antigen through the introduction of stabilizing mutations to preserve the pre-F conformation. The resulting antigen, DS-Cav1, is a stabilized trimer of the pre-F RSV F glycoprotein. This protein subunit vaccine elicited neutralizing titers 70 to 80 times greater than post-F antigens in mice and nonhuman primates. Advancing to phase 1 trials, DS-Cav1 was safe and tolerable in healthy adults and elicited neutralizing antibody responses with and without adjuvant.[25] The identification and stabilization of the pre-F conformation have led to the development of several types of vaccine candidates designed specifically for infants, the elderly, and pregnant women in the third trimester intended to provide passive immunity to the infant through the first months of life.[26] At the time of this writing, further clinical development and vaccine efficacy trial outcomes are highly anticipated.

Concurrent with the advances in structural biology that enabled progress in RSV vaccine development, advances in genetics and molecular biology allowed for gene cloning and expression in recombinant molecular systems, revolutionizing vaccine development. Vaccines can now be designed based on the *in vitro* expression of one or a few genes. For example, the hepatitis B vaccine, originally developed by Hilleman, was purified hepatitis B surface antigen (HBsAg) from the blood of chronically infected humans. But soon thereafter, a second licensed hepatitis B vaccine was produced in yeast cells through recombinant DNA methods that inserted the HBsAg gene into yeast organisms for expression and purification. In 1986, this hepatitis B vaccine was the first approved recombinant vaccine in the US (RECOMBIVAX HB) (https://www.fda.gov/media/74274/download). Newer generation vaccines have since been developed (ENGERIX-B [https://www.fda.gov/media/119403/download] and HEPLISAV-B [https://www.fda.gov/media/108745/download]) and are widely used today. This platform offers advantages including protein purity, as the genes of interest are expressed in relative isolation, and vaccine safety, as it is no longer necessary to derive vaccines by partially purifying the HBsAg from paid-donor plasma of humans chronically infected with hepatitis B virus (and potentially other viruses).

Novel recombinant vaccines and recombinant viral-vectored vaccines have been approved or recommended for human use for a number of pathogens, including HPV, malaria, and dengue, as discussed below.

Human Papillomavirus

The HPV vaccine is a highly effective recombinant VLP vaccine; its public health impacts were highlighted earlier in this chapter. Recombinant HPV L proteins expressed in recombinant systems form VLPs that are purified and formulated with or without an adjuvant. The most recent polyvalent vaccine expresses VLPs representing nine HPV serotypes (GARDASIL 9).[27] HPV VLP vaccines are remarkable for their efficacy and safety and because they are primary prevention for several types of cancer in both boys and girls.[28]

Malaria

Malaria causes an annual global disease burden of 220 million cases and 400,000 deaths, with the vast majority of cases concentrated in Africa. Pregnant women and children under 5 are the two highest-risk populations, and the development of an effective malaria vaccine remains a global health priority (https://www.who.int/publications/i/item/world-malaria-report-2019). The most advanced malaria vaccine, RTS,S, is a recombinant protein subunit vaccine that targets the pre-erythrocytic stage of the *Plasmodium falciparum* parasite. It was evaluated in combination with the adjuvant AS01 in a Phase 3 trial in which RTS,S/AS01 (Mosquirix) demonstrated 36.3% vaccine efficacy four years after first vaccination in children aged 5 to 17 months who received the four recommended doses.[29] Following the phase 3 results, two WHO advisory groups jointly called for pilot implementation of the vaccine in 3 to 5 African nations. In April 2017, the WHO approved the joint recommendation and established the Malaria Vaccine Implementation Programme (MVIP) to further evaluate the vaccine's safety profile and assess the feasibility of a four-dose vaccine administration before broader use across sub-Saharan Africa (https://www.who.int/immunization/sage/meetings/2018/april/2_WHO_MalariaMVIP-update_SAGE_Apr2018.pdf?ua=1). Three pilot countries, Malawi, Ghana, and Kenya, were selected based on pre-specified criteria. In May 2018, the vaccine was approved by each country's national regulatory agency, and the first round of administration began in April 2019. In October 2021, after more than 2.3 million doses of the vaccine had been administered to over 800,000 children in the pilot nations, the WHO recommended RTS,S/AS01 for broad use in children in sub-Saharan Africa and other regions with moderate to high Plasmodium falciparum malaria transmission. The MVIP is anticipated to conclude in 2023 once the potential benefits of a 4th dose and longer-term effects on childhood deaths have been assessed (https://www.who.int/news/item/06-10-2021-who-recommends-groundbreaking-malaria-vaccine-for-children-at-risk).

Dengue

There are an estimated 390 million infections of the dengue virus globally each year, with 95 million of those infections resulting in clinical disease.[30] In 2019, the first dengue vaccine was approved in several countries, including by the US FDA, for use in dengue-endemic regions. This vaccine, Dengvaxia (CYD-TDV), is a recombinant live tetravalent viral vector based on the yellow fever virus vaccine strain 17D expressing the envelope and pre-membrane genes of all four dengue serotypes. Dengvaxia has been administered to more than 41,000 individuals across 26 clinical trials, with a favorable safety and immunogenicity profile.[31] Based on promising results, vaccination campaigns were launched in both Brazil and the Philippines, which included school-aged children. However, long-term observation of vaccine recipients revealed an increased risk of severe dengue disease in individuals who had no previous exposure to dengue at the time of vaccination (*i.e.*, baseline seronegative individuals) and in young children (regardless of serostatus).[31] During the vaccination campaigns in Brazil and the Philippines, 87 cases of dengue infection were reported, with 14 resulting in fatalities. Following an additional investigation by the WHO Global Advisory Committee on Vaccine Safety, no causality determination could be made for these fatalities.[32] Based on this increased risk of severe dengue infection in seronegative vaccine recipients, Dengvaxia is indicated only for seropositive individuals aged 9 to 45 years.[33] The results from the long-term follow-up of Dengvaxia had a detrimental impact on vaccine confidence, particularly in the Philippines. In that country, increased vaccine hesitancy is believed to have contributed to a widespread measles outbreak in 2019.[34]

Recombinant viral vectored vaccines, such as Dengvaxia, the Ebola vaccine Ervebo, which will be discussed later in this chapter, and

nucleic acid vaccines represent the next generation of vaccine technology. DNA vaccines utilize DNA plasmids as a vector for expressing pathogen antigens *in vivo*, while mRNA is packaged in a carrier molecule for cellular delivery, most often a lipid nanoparticle. The technology has been utilized for rapid vaccine production in response to outbreaks such as Zika and the COVID-19 global pandemic, which began in 2019.[35–37] In August 2021, COMIRNATY®, a COVID-19 mRNA vaccine developed by BioNTech and Pfizer, became the first nucleic acid-based vaccine to receive approval from the US FDA for use in humans (https://www.fda.gov/news-events/press-announcements/fda-approves-first-covid-19-vaccine). A second vaccine for COVID-19 using mRNA technology, SPIKEVAX® manufactured by Moderna, received FDA approval shortly after in January 2022 (https://www.fda.gov/emergency-preparedness-and-response/coronavirus-disease-2019-covid-19/spikevax-and-moderna-covid-19-vaccine). Fig. 87.1 portrays a schematic of both classical and next-generation platforms in the context of COVID-19 vaccine development, and additional discussion of the pandemic and the response is included later in this chapter.

Adjuvants

The magnitude of immune responses can be improved by adding compounds called *adjuvants* to vaccine formulations. Recent advances in our understanding of the innate immune system have led to an appreciation that adjuvants act primarily through their effects on innate immunity. Adjuvant-triggered innate signals enhance the quantity, quality, and specificity of the downstream adaptive immune responses to the vaccine antigen. Adjuvants are also used to promote increased rates of seroconversion and induce immunity even in populations with less responsive immune systems such as the elderly, infants, and immunocompromised. Another advantage of adjuvants relates to dose-sparing, or their ability to reduce the amount of antigen used or the number of vaccine administrations given to produce comparable immune responses.[38] Some of the most widely used, clinically approved adjuvants are aluminum-based, such as aluminum hydroxide (AH) and aluminum phosphate (AP). These adjuvants primarily function to amplify antibody production in response to vaccine antigens. Although aluminum adjuvants have been used for many decades, the exact mechanism underlying their immune enhancement properties is not fully understood. Aluminum adjuvants generally have safe profiles and are included in vaccine formulations at very low doses (0.85 to 1.25 mg).[39,40] Novel adjuvants in various stages of development include oil-in-water emulsions (*e.g.*, MF059 and AS03), saponin-based adjuvants (*e.g.*, QS-21), adjuvants targeting pattern recognition (*e.g.*, CpG-ODN), and Toll-like receptor and RIG-I-like receptor ligand-specific adjuvants (*e.g.*, TLR4).[38]

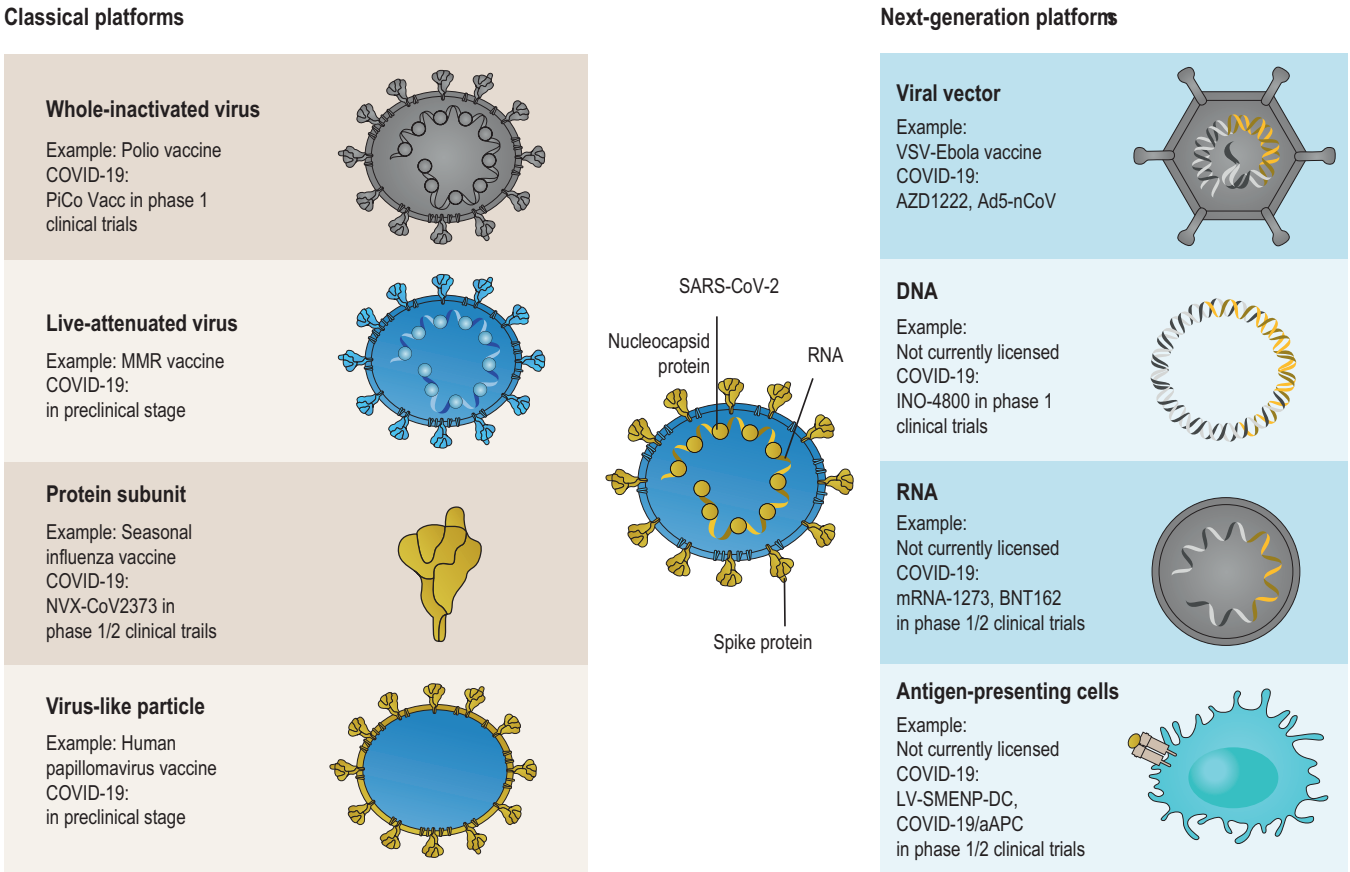


FIG. 87.1 An Overview of the Different Vaccine Platforms in Development Against COVID-19. A schematic representation is shown of the classical vaccine platforms that are commonly used for human vaccines, and next-generation platforms, where very few have been licensed for use in humans. The stage of development for each of these vaccine platforms for COVID-19 vaccine development is shown; online vaccine trackers are available to follow these vaccines through the clinical development and licensing process. (Reproduced from van Riel D, de Wit E. Next-generation vaccine platforms for COVID-19. *Nat Mater*. 2020;19[8]:810–812.) As of March 2022, two RNA vaccines are currently approved by the FDA for use in humans: COMIRNATY® (NCT04368728) and SPIKEVAX® (NCT04470427).

Systems Biology Approach to Vaccination

Over the last decade, systems biology, or systems vaccinology, approaches to vaccine development have captured considerable interest.[41] The principal objectives of a systems vaccinology approach are to elucidate complex immunological pathways that generate long-lasting immunological memory and to provide new insights into molecular and cellular signatures that can predict vaccine efficacy. In addition to the traditional cellular and humoral immunological assays (antibodies, T cells, B cells), multiple "omics" assays may be performed and used to generate computer models or algorithms describing the immune response to vaccination. Some of the applications of "omics" assays include profiling of T-cell epitopes and antibody specificity by proteomics, the discovery of predictive biomarkers by metabolomics and lipidomics, assessment of host-pathogen interactions and infection-induced immune responses by transcriptomics, and mapping of antibody glycosylation by glycomics.[42] Recent technological advances bring novel methods to systems vaccinology that include single-cell genomics and epigenomics. Single-cell technologies enable deconvolution and more in-depth resolution of immune responses by identifying cellular heterogeneity, rare cell subtypes, and unique biomarkers.[43] The use of new high-throughput assays to assess multiple dimensions of innate and adaptive immune responses generates very large data sets. Analyses and integration of these huge data sets require multidisciplinary collaboration with computational biologists and informaticians. These detailed assessments are being applied in a variety of infectious and non-infectious disease states. The impact of systems vaccinology on public health is yet to be fully realized; as the technologies supporting this approach continue to improve, new progress on vaccine development and utilization is anticipated.

Recent Responses to Epidemics and Pandemics

Despite all the advances and accomplishments of vaccine science, there remains a pressing public health concern that resonates around the globe, that is: when major epidemics of lethal and highly infectious diseases occur, can protective vaccines be developed quickly enough to respond? Advances in next-generation vaccine technology have allowed for record-speed product development over the last several years, most recently with the development of multiple COVID-19 vaccine candidates in a matter of months. Demonstrating vaccine efficacy during an ongoing epidemic or pandemic remains a challenge, and each outbreak presents unique hurdles. Below are three recent case studies of diseases that caused epidemics or pandemics and examples of vaccine development that occurred in response to these global events.

Ebola

Ebola was first discovered in 1976, and vaccine development began in the late 1990s with an initial phase 1 clinical trial of the first candidate vaccine in 2003.[44] Multiple iterations of the Ebola vaccine were developed and tested in Phase 1 clinical trials leading to refinement of the antigen design and platform approach, and testing of advanced candidates starting in 2014.[45,46] One of these candidates, rVSV-ZEBOV (Ervebo), was approved by the US FDA in 2019 for the prevention of Ebola virus disease (EVD) after demonstration of efficacy in a ring-vaccination clinical trial in 2015–16 during the West Africa Ebola outbreak (https://www.fda.gov/vaccines-blood-biologics/ervebo).[47] The VSV-EBOV vaccine is unique in that it represents the first vaccine against a filovirus to be approved in the United States and is from a novel class of vaccine based on a viral vector. VSV-EBOV is a live, attenuated recombinant vesicular stomatitis virus (VSV) in which the gene for the native envelope glycoprotein is replaced with the gene from the Ebola virus glycoprotein.[46] Additional candidate vaccines designed to prevent EVD and other filoviruses are under evaluation in clinical testing (NCT04041570, NCT03475056),[48] and a prime-boost vaccine regimen, Zabdeno® (Ad26.ZEBOV) and Mvabea® (MVA-BN-Filo), was granted Marketing Authorization from the European Medicines Agency for prophylactic use in individuals ages 1 and older in May 2020 (https://www.who.int/news-room/questions-and-answers/item/ebola-vaccines).

Zika Virus

Zika virus is a mosquito-borne flavivirus closely related to dengue that was first discovered in 1947 in Zika Forest, Uganda. This single-stranded positive sense RNA virus resulted in small human disease outbreaks over the years, but from 2015 to 2016 emerged and spread across the Americas, Africa, and other parts of the world (https://www.who.int/news-room/fact-sheets/detail/zika-virus). To date, a total of 86 countries and territories have reported evidence of mosquito-transmitted Zika infection.[49] In pregnant women, infections resulted in fetal microcephaly or other birth anomalies.[50] In general, healthy adults with symptomatic infections experience a mild to moderate self-limited viral illness which has been described as "mild dengue" and is mostly characterized by fever, rash, conjunctivitis, and arthritis. An increased association of Guillain-Barré syndrome with Zika infections has also been reported in multiple countries. Interestingly, it is believed that 80% of Zika infections are asymptomatic. In symptomatic infected adults, viremia persists for less than a week in most cases, but longer durations of viral RNA detection are reported in semen and urine.[51,52]

The development of a vaccine emerged as a top priority of the US government's response to the epidemic in 2015. Several leading candidates, including both inactivated and DNA vaccine platforms, were rapidly developed and evaluated in early phase clinical trials.[36,53] One of the candidates progressed into a multinational efficacy trial in early 2017, but the epidemic waned before an efficacy signal could be detected.[54] However, these vaccine candidates, along with others that have shown promise in preclinical studies, remain in development in preparation for another Zika epidemic.

SARS-CoV-2

In January 2020, a novel coronavirus was identified as the cause of an outbreak in China. By late September 2020, SARS-CoV-2 quickly spread worldwide with over 1 million documented deaths due to the clinical disease, COVID-19.[55] Using techniques and expertise garnered from prior pandemic responses and pre-existing coronavirus (SARS-CoV-1 and MERS) vaccine research, publicly and privately funded vaccine research teams promptly developed candidate SARS-CoV-2 vaccines for the prevention of COVID-19 disease.[56,57] The first documented COVID-19 vaccine clinical trial launched in the United States in March 2020 with multiple candidates entering clinical trials shortly after, quickly demonstrating safety and immunogenicity[37,58–60] resulting in the launch of multiple phase 3 efficacy trials by mid to late 2020 (NCT04505722, NCT04516746, NCT04470427, NCT04368728).[61] A variety of established and novel vaccine

platforms were developed predominantly expressing the SARS-CoV-2 spike protein with many specifically encoding for a stabilized version of the spike protein described in early 2020.[56] To enable rapid deployment of safe and efficacious vaccines, multiple international governments established vaccine research and production programs and the US government launched Operation Warp Speed (OWS) in May 2020, designed to utilize expertise and resources from the US government and private sectors working rapidly to develop and produce a vaccine for the US public, specifically to produce over 300 million safe and effective vaccine regimens for the US public by January of 2021. This effort funded and enabled the development of multiple candidate vaccines of various platform types (including nucleic acid, viral vector, and protein subunit).[61] At the time of this writing, two nucleic acid vaccines have received FDA approval, COMIRNATY® and SPIKEVAX®, after demonstrating 93 to 95% efficacy against symptomatic disease in final analyses of the phase 3 trials (https://www.fda.gov/emergency-preparedness-andresponse/coronavirus-disease-2019-covid-19/covid-19-vaccines).

Current Recommendations

Today in the United States, there are clear national guidelines that recommend vaccines for children, adolescents, and adults. Each February, the CDC publishes two recommended immunization schedules based on the recommendations of the CDC-appointed ACIP. One ACIP schedule of immunizations provides the adult immunization recommendations (Table 87.4). The adult schedule offers recommendations for each vaccine based on the age of the patient. For example, the ACIP recommends that all adults (persons aged 19 years and over) receive: annual influenza vaccination; tetanus-diphtheria-acellular pertussis (Tdap) vaccine once, followed by tetanus boosters every 10 years; a 2-dose recombinant zoster vaccine (RZV) administered 2 to 6 months apart for individuals aged ≥50 years; and a pneumococcal vaccination at age 65 years (https://www.cdc.gov/vaccines/schedules/hcp/index.html).

The adult schedule also provides recommendations for vaccines indicated for certain risk factors, including medical conditions (*e.g.*, immunocompromising conditions, kidney failure, diabetes), pregnancy, and certain occupations. Importantly, live

TABLE 87.4 Parts A and B: Recommended Adult Immunization Table, United States, 2020, From the Advisory Committee on Immunization Practice of the Centers for Disease Control and Prevention

A

Vaccine	19–26 years	27–49 years	50–64 years	≥65 years
Influenza inactivated (IIV4) or **Influenza recombinant** (RIV4)	1 dose annually			
or		or		
Influenza live, attenuated (LAIV4)	1 dose annually			
Tetanus, diphtheria, pertussis (Tdap or Td)	1 dose Tdap each pregnancy; 1 dose Td/Tdap for wound management (see notes)			
	1 dose Tdap, then Td or Tdap booster every 10 years			
Measles, mumps, rubella (MMR)	1 or 2 doses depending on indication (if born in 1957 or later)			
Varicella (VAR)	2 doses (if born in 1980 or later)		2 doses	
Zoster recombinant (RZV)	2 doses for immunocompromising conditions (see notes)		2 doses	
Human papillomavirus (HPV)	2 or 3 doses depending on age at initial vaccination or condition	27 through 45 years		
Pneumococcal (PCV15, PCV20, PPSV23))	1 dose PCV15 followed by PPSV23 or 1 dose PCV20 (see notes)			1 dose PCV15 followed by PPSV23 or 1 dose PCV20
Hepatitis A (HepA)	2 or 3 doses depending on vaccine			
Hepatitis B (HepB)	2, 3, or 4 doses depending on vaccine or condition			
Meningococcal A, C, W,Y (MenACWY)	1 or 2 doses depending on indication, see notes for booster recommendations			
Meningococcal B (MenB)	2 or 3 doses depending on vaccine and indication, see notes for booster recommendations			
	19 through 23 years			
***Haemophilus influenzae* type b** (Hib)	1 or 3 doses depending on indication			

Recommended vaccination for adults who meet age requirement, lack documentation of vaccination, or lack evidence of past infection

Recommended vaccination for adults with an additional risk factor or another indication

Recommended vaccination based on shared clinical decision-making

No recommendation/ Not applicable

TABLE 87.4 Parts A and B: Recommended Adult Immunization Table, United States, 2020, From the Advisory Committee on Immunization Practice of the Centers for Disease Control and Prevention—cont'd

B

Vaccine	Pregnancy	Immuno-compromised (excluding HIV infection)	HIV infection CD4 percentage and count: <15% or <200 mm³	HIV infection CD4 percentage and count: ≥15 and ≥200 mm³	Asplenia, complement deficiencies	End-stage renal disease, or on hemodialysis	Heart or lung disease; alcoholism[1]	Chronic liver disease	Diabetes	Health care personnel[2]	Men who have sex with men
IIV4 or RIV4	1 dose annually										
or										or	
LAIV4	Contraindicated					Precaution				1 dose annually	
Tdap or Td	1 dose Tdap each pregnancy	1 dose Tdap, then Td or Tdap booster every 10 years									
MMR	Contraindicated*	Contraindicated	1 or 2 doses depending on indication								
VAR	Contraindicated*	Contraindicated		2 doses							
RZV		2 doses at age ≥19 years		2 doses at age ≥50 year							
HPV	Not Recommended*	3 doses through age 26 years	2 or 3 doses through age 26 years depending on age at initial vaccination or condition								
Pneumococcal (PCV15, PCV20, PPSV23)		1 doses PCV15 followed by PPSV23 OR 1 dose PCV20(see notes)									
HepA				2 or 3 doses depending on vaccine							
HepB	3 doses (see notes)	2, 3, or 4 doses depending on vaccine or condition									
MenACWY	1 or 2 doses depending on indication, see notes for booster recommendations										
MenB	Precaution	2 or 3 doses depending on vaccine and indication, see notes for booster recommendations									
Hib		doses HSCT[3] recipients only		1 dose							

Recommended vaccination for adults who meet age requirement, lack documentation of vaccination, or lack evidence of past infection

Recommended vaccination for adults with an additional risk factor or another indication

Recommended vaccination based on shared clinical decision-makin

Precaution—vaccination might be indicated if benefit of protection outweighs risk of adverse reaction

Contraindicated or not recommended—vaccine shoulde not be administered.

*Vaccinate after pregnancy.

No recommendation/ Not applicable

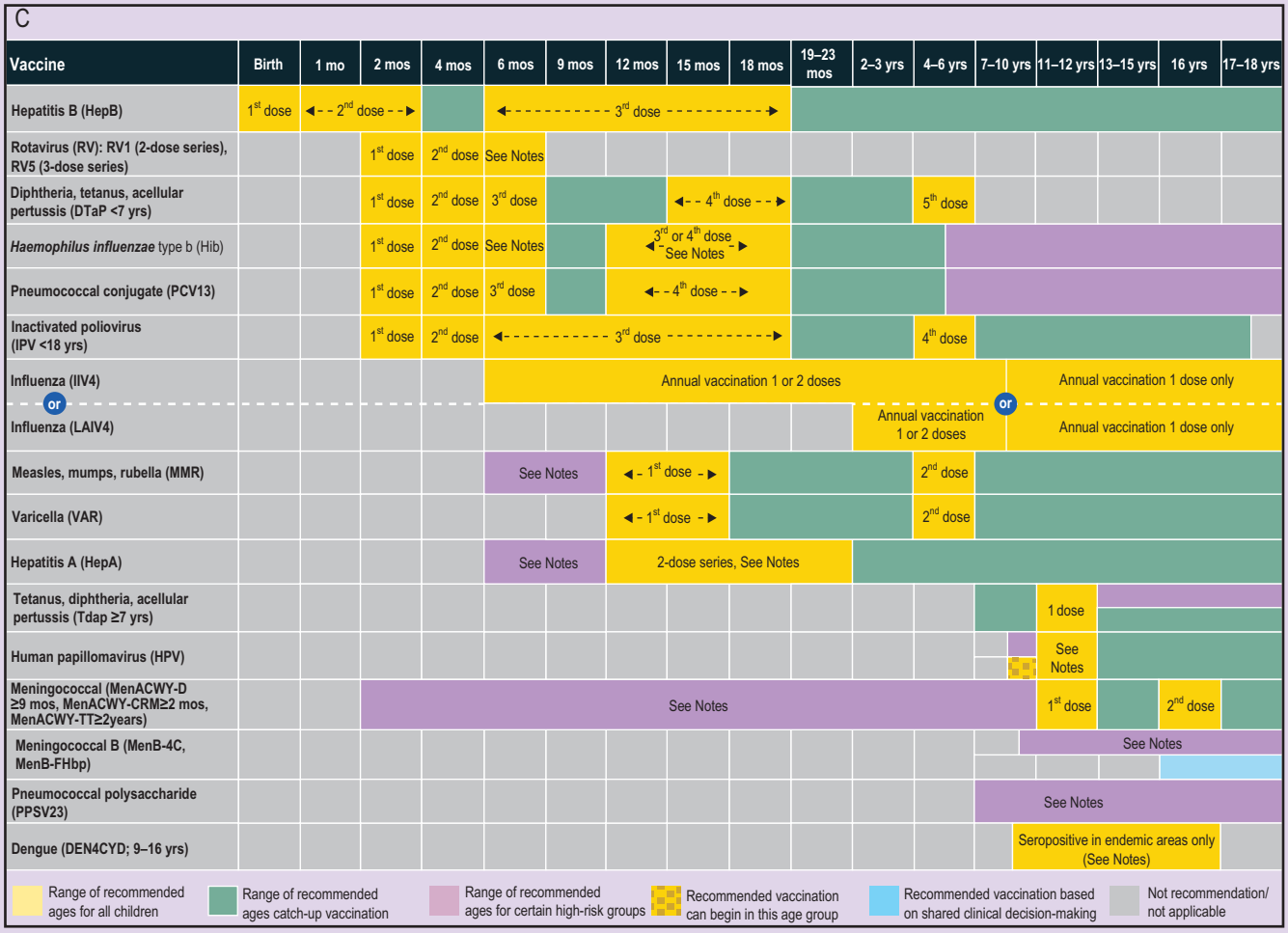
1. Precaution for LAIV does not apply to alcoholism. 2. See notes for influenza; hepatitis B; measles, mumps, and rubella; and varicella vaccinations.
3. Hematopoietic stem cell transplant.

Notes: Tetanus, diphtheria, and pertussis vaccination: Pregnancy: 1 dose Tdap during each pregnancy, preferable in early part of gestational weeks 27-36; wound management: Persons with 3 or more doses of tetanus-toxoid-containing vaccine: For clean and minor wounds, administer Tdap or Td if more than 10 years since last dose of tetanus-toxoid-containing vaccine; for all other wounds, administer Tdap or Td if more than 5 years since last dose of tetanus-toxoid-containing vaccine. Tdap is preferred for persons who have not previously received Tdap or whose Tdap history is unknown. If a tetanus-toxoid-containing vaccine is indicated for a pregnant woman, use Tdap. **Zoster vaccination:** Immunocompromising conditions (including HIV): RZV recommended for use for persons 19 years or older who are or will be immunodeficient of immunosuppressed because of disease or therapy. **Pneumococcal vaccination:** Age 19–64 years with certain underlying medical conditions or other risk factors who have not previously received a pneumococcal conjugate vaccine or whose previous vaccination history is unknown: 1 dose PCV15 or 1 dose of PCV20. If PCV15 is used, this should be followed by a dose of PPSV23 given at least 1 year after PCV15 dose. A minimum interval of 8 weeks between PCV15 and PPSV23 can be considered for adults with an immunocompromised condition, cochlear implant, or cerebrospinal fluid leak to minimize the risk of invasive pneumococcal disease caused by serotypes unique to PPSV23 in these vulnerable groups. **Hepatitis B vaccination:** Heplisay-B is not recommended in pregnancy due to lack of safety data in pregnant women. **Meningococcal vaccination:** Booster dose is recommended for those at increased risk due to an outbreak and if 5 or more years have passed since receiving MenACWY and 1 year or more since receiving MenB. Detailed information could be found at https://www.cdc.gov/vaccines/schedules/downloads/adult/adult-combined-schedule.pdf.

vaccines (varicella and MMR) are contraindicated for pregnant women, immunocompromised hosts, and HIV-infected individuals when the CD4+ T-cell absolute count is below 200 cells/mcl.

A second ACIP immunization schedule of immunizations covers birth to 18 years of age and catch-up recommendations for children or adolescents who have not received recommended vaccines (Table 87.5). The ministries of health for many European countries publish their own country-specific immunization schedules, and vaccination guidelines published by the WHO are utilized by many developing countries. The schedules are generally similar but with some region-specific differences. For example, the 2020 US ACIP immunization schedule for children recommends vaccinations against ten viral diseases: hepatitis B, rotavirus, poliovirus, influenza, measles, mumps, rubella, varicella, hepatitis A, and human papilloma virus (HPV) (https://www.cdc.gov/vaccines/schedules/hcp/index.html). Preventive viral vaccines in the WHO-recommended routine immunization schedule for children include the same 10 viral vaccines (although four—mumps, influenza, varicella, and hepatitis A vaccines—are recommended only for country immunization programs with certain characteristics). The WHO schedule also recommends additional vaccines, for example, rabies, yellow fever, Japanese encephalitis, and tick-borne encephalitis vaccines for certain high-risk populations (https://www.who.int/immunization/policy/immunization_tables/en/).

TABLE 87.5 Recommended Immunization Schedules for Persons Aged 0 Through 18 Years, United States, 2020, From the Advisory Committee on Immunization Practice of the Centers for Disease Control and Prevention



Notes: Measles, mumps, rubella (MMR): during international travel infants age 6–11 months 1 dose before departure; revaccinate with 2-dose series at age 12–15 months (12 months for children in high-risk areas) and dose 2 as early as 4 weeks later. **Hepatitis A:** during international travel infants age 6–11 months 1 dose before departure; revaccinate with 2 doses, separated by at least 6 months, between age 12–23 months. For detailed information on **Meningococcal Vaccination:** MenACWY-D, MenACWY-TT, MenB-4C, MenB-FHbp) and **Pneumococcal vaccination** see https://www.cdc.gov/vaccines/schedules/hcp/imz/child-adolescent.html#note-mening.

KEY CONCEPTS

Current Areas of Vaccine Need

- Human immunodeficiency virus (HIV)
- Lyme disease
- Malaria
- Powassan disease
- Rocky Mountain spotted fever
- Universal coronavirus (SARS-CoV-1, MERS, SARS-CoV-2)
- Tuberculosis (TB)
- Tularemia
- Zika

Present and Future Challenges

A few specific challenges facing those involved in vaccine research and discovery are highlighted below to illustrate the ongoing needs in the area of public health.

A Vaccine for HIV. The development of an HIV/AIDS vaccine has long been recognized as a top HIV research global priority at the US National Institutes of Health. Strong and simple treatments for those who are living with HIV infection are now available and have even been rolled out to developing countries. It has been shown that treatment-as-prevention, wherein the viral load is lowered to an undetectable level by antiretroviral treatment of infected persons, results in a benefit to the infected patient and up to 96% reduction in HIV incidence in sexual partners.[62]

More recently, antiretroviral agents have been tested globally and licensed in the United States as a once-daily pill (a combination of tenofovir and emtricitabine) for HIV/AIDS prevention in higher-risk individuals.[63] Known as pre-exposure prophylaxis (PrEP), this approach, in an idealized setting where resources and human adherence were not limiting, could have a truly dramatic impact on HIV incidence. However, to date, uptake has been low, and adherence has been a concern. A federal initiative, Ending the HIV Epidemic: A Plan for America,

which includes a plan to make PrEP medication available without cost for up to 200,000 people a year for 11 years, was announced in 2019 to help combat these issues and reduce the number of new HIV infections in the United States.[64]

While the advances in HIV treatment, treatment-as-prevention, and PrEP have been significant, the numbers of new infections globally remain unacceptably high, with 1.7 million new infections in 2019 and a total of 38 million people living with HIV infection.[65] In the United States, progress on prevention of HIV infections through condoms, education, and evidence-based interventions plateaued new infections to ~36,400 in 2018 (http://www.cdc.gov/hiv/library/reports/hiv-surveillance.html).

The global need for an HIV vaccine remains. However, the scientific challenges have proven significant, and despite more than 30 years of major effort since the identification of the viral etiologic agent of AIDS in 1984, there is no approved HIV vaccine with proven protective efficacy. The majority of efficacy trials completed to date have not achieved protection of higher-risk vaccinated subjects relative to placebo recipients.[66] Additionally, in two of the trials (which tested replication-deficient adenovirus serotype 5 recombinant vaccine vectors expressing HIV Gag, Pol, and Nef but not Env), the studies were halted early with either concern over possible enhanced HIV acquisition in a small subset of participants or lack of prevention efficacy in the vaccine groups relative to the placebo groups.[66]

Importantly, modest vaccine efficacy was observed in the RV144 efficacy trial reported in 2009.[67] Conducted by the US Army in collaboration with the government of Thailand. This 16,000-person study evaluated a prime-boost regimen of a non-replicating canarypox vector prime (expressing HIV Gag, protease, and gp120) followed by boosting with the same vector, plus a bivalent gp120 protein adjuvanted in alum. The RV144 regimen produced 61% protection in the first year post-vaccination and modest (31.2%) protection at 3.5 years post-vaccination.[67]

This first evidence of human protection by an HIV vaccine proved that the development of an HIV vaccine may be possible and has re-energized the field. The US HIV Vaccine Trials Unit Network (HVTN), National Institute of Allergy and Infectious Diseases (NIAID), National Institutes of Health (NIH), US Army, and country-level and industry collaborators have formed the Pox-Protein Public-Private partnership (also known as P5) to plan an intensive series of follow-up human studies to confirm and fully investigate the important leads provided to the field by RV144.[68] One such follow-up efficacy trial of 5407 participants in South Africa (HVTN 702) was unfortunately not as successful as RV144. This trial was investigating a poxvirus vector and a bivalent gp120 protein adjuvanted with MF59, both modified to express clade C. The HVTN 702 trial was recently halted when the data and safety monitoring board (DSMB) found that the regimen did not prevent infection compared to placebo recipients.[69] Additional analysis is ongoing to discover the reason behind these discrepant outcomes.

A holy grail for HIV vaccines remains the discovery of a vaccine immunogen that induces broadly neutralizing, protective antibodies. Such antibodies occur naturally in up to 15% of chronically infected persons, usually after years of infection. Although seen in natural infection, no vaccine has been able to readily induce these broadly neutralizing antibodies in vaccinated humans. However, several broadly neutralizing monoclonal antibodies (bnAbs) have been isolated and cloned, and several have been tested for safety and pharmacokinetics in human trials. To date, these bnAbs have proven safe and well-tolerated in both healthy and HIV-infected recipients.[70–72] In viremic recipients who do not have resistant viruses present prior to infusion, receipt of either a single or combination of bnAbs results in a temporary decrease in circulating viral load, which typically rebounds once the serum levels of the bnAb lower below a protective concentration.[73–75]

NIAID, through two of its HIV/AIDS clinical trials networks (the HVTN and the HIV Prevention Trials Network (HPTN)), is conducting two phase 2B efficacy trials of a broadly neutralizing monoclonal antibody, VRC01 (Clinical trials HVTN 703/HPTN 081 and HVTN 704/HPTN 085).[70,75] The VRC01 efficacy trials (or AMP studies for antibody-mediated protection) are randomized, double-blind, placebo-controlled clinical trials. In the AMP studies, VRC01 was infused IV every eight weeks for 18 months at doses of 0 mg/kg (placebo), 10 mg/kg, or 30 mg/kg. While VRC01 was generally well-tolerated and demonstrated a favorable safety profile, it did not prevent acquisition of resistant viral strains. It did, however, protect against sensitive isolates, providing 75% protection over the 20- month trial to at-risk populations exposed to sensative subtype B and C variants.

These results support what many experts suspected: rather than a single antibody, a combination of potent monoclonal antibodies targeting different epitopes on the gp120 envelope protein structure may be required to produce broad protection across a range of diverse subtypes. Combinations of two or three bnAbs are being evaluated in early-phase trials (NCT04173819, NCT04212091, NCT03928821).[75a]

Improved Influenza Vaccines

The disease burden due to seasonal influenza A is significant, with the highest morbidity and mortality occurring in children, older adults, pregnant women, and persons with chronic medical conditions.[76] During an average year, seasonal infections result in an estimated 3 to 5 million severe cases and 291,000 to 645,000 influenza-associated deaths worldwide.[77] In the United States, it is currently recommended that all persons 6 months or older receive an annual influenza vaccine.[76] This recommendation serves to protect the vaccinated individual as well as those in the community who cannot be vaccinated themselves.

Influenza A and influenza B viruses are responsible for the majority of human infections. Multiple subtypes of influenza A are categorized based on the amino acid (AA) sequence homology within the viral surface proteins, hemagglutinin (HA), and neuraminidase (NA). So far, 18 HA and 11 NA subtypes have been discovered. Currently, two influenza A subtypes (H1N1 and H3N2) and two antigenically distinct lineages of influenza B (Yamagata and Victoria) co-circulate in humans,[76] with one strain of each represented in the quadrivalent seasonal vaccine developed each year.

Influenza is a segmented negative-stranded RNA virus of the Orthomyxoviridae family that lacks a proof-reading function in its viral polymerase and therefore mutates rapidly. These mutations result in an antigenic drift of the surface proteins, requiring an annual vaccine reformulation. Currently, licensed vaccines are produced in either embryonated eggs or cell culture and include inactivated influenza vaccines (IIVs), recombinant influenza vaccines, and live attenuated vaccines.[76] WHO issues a new vaccine strain recommendation for the vaccine each February, and vaccine manufacturers then race to produce the year's seasonal vaccine by late summer in order to be ready for the winter influenza season. There are multiple challenges and

needs with regard to this annual process of influenza vaccine prediction, production, distribution, implementation, uptake, and protection (Fig. 87.2).

Young children (particularly those between 6 months and 5 years of age) and older adults have a higher risk of severe illness during influenza infection. Children between 6 months and eight years of age should receive two doses of vaccine administered at least four weeks apart during the first season they receive vaccination for optimal protection.[76] Quadrivalent inactivated influenza vaccines (IIVs) are approved for all ages, while live attenuated influenza vaccines (LAIVs) should only be administered to children over 2 years of age, and quadrivalent recombinant influenza vaccines (RIVs) to children over the age of 4 years. For older adults, currently licensed vaccines provide relatively weak protection overall but remain an important public health measure. Immunosenescence is a large contributor to this reduced vaccine efficacy, resulting in increased disease susceptibility and severity. One solution to this challenge has been the approval of a high-dose vaccine, containing a fourfold higher dose of antigen, for a total of 60 mcg (compared to the standard 15 mcg) of each viral HA protein. This high-dose vaccine was shown to increase efficacy and was approved for use in older adults in the United States in 2009 (trivalent) and 2019 (quadrivalent).[2,78] A vaccine formulated with the oil-in-water adjuvant, MF59, was approved for use in the United States and may increase immunogenicity in older adults (https://www.fda.gov/vaccines-blood-biologics/approved-products/fluad).[3]

An ongoing issue with seasonal influenza vaccines involves the variable vaccine effectiveness for each viral antigen within the vaccine, which is partially dependent on the degree of match between the vaccine strains and the circulating strains. In order to allow manufacturers the six months currently required for egg-based vaccine production, the vaccine strains for each subtype and lineage must be selected in February of each year for the following season's vaccine campaign. The burden of annual revaccination of the entire population against a variable viral target is high, both logistically and financially. Furthermore, uptake of the annual seasonal influenza vaccine in the general population remains suboptimal. Over the past decade, considerable effort has been put into the development of universal influenza vaccines that would produce a broad immune response capable of protecting an individual against more antigenically drifted viruses and should ideally mean protection for more than one influenza season. A common approach to the rational design of such a vaccine involves selecting antigens in the more conserved regions of the virus, including the highly conserved HA stalk rather than the hypervariable HA head and other conserved internal proteins. Some of the universal influenza vaccine candidates are moving into early phase human safety and immunogenicity clinical trials, with the ultimate goal to improve or possibly supplant the current annual vaccination

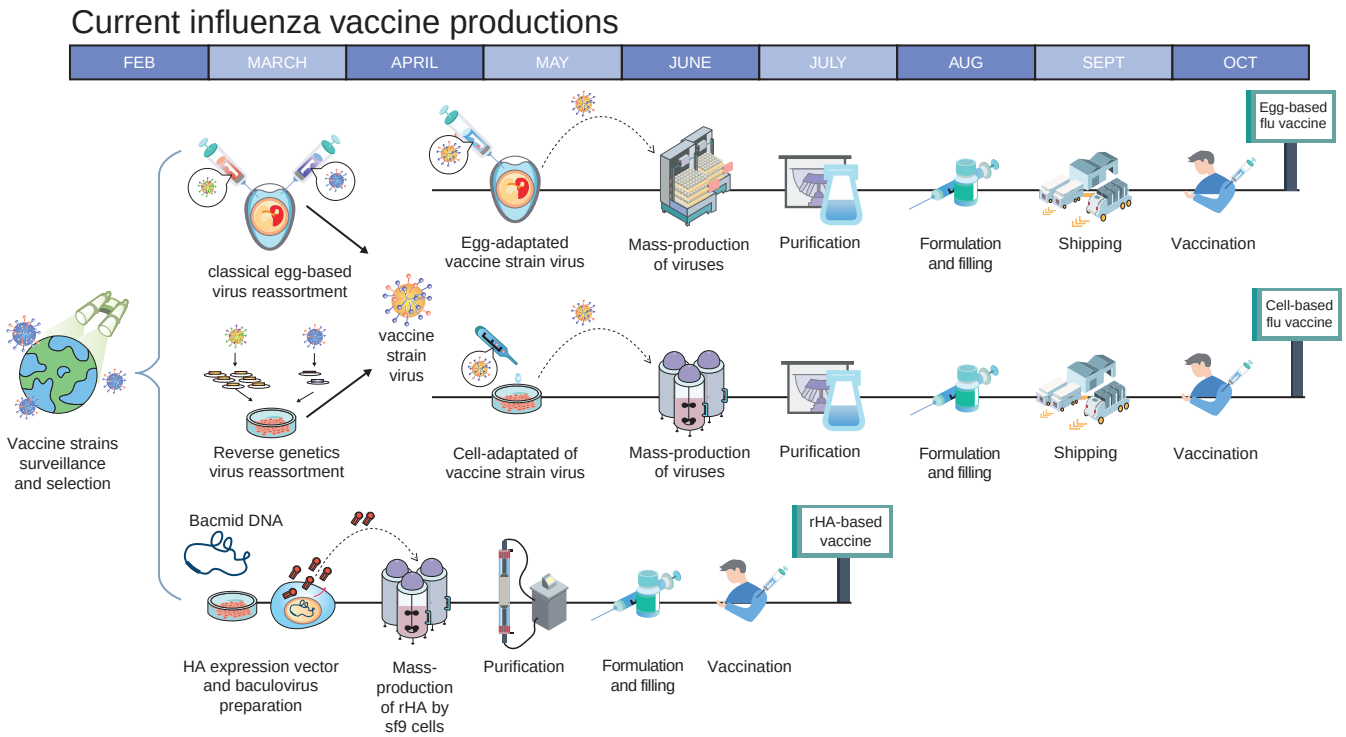


FIG. 87.2 Current Influenza Vaccine Productions. Timeline of current influenza vaccine production methods. Schematic overview of egg-based, cell-based and recombinant protein-based influenza vaccine production. Vaccine strains that match circulating influenza viruses for the upcoming flu season are selected by the World Health Organization (WHO) Global Influenza Surveillance and Response System (GISRS). High yielding vaccine strains for egg- or cell-based production are generated by either classic or reverse genetic reassortment. These adapted viruses go into mass production, either in embryonated eggs or MDCK cells with a production timeline of approximately 6 to 8 months. In recombinant hemagglutinin (HA) vaccines (rHA), the *HA* sequence is cloned into baculovirus and expressed by insect cells, significantly shortening production time. (Reproduced from Chen JR, Liu YM, Tseng YC, Ma C. Better influenza vaccines: an industry perspective. *J Biomed Sci.* 2020;27[1]:33.)

with an initial vaccination and periodic boosting. A further advantage to universal vaccines would be pandemic preparedness since the universal influenza vaccine, by targeting conserved viral domains present on all influenza viruses, may also protect against novel emerging subtypes.[79]

Selected Unmet Vaccine Needs

Along with HIV/AIDS and malaria, tuberculosis is one of the three top infectious disease killers each year, and therefore remains a major target for the development of a vaccine to reduce mortality and morbidity.[80–82] Currently, the WHO's list of diseases which "pose the greatest public health risk due to their epidemic potential and/or [the existence of] no or insufficient countermeasures" includes COVID-19, Crimean-Congo hemorrhagic fever, Ebola virus diseases, and Marburg virus disease, Lassa fever, Middle East respiratory coronavirus (MERS-CoV) and severe acute respiratory syndrome (SARS), Nipah and henipaviral diseases, Rift Valley fever, Zika, and a yet to be discovered "Disease X" (https://www.who.int/activities/prioritizing-diseases-for-research-and-development-in-emergency-contexts). There are also a number of tick-borne diseases with potential to cause significant disease burden in the United States for which there are no human vaccines including Lyme and Powassan diseases, Tularemia and Rocky Mountain spotted fever (https://www.cdc.gov/ticks/diseases/index.html).

Reflections on the Future of Vaccinology

The science of vaccinology has entered a new era with the high-throughput analytic and data approaches being applied today. Whether these advances in technology, computation, and science will result in improvements in human health remains to be seen. The potential of the big data systems biology approach has yet to be realized within vaccine science. The field must show that the massive data can be analyzed, integrated, and can produce new knowledge leading to improved vaccines that benefit human health.

In order for outstanding scientists to be retained and attracted to the field of vaccinology, it is essential that science funding agencies provide and maintain robust support for the critical discovery research projects (investigator-initiated research projects) that form the engine of innovation that drives all of science. At the same time, targeted big science vaccine program projects and networks have the potential to synergistically pool approaches in intensely focused efforts, tackling major vaccine needs and challenges, for example, HIV and TB vaccine development. In order to translate basic advances and laboratory science into improved health for patients, having well-trained translational physician-scientists leading clinical-translational human research programs is an additional component of essential infrastructure. Post-doctoral training programs that focus on vaccinology are needed, but few are in existence. The future of the field depends on attracting and training a highly qualified next generation of vaccinology research leaders.

Patients' and parents' confidence in and uptake of vaccines is derived primarily from the advice and education provided by their trusted physicians and other clinicians. But in the time-pressured office or hospital encounter, clinicians often find it challenging to prioritize discussions around vaccination. The clinician-patient interface is the most potent opportunity for education and influence, and this crucial moment must be seized in order for vaccine programs to be successful. Every clinician encounter is an opportunity to review vaccination history, recommend and discuss needed vaccines, and bring patients' vaccinations up to date. These seemingly simple, everyday actions hold incredible power to use vaccination to prevent disease and protect the health of all members of our global community.

ON THE HORIZON

- As valuable as vaccines have been in improving health in the past century, they are likely to be even more valuable in the next century.
- Emerging infections capable of severe morbidity (*e.g.*, Zika and SARS-CoV-2 viruses) or mortality (*e.g.*, Ebola virus and avian influenza A/H7N9) continue to emerge and threaten global human health with alarming regularity.
- Having vaccine platforms "on the shelf"—to enable rapid production and facilitate immediate preclinical testing for safety, immunogenicity and efficacy will be critical for timely and effective responses to global outbreaks that result from international personal contacts and air travel.
- Vaccines moving forward will depend on newer, improved methods, including cell culture-derived recombinant antigens and nucleic acid vaccines as well as automated high-throughput neutralization assays.
- There will be an increased application of new molecular adjuvants (*e.g.*, Toll-like receptor-7 [TLR-7]) that target specific biochemical pathways to direct and shape the cascade from innate immune response to desired adaptive immunity protective response.
- Nanoparticle delivery approaches will be increasingly used to prevent viral infections by targeting molecular structures of viral antigens.

ACKNOWLEDGMENTS

The authors thank Julie E. Ledgerwood, DO; Robert A. Seder, MD; Mark J. Mulligan, MD; and Larisa Strom, MPH for their editorial contributions to this chapter.

REFERENCES

1. Steinhoff MC, Omer SB, Roy E, Arifeen SE, Raqib R, Altaye M, et al. Influenza immunization in pregnancy--antibody responses in mothers and infants. *N Engl J Med.* 2010;362(17):1644–1646.
2. DiazGranados CA, Dunning AJ, Kimmel M, Kirby D, Treanor J, Collins A, et al. Efficacy of high-dose versus standard-dose influenza vaccine in older adults. *N Engl J Med.* 2014;371(7):635–645.
3. Black S. Safety and effectiveness of MF-59 adjuvanted influenza vaccines in children and adults. *Vaccine.* 2015;33(Suppl 2). B3-5.
4. Willis NJ. Edward Jenner and the eradication of smallpox. *Scott Med J.* 1997;42(4):118–121.
5. Memoli MJ, Shaw PA, Han A, Czajkowski L, Reed S, Athota R, et al. Evaluation of Antihemagglutinin and Antineuraminidase Antibodies as Correlates of Protection in an Influenza A/H1N1 Virus Healthy Human Challenge Model. *MBio.* 2016;7(2).
6. Kirkpatrick BD, Whitehead SS, Pierce KK, Tibery CM, Grier PL, Hynes NA, et al. The live attenuated dengue vaccine TV003 elicits complete protection against dengue in a human challenge model. *Sci Transl Med.* 2016;8(330):330ra36.
7. Bernstein DI, Atmar RL, Lyon GM, Treanor JJ, Chen WH, Jiang X, et al. Norovirus vaccine against experimental human GII.4 virus illness: a challenge study in healthy adults. *J Infect Dis.* 2015;211(6):870–878.
8. Roestenberg M, de Vlas SJ, Nieman AE, Sauerwein RW, Hermsen CC. Efficacy of preerythrocytic and blood-stage malaria vaccines can be

assessed in small sporozoite challenge trials in human volunteers. *J Infect Dis*. 2012;206(3):319–323.
9. Chen WH, Cohen MB, Kirkpatrick BD, Brady RC, Galloway D, Gurwith M, et al. Single-dose Live Oral Cholera Vaccine CVD 103-HgR Protects Against Human Experimental Infection With Vibrio cholerae O1 El Tor. *Clin Infect Dis*. 2016;62(11):1329–1335.
10. Gaynes RP. *Germ Theory: Medical Pioneers in Infectious Diseases*. ASM Press; 2011.
11. Grabenstein JD, Klugman KP. A century of pneumococcal vaccination research in humans. *Clin Microbiol Infect*. 2012;18(Suppl 5):15–24.
12. Swift HF, Wilson AT, Lancefield RC. Typing Group a Hemolytic Streptococci by M Precipitin Reactions in Capillary Pipettes. *J Exp Med*. 1943;78(2):127–133.
13. Plotkin S. History of vaccination. *Proc Natl Acad Sci U S A*. 2014;111(34):12283–12287.
14. The Global Eradication of Smallpox: Final Report of the Global Commission for the Certification of Smallpox Eradication, Geneva, December 1979: World Health Organization. Global Commission for the Certification of Smallpox Eradication World Health Organization; 1980.
15. Ten great public health achievements–United States, 1900-1999. Centers for Disease Control and Prevention, US Department of Health and Human Services; 1999 Apr 2. Report No.: 0149-2195 (Print) 0149-2195 (Linking) Contract No.: 12.
16. Roush SW, Murphy TV. Vaccine-Preventable Disease Table Working G. Historical comparisons of morbidity and mortality for vaccine-preventable diseases in the United States. *JAMA*. 2007;298(18):2155–2163.
17. Ten great public health achievements–United States, 2001-2010. Centers for Disease Control and Prevention, US Department of Health and Human Services; 2011 May 20. Contract No.: 19.
18. Patel M, Lee AD, Clemmons NS, Redd SB, Poser S, Blog D, et al. National Update on Measles Cases and Outbreaks - United States, January 1-October 1, 2019. *MMWR Morb Mortal Wkly Rep*. 2019;68(40):893–896.
19. Kyaw MH, Lynfield R, Schaffner W, Craig AS, Hadler J, Reingold A, et al. Effect of introduction of the pneumococcal conjugate vaccine on drug-resistant Streptococcus pneumoniae. *N Engl J Med*. 2006;354(14):1455–1463.
20. Daniel C. Payne, Umesh D. Parashar. Rotavirus. Manual for Surveillance of Vaccine-Preventable Diseases Center for Disease Control and Prevention, 2018.
21. Drolet M, Benard E, Perez N, Brisson M. Group HPVVIS. Population-level impact and herd effects following the introduction of human papillomavirus vaccination programmes: updated systematic review and meta-analysis. *Lancet*. 2019;394(10197):497–509.
22. Nair H, Nokes DJ, Gessner BD, Dherani M, Madhi SA, Singleton RJ, et al. Global burden of acute lower respiratory infections due to respiratory syncytial virus in young children: a systematic review and meta-analysis. *Lancet*. 2010;375(9725):1545–1555.
23. Kim HW, Canchola JG, Brandt CD, Pyles G, Chanock RM, Jensen K, et al. Respiratory syncytial virus disease in infants despite prior administration of antigenic inactivated vaccine. *Am J Epidemiol*. 1969;89(4):422–434.
24. Murphy BR, Walsh EE. Formalin-inactivated respiratory syncytial virus vaccine induces antibodies to the fusion glycoprotein that are deficient in fusion-inhibiting activity. *J Clin Microbiol*. 1988;26(8):1595–1597.
25. Crank MC, Ruckwardt TJ, Chen M, Morabito KM, Phung E, Costner PJ, et al. A proof of concept for structure-based vaccine design targeting RSV in humans. *Science*. 2019;365(6452):505–509.
26. Thornhill EM, Salpor J, Verhoeven D. Respiratory syntycial virus: Current treatment strategies and vaccine approaches. *Antivir Chem Chemother*. 2020;28:2040206620947303.
27. Iversen OE, Miranda MJ, Ulied A, Soerdal T, Lazarus E, Chokephaibulkit K, et al. Immunogenicity of the 9-Valent HPV Vaccine Using 2-Dose Regimens in Girls and Boys vs a 3-Dose Regimen in Women. *JAMA*. 2016;316(22):2411–2421.
28. Markowitz LE, Liu G, Hariri S, Steinau M, Dunne EF, Unger ER. Prevalence of HPV After Introduction of the Vaccination Program in the United States. *Pediatrics*. 2016;137(3):1–9.
29. Rts SCTP. Efficacy and safety of RTS,S/AS01 malaria vaccine with or without a booster dose in infants and children in Africa: final results of a phase 3, individually randomised, controlled trial. *Lancet*. 2015;386(9988):31–45.
30. Bhatt S, Gething PW, Brady OJ, Messina JP, Farlow AW, Moyes CL, et al. The global distribution and burden of dengue. *Nature*. 2013;496(7446):504–507.
31. Thomas SJ, Yoon IK. A review of Dengvaxia(R): development to deployment. *Hum Vaccin Immunother*. 2019;15(10):2295–2314.
32. WHO. Global Advisory Committee on Vaccine Safety, 6-7 June 2018. 2018.
33. Dengue vaccine: WHO position paper, September 2018 - Recommendations. Vaccine. 2019;37(35):4848-9.
34. Philippines: Situation Report Measles Outbreak. UNICEF; World Health Organization; 2019.
35. van Riel D, de Wit E. Next-generation vaccine platforms for COVID-19. *Nat Mater*. 2020;19(8):810–812.
36. Gaudinski MR, Houser KV, Morabito KM, Hu Z, Yamshchikov G, Rothwell RS, et al. Safety, tolerability, and immunogenicity of two Zika virus DNA vaccine candidates in healthy adults: randomised, open-label, phase 1 clinical trials. *Lancet*. 2018;391(10120):552–562.
37. Jackson LA, Anderson EJ, Rouphael NG, Roberts PC, Makhene M, Coler RN, et al. An mRNA Vaccine against SARS-CoV-2 - Preliminary Report. *N Engl J Med*. 2020
38. Coffman RL, Sher A, Seder RA. Vaccine adjuvants: putting innate immunity to work. *Immunity*. 2010;33(4):492–503.
39. HogenEsch H, O'Hagan DT, Fox CB. Optimizing the utilization of aluminum adjuvants in vaccines: you might just get what you want. *NPJ Vaccines*. 2018;3:51.
40. Mitkus RJ, King DB, Hess MA, Forshee RA, Walderhaug MO. Updated aluminum pharmacokinetics following infant exposures through diet and vaccination. *Vaccine*. 2011;29(51):9538–9543.
41. Mooney M, McWeeney S, Canderan G, Sekaly RP. A systems framework for vaccine design. *Curr Opin Immunol*. 2013;25(5):551–555.
42. Raeven RHM, van Riet E, Meiring HD, Metz B, Kersten GFA. Systems vaccinology and big data in the vaccine development chain. *Immunology*. 2019;156(1):33–46.
43. Wimmers F, Pulendran B. Emerging technologies for systems vaccinology - multi-omics integration and single-cell (epi)genomic profiling. *Curr Opin Immunol*. 2020;65:57–64.
44. Martin JE, Sullivan NJ, Enama ME, Gordon IJ, Roederer M, Koup RA, et al. A DNA vaccine for Ebola virus is safe and immunogenic in a phase I clinical trial. *Clin Vaccine Immunol*. 2006;13(11):1267–1277.
45. Ledgerwood JE, DeZure AD, Stanley DA, Coates EE, Novik L, Enama ME, et al. Chimpanzee Adenovirus Vector Ebola Vaccine. *N Engl J Med*. 2017;376(10):928–938.
46. Regules JA, Beigel JH, Paolino KM, Voell J, Castellano AR, Hu Z, et al. A Recombinant Vesicular Stomatitis Virus Ebola Vaccine. *N Engl J Med*. 2017;376(4):330–341.
47. Henao-Restrepo AM, Camacho A, Longini IM, Watson CH, Edmunds WJ, Egger M, et al. Efficacy and effectiveness of an rVSV-vectored vaccine in preventing Ebola virus disease: final results from the Guinea ring vaccination, open-label, cluster-randomised trial (Ebola Ca Suffit!). *Lancet*. 2017;389(10068):505–518.
48. Mutua G, Anzala O, Luhn K, Robinson C, Bockstal V, Anumendem D, et al. Safety and Immunogenicity of a 2-Dose Heterologous Vaccine Regimen With Ad26.ZEBOV and MVA-BN-Filo Ebola Vaccines: 12-Month Data From a Phase 1 Randomized Clinical Trial in Nairobi, Kenya. *J Infect Dis*. 2019;220(1):57–67.
49. Zika Virus: World Health Organization; 2020 [Available from: https://www.who.int/news-room/fact-sheets/detail/zika-virus.
50. Rasmussen SA, Jamieson DJ, Honein MA, Petersen LR. Zika Virus and Birth Defects--Reviewing the Evidence for Causality. *N Engl J Med*. 2016;374(20):1981–1987.
51. Interim Guidance for Zika Virus Testing of Urine - United States, 2016. Centers for Disease Control and Prevention, US Department of Health and Human Services; 2016. Report No.: 1545-861X (Electronic) 0149-2195 (Linking) Contract No.: 18.

52. Matheron S, D'Ortenzio E, Leparc-Goffart I, Hubert B, de Lamballerie X, Yazdanpanah Y. Long Lasting Persistence of Zika Virus in Semen. *Clin Infect Dis*. 2016
53. Modjarrad K, Lin L, George SL, Stephenson KE, Eckels KH, De La Barrera RA, et al. Preliminary aggregate safety and immunogenicity results from three trials of a purified inactivated Zika virus vaccine candidate: phase 1, randomised, double-blind, placebo-controlled clinical trials. *Lancet.* 2018;391(10120):563–571.
54. Cohen J. Steep drop in Zika cases undermines vaccine trial. *Science.* 2018;361(6407):1055–1056.
55. Coronavirus Disease (COVID-19): World Health Organization; 2020.
56. Corbett KS, Edwards DK, Leist SR, Abiona OM, Boyoglu-Barnum S, Gillespie RA, et al. SARS-CoV-2 mRNA vaccine design enabled by prototype pathogen preparedness. *Nature.* 2020
57. Corey L, Mascola JR, Fauci AS, Collins FS. A strategic approach to COVID-19 vaccine R&D. *Science*. 2020;368(6494):948–950.
58. Folegatti PM, Ewer KJ, Aley PK, Angus B, Becker S, Belij-Rammerstorfer S, et al. Safety and immunogenicity of the ChAdOx1 nCoV-19 vaccine against SARS-CoV-2: a preliminary report of a phase 1/2, single-blind, randomised controlled trial. *Lancet.* 2020;396(10249):467–478.
59. Sahin U, Muik A, Derhovanessian E, Vogler I, Kranz LM, Vormehr M, et al. COVID-19 vaccine BNT162b1 elicits human antibody and TH1 T cell responses. *Nature.* 2020
60. Keech C, Albert G, Cho I, Robertson A, Reed P, Neal S, et al. Phase 1-2 Trial of a SARS-CoV-2 Recombinant Spike Protein Nanoparticle Vaccine. *N Engl J Med*. 2020
61. Slaoui M, Hepburn M. Developing Safe and Effective Covid Vaccines - Operation Warp Speed's Strategy and Approach. *N Engl J Med.* 2020
62. Cohen MS, Chen YQ, McCauley M, Gamble T, Hosseinipour MC, Kumarasamy N, et al. Prevention of HIV-1 infection with early antiretroviral therapy. *N Engl J Med*. 2011;365(6):493–505.
63. Koester KA, Grant RM. Editorial Commentary: Keeping Our Eyes on the Prize: No New HIV Infections With Increased Use of HIV Pre-exposure Prophylaxis. *Clin Infect Dis*. 2015;61(10):1604–1605.
64. Heberlein-Larson LA, Tan Y, Stark LM, Cannons AC, Shilts MH, Unnasch TR, et al. Complex Epidemiological Dynamics of Eastern Equine Encephalitis Virus in Florida. *Am J Trop Med Hyg*. 2019;100(5):1266–1274.
65. Logunov DY, Dolzhikova IV, Tukhvatullin AI, Shcheblyakov DV. Safety and efficacy of the Russian COVID-19 vaccine: more information needed-Authors' reply. *Lancet.* 2020;396(10256):e54–e55.
66. Excler JL, Michael NL. Lessons from HIV-1 vaccine efficacy trials. *Curr Opin HIV AIDS*. 2016;11(6):607–613.
67. Rerks-Ngarm S, Pitisuttithum P, Nitayaphan S, Kaewkungwal J, Chiu J, Paris R, et al. Vaccination with ALVAC and AIDSVAX to prevent HIV-1 infection in Thailand. *N Engl J Med*. 2009;361(23):2209–2220.
68. Corey L, Gilbert PB, Tomaras GD, Haynes BF, Pantaleo G, Fauci AS. Immune correlates of vaccine protection against HIV-1 acquisition. *Sci Transl Med*. 2015;7(310):310rv7.
69. Experimental HIV vaccine regimen ineffective in preventing HIV [press release]. National Institutes of Health, US Department of Health and Human Services 2020.
70. Ledgerwood JE, Coates EE, Yamshchikov G, Saunders JG, Holman L, Enama ME, et al. Safety, pharmacokinetics and neutralization of the broadly neutralizing HIV-1 human monoclonal antibody VRC01 in healthy adults. *Clin Exp Immunol*. 2015;182(3):289–301.
71. Cohen YZ, Butler AL, Millard K, Witmer-Pack M, Levin R, Unson-O'Brien C, et al. Safety, pharmacokinetics, and immunogenicity of the combination of the broadly neutralizing anti-HIV-1 antibodies 3BNC117 and 10-1074 in healthy adults: A randomized, phase 1 study. *PLoS One*. 2019;14(8):e0219142.
72. Gaudinski MR, Houser KV, Doria-Rose NA, Chen GL, Rothwell RSS, Berkowitz N, et al. Safety and pharmacokinetics of broadly neutralising human monoclonal antibody VRC07-523LS in healthy adults: a phase 1 dose-escalation clinical trial. *Lancet HIV*. 2019;6(10):e667–e679.
73. Caskey M, Schoofs T, Gruell H, Settler A, Karagounis T, Kreider EF, et al. Antibody 10-1074 suppresses viremia in HIV-1-infected individuals. *Nat Med*. 2017;23(2):185–191.
74. Caskey M, Klein F, Lorenzi JC, Seaman MS, West Jr. AP, Buckley N, et al. Viraemia suppressed in HIV-1-infected humans by broadly neutralizing antibody 3BNC117. *Nature*. 2015;522(7557):487–491.
75. Lynch RM, Boritz E, Coates EE, DeZure A, Madden P, Costner P, et al. Virologic effects of broadly neutralizing antibody VRC01 administration during chronic HIV-1 infection. *Sci Transl Med*. 2015;7(319):319ra206.

75a. Corey L, Gilbert PB, Juraska M, Montefiori DC, Morris L, Karuna ST, et al. HVTN 704/HPTN 085 and HVTN 703/HPTN 081 Study Teams. Two Randomized Trials of Neutralizing Antibodies to Prevent HIV-1 Acquisition. *N Engl J Med*. 2021;384(11):1003–1014. doi:10.1056/NEJMoa2031738. PMID: 33730454; PMCID: PMC8189692.

76. Grohskopf LA, Alyanak E, Broder KR, Blanton LH, Fry AM, Jernigan DB, et al. Prevention and Control of Seasonal Influenza with Vaccines: Recommendations of the Advisory Committee on Immunization Practices - United States, 2020-21 Influenza Season. *MMWR Recomm Rep.*. 2020;69(8):1–24.
77. Iuliano AD, Roguski KM, Chang HH, Muscatello DJ, Palekar R, Tempia S, et al. Estimates of global seasonal influenza-associated respiratory mortality: a modelling study. *Lancet.*. 2018;391(10127):1285–1300.
78. Dunning AJ, DiazGranados CA, Voloshen T, Hu B, Landolfi VA, Talbot HK. Correlates of Protection against Influenza in the Elderly: Results from an Influenza Vaccine Efficacy Trial. *Clin Vaccine Immunol*. 2016;23(3):228–235.
79. Krammer F. Novel universal influenza virus vaccine approaches. *Curr Opin Virol*. 2016;17:95–103.
80. Evans TG, Schrager L, Thole J. Status of vaccine research and development of vaccines for tuberculosis. *Vaccine.*. 2016;34(26):2911–2914.
81. Hoft DF. Tuberculosis vaccine development: goals, immunological design, and evaluation. *Lancet.*. 2008;372(9633):164–175.
82. Kaplan G. Rational vaccine development—a new trend in tuberculosis control. *N Engl J Med*. 2005;353(15):1624–1625.

Allergen Immunotherapy for Allergic Diseases

Joana Cosme and Stephen R. Durham

Allergen immunotherapy involves the administration of allergen extracts or allergen products to IgE-sensitized allergic individuals in order to induce a state of durable clinical and immunologic tolerance following allergen re-exposure.[1] This chapter gives a historical perspective and review of indications and contraindications for allergen immunotherapy. There follows a review of the evidence for subcutaneous and sublingual immunotherapy for inhalant allergies. Immunotherapy for Hymenoptera venom anaphylaxis and food allergy are briefly considered. Mechanisms of immunotherapy are reviewed along with implications for biomarker discovery and novel therapeutic approaches.

HISTORICAL PERSPECTIVE

In 1911, Leonard Noon[2] published in The Lancet that extracts of grass pollen injected into patients with seasonal pollinosis resulted in a marked reduction in conjunctival allergen sensitivity with a 10 to 30-folds increase in the concentration of grass pollen allergen necessary to provoke an immediate conjunctival response. Noon died that year, and his colleague John Freeman reported that this was accompanied by an improvement in hay fever symptoms during subsequent natural seasonal exposure. In 1921, Heintz Kustner, who was fish-allergic, demonstrated that following injection of his serum into the skin of his colleague Carl Prausnitz, a subsequent skin prick test with fish extract resulted in a weal and flare, thereby demonstrating passive transfer of allergen hypersensitivity by serum.[3] The nature of this serum factor (coined "reagin") was unknown until 45 years later when Hans Bennich and Gunnar Johansson in Sweden[4] and Kimi and Teruko Ishizaka in USA[5] officially identified immunoglobulin E as a novel class of antibody responsible for this "reaginic" activity. In 1935 Robert Cooke showed that the intradermal injection of serum obtained post-immunotherapy from a ragweed-allergic patient into the skin of a non-allergic individual conferred protection against passive transfer of immediate ragweed skin test hypersensitivity that followed intradermal injection of the same individual's pre-immunotherapy serum.[6] In 1940, Mary Loveless showed that this "blocking" serum factor resided within the immunoglobulin (presumed IgG) fraction of serum,[7] but still many years before the discovery of IgE in 1967. In 1952 William Frankland, a student of Freeman and also based at St Mary's hospital Paddington UK, published the first randomized, blinded controlled trial of grass pollen immunotherapy.[8] He showed that compared to the control diluent, a crude extract of grass pollen injected subcutaneously was effective in improving rhinitis and asthma symptoms during the pollen season. Remarkably, in the same study, he showed that it was the high-molecular-weight protein fraction rather than the non-protein-containing low-molecular-weight fraction that was responsible for the therapeutic activity of the crude grass pollen extract. In the 1960s to 1980s, there followed a series of controlled trials confirming the efficacy of subcutaneous immunotherapy (SCIT) for pollen allergy[9,10] and in mite-allergic children.[11] In 1986, a report from the Committee of Safety of Medicines questioned the safety of subcutaneous immunotherapy following a series of deaths in the UK. Fortunately, immunotherapy practice has moved on, and national and international guidelines now provide guidance for best practice. Together with the publication of large randomized controlled trials, subcutaneous immunotherapy is now recognized as safe and effective.

ALLERGEN IMMUNOTHERAPY FOR ALLERGIC RHINOCONJUNCTIVITIS AND ASTHMA: PLACE IN THERAPY

Allergen Avoidance and Pharmacotherapy

The role of allergy should be critically evaluated in all cases of rhinoconjunctivitis and in bronchial asthma, which is very commonly associated with rhinitis. An allergy history combined with either skin prick testing and/or serum IgE determinations for relevant allergens should always be part of the diagnostic work-up. It is important to identify and, where possible, avoid provoking allergens such as perennial exposure to the dander of domestic pets and house dust mites (HDM). Unfortunately, this is often very difficult in view of the psychosocial implications of owning a family pet and the relative lack of efficacy of even rigorous HDM avoidance measures.

According to ARIA (Allergic rhinitis and its impact on asthma) guidelines,[12] allergic rhinoconjunctivitis is classified according to symptom severity as mild or moderate/severe depending on whether symptoms impact quality of life and usual daily activities. Allergic rhinitis is further classified according to the duration of symptoms as intermittent (less than 4 days per week and/or for less than 4 weeks/year) and persistent (greater than 4 days per week for more than 4 weeks/year. A simplified guide of treatment according to ARIA[12] is summarized in Fig. 88.1. For mild intermittent or persistent symptoms, non/low sedating once-daily oral antihistamines, for example, loratadine or cetirizine (also intranasal antihistamines), have been shown to be effective. For moderate/severe and persistent symptoms, intranasal corticosteroids are first-line treatment and include fluticasone propionate or mometasone furoate once daily. When these treatments are not fully effective, it is

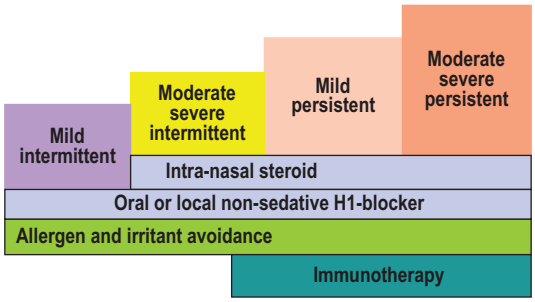


FIG. 88.1 Stepwise guide to treatment of rhinoconjunctivitis. (Simplified from ARIA guideline.[12])

essential to check compliance with therapy and intranasal spray technique. Recently a nasal spray comprising the combination of an intranasal corticosteroid (fluticasone) and antihistamine (azelastine) has been shown to be more effective than either treatment alone.[13] When patients fail to respond to these measures or develop unacceptable side effects, the diagnosis should be checked and adherence to treatment reviewed.

Allergen Immunotherapy

Allergen immunotherapy (AIT) should be considered in patients aged 5 years or older, with allergic rhinoconjunctivitis in whom a dominant allergen or limited spectrum of allergens is mainly responsible for symptoms and in whom the combination of avoidance measures and pharmacotherapy have not been adequately effective and/or resulted in unacceptable side effects.[1] General Indications and contraindications for rhinitis and asthma are summarized in Table 88.1. Immunotherapy is indicated in patients with rhinoconjunctivitis with/without mild asthma who have evidence of symptoms *on exposure* to a relevant allergen and documented *IgE sensitization* to a relevant allergen (SPT and/or Sp-IgE). Contraindications include uncontrolled asthma, malignancy, and active autoimmune disease. Immunotherapy should not be initiated during pregnancy. Recent data suggest that with extracts of proven value, polysensitized patients do as well as monosensitized patients.[14,15]

Allergen immunotherapy may be indicated for children and adults with seasonal asthma complicating rhinitis due to pollens.[14] Sublingual HDM tablet immunotherapy is indicated in adults with HDM-driven asthma as an add-on treatment to regular therapy in order to reduce exacerbations and to decrease symptoms and corticosteroid use.[15] Due to the lack of robust evidence, there are no current indications for the prescription of AIT in asthma driven by other aeroallergens.

For both rhinoconjunctivitis and asthma, a 3-year period of treatment is generally recommended in order to achieve long-term efficacy.[14] In rhinoconjunctivitis, this period can be extended up to 5 years, while for HDM-driven asthma, there does not appear to be an additional benefit of 5-year therapy compared to a 3-year period.[15]

The role of allergen immunotherapy for children and adults with atopic eczema is less clear. In patients with inhalant allergy, immunotherapy is not contraindicated in those with mild-moderate atopic eczema. In severe eczema, evidence for efficacy is weak, and the disease may be exacerbated.[16]

Allergen immunotherapy is highly effective and may be lifesaving in patients who develop anaphylaxis to insect stings of Hymenoptera species (bees, wasps, and hornets).

TABLE 88.1 Allergen Immunotherapy for Inhalant Allergens

Indications for Immunotherapy	Contraindications
• Rhinoconjunctivitis /mild asthma • Symptoms on exposure to relevant allergen • IgE sensitization to relevant allergen • Inadequate response to anti-allergic drugs • Unacceptable drug side effects • Polysensitization not a contraindication	• Severe or uncontrolled asthma • Autoimmunity • Immunodeficiency • Malignancy • Pregnancy—initiation (maintenance OK) • Lack of understanding/poor adherence

Evidence of efficacy for certain food allergies, particularly peanut, is emerging—but immunotherapy for food allergy is not yet recommended as part of routine practice.

Evidence for Efficacy of Allergen Immunotherapy

Allergic rhinoconjunctivitis. Recently the European Academy of Allergy and Clinical Immunology published a systematic review of immunotherapy for allergic rhinoconjunctivitis,[17] followed by a guideline based on this review.[14] 5932 studies were reviewed, of which 160 were suitable for systematic review. Summary data for subcutaneous and sublingual immunotherapy are presented in Fig. 88.2.[17] Efficacy measures are expressed as symptom scores, "rescue medication" scores, and combined scores—the combination of symptom and medication scores. Data are represented as standardized mean differences and 95% confidence intervals compared with placebo. The number of studies analyzed for which data were available for each category, the total number of participants receiving active or placebo treatment, the I^2 value (a measure of the heterogeneity of the data), and levels of statistical significance are shown. Overall, the level of heterogeneity was moderate to substantial for all categories. Confidence intervals were greater for the combined scores since there were fewer studies that included this recently introduced measure of efficacy. The mean differences compared to placebo varying between 0.4 and 0.6 indicate a moderate level of efficacy for both subcutaneous and sublingual immunotherapy routes.

Subgroup analyses showed that mono/polysensitization and presence/absence of asthma did not influence the level of efficacy of immunotherapy for rhinoconjunctivitis. The treatment was effective in all age groups studied, although less data were available for children and the elderly. Based on preventative studies, allergen immunotherapy initiated below the age of 5 years could be effective (or more effective). Few studies are available, and this remains an area of high priority. Immunotherapy should not be initiated in pregnancy, although effective maintenance immunotherapy may be continued during pregnancy.

Examples of individual studies include a phase III trial in the UK of 410 grass pollen allergics with moderate-severe hay fever who were randomized to receive subcutaneous alum-based grass pollen extract in two doses (containing 2 or 20 µg major allergen Phleum p 5 in maintenance injections) per week for a month, and then monthly over approximately 8 months. There was a dose- and time-dependent reduction in seasonal hay fever symptoms, with a 30% average reduction for the 20 µg group, accompanied by dose-dependent increases in specific IgG IgE-blocking antibodies.[18] In another randomized placebo-con-

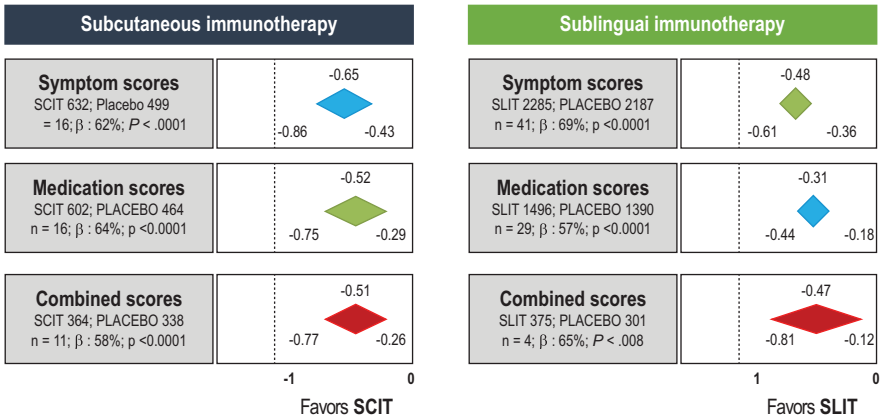


FIG. 88.2 Meta-analysis of subcutaneous immunotherapy v placebo and sublingual immunotherapy v placebo for allergic rhinoconjunctivitis. A comparison of mean differences and confidence intervals of active v placebo treatments for symptom scores, rescue medication scores, and combined scores. (Data derived from Dhami et al.[17])

trolled trial of sublingual immunotherapy in 992 HDM-allergic rhinitics, participants were randomized to receive a sublingual tablet containing approximately 15 mcg of Der p1 and Der p 2 major allergen daily (or half these doses).[19] There was a 20% reduction in total combined symptom-rescue medication scores at the end of 12 months of treatment. It is noteworthy that the World Allergy Organization criteria for a clinically meaningful reduction in the combined score is 20%. For both studies, clinical improvement was measured over and above usual treatment with anti-allergic drugs available to all participants in both active and placebo-treated arms.

Based on a systematic review of evidence using "Agree" criteria, the recent EAACI Guideline on allergen immunotherapy for allergic rhinoconjunctivitis[14] recommended that an *individual product-based* evaluation of evidence for short-term efficacy (i.e., for the duration whilst on treatment) is indicated before treatment with a specific product is initiated—with continuous subcutaneous treatment for seasonal rhinitis (Grade A evidence), pre and pre/co-seasonal SCIT, for both modified (allergoids) and non-modified allergen extracts and for SLIT aqueous solutions for grass/tree pollen. Recent evidence has also confirmed efficacy during treatment with sublingual tablets for seasonal rhinitis (grass, tree, ragweed pollens) and perennial rhinoconjunctivitis (HDM) in adults and children.

Allergic Bronchial Asthma. Allergen immunotherapy for bronchial asthma was evaluated in a recent EAACI systematic review.[20] Of 5997 records screened, 98 were suitable for qualitative review and 15 for quantitative meta-analysis with a standardized mean difference compared to placebo of −1.11 (confidence intervals −1.66, −0.56) for symptom scores and −1.21 (−1.87, −0.54) for medication scores but both with substantial evidence of publication bias. The conclusion was that immunotherapy can achieve substantial reductions in short-term symptoms and medication use for allergic asthma, although at the expense of a modest increase in the risk of systemic and local allergic reactions. From limited data on cost-effectiveness, sublingual immunotherapy was considered cost-effective for asthma, although long-term studies are required.

The EAACI guideline based on this meta-analysis included Grading of Recommendations Assessment, Development, and Evaluation approach (GRADE) criteria.[15] The expert group focused on immunotherapy for HDM allergic asthma and concluded there was insufficient evidence to make recommendations for other allergens. The committee concluded that the important prerequisites were (1) optimal selection of patients based on a history of HDM-driven asthma (with HDM provocation testing if in doubt) as confirmation of IgE-sensitization (2) use of individual allergen extracts and desensitization procedures of proven value. Only HDM tablet immunotherapy was found to have a robust effect for relevant asthma critical endpoints (exacerbations, asthma control, and safety) and was recommended for partially controlled HDM-driven asthma (conditional recommendation, moderate quality evidence).[15]

Evidence for the EAACI recommendation was from a randomized controlled trial of 12 months duration of two doses of HDM tablet immunotherapy in 834 participants with partially controlled asthma. The participants' inhaled corticosteroid treatment was withdrawn 50% after 6 months and completely at 9 months. The primary outcome was time to first moderate-severe asthma exacerbation up to 12 months. There was observed a 31% reduction for the six subcutaneous HDM tablets and a 34% reduction for the 12 subcutaneous tablets compared to placebo treatment. Side-effects were mainly local irritation and swelling in the mouth, and an overall 6% withdrawal rate due to side effects was noted.[21] HDM tablet immunotherapy now appears in the most recent GINA asthma guideline as an alternative add-on treatment at steps 3 and 4 in HDM-driven asthma. HDM SCIT was recommended in adults and children for reduction of asthma symptoms and medication needs (conditional recommendation low-quality evidence).[22]

Long-Term Benefits of Allergen Immunotherapy

Persistence of Benefit

A major advantage of allergen immunotherapy is its potential disease-modifying effects with persistence of benefit for years after its discontinuation.[23] In contrast, whereas pharmacotherapy for allergic rhinitis and asthma is effective in improving symptoms and quality of life, there are no long-term benefits. This is true not only for antihistamines and beta-sympathomimetic drugs but also for inhaled/oral corticosteroids, anti-IgE monoclonal antibodies, and modern anti-Th2 biologic therapies

In a study of 47 patients with moderate-severe grass pollen allergy, 32 had completed 3 to 4 years of high-dose alum-based subcutaneous immunotherapy and were randomized to either continue for a total of 6 to 7 years on active injections or withdrawn at 3 years of matched placebo injections of identical appearance. Fifteen matched, but non-randomized hay fever sufferers were followed during the withdrawal phase for a total of 3 years. The group that withdrew at 3 years had comparable efficacy (>30% mean difference from placebo) as those for 3 to 4 years on immunotherapy and comparable to the continued efficacy during 3 years of those that continued immunotherapy for a total of 6 to 7 years. Continued clinical improvement was accompanied by a maintained reduction in immediate cutaneous and conjunctival allergen sensitivity, marked and persistent suppression of intradermal allergen-induced late-phase skin response throughout the 7 years. There was also a reduction in allergen-stimulated CD4 T cells and IL-4 mRNA$^+$ cells as determined by immunohistochemistry and *in situ* hybridization of microscopic sections of biopsies taken from the sites of suppressed cutaneous late-phase responses at 7 years for both those maintained and withdrawn from treatment.[24]

Two independent large randomized, double-blind placebo-controlled trials of sublingual tablet immunotherapy in adults with moderate-severe grass pollen seasonal rhinitis involved either continuous treatment or a pre/co-seasonal regimen over 3 years, with blinded follow up for 2 years off treatment. Both regimens were highly effective (>30% mean difference compared to placebo) in reducing seasonal symptoms and medication scores throughout the 3 years of treatment. There was also persistence of benefit (approximately 30% reduction in symptom and medication scores) during successive pollen seasons during the 2 years' follow up off treatment.[23]

In one large randomized controlled trial of HDM sublingual tablets in adults with perennial allergic rhinoconjunctivitis, 1 year's treatment was effective in reducing symptoms and rescue medication, and the effects persisted during a second year of treatment. This raises the question of whether continuous natural exposure to a perennial allergen following a shorter course of immunotherapy may be more effective in maintaining tolerance. This requires confirmation in a prospective trial with tolerance as the primary outcome.[23]

EAACI guidelines recommend continuous grass pollen SCIT and grass pollen SLIT tablets and SLIT aqueous solution for both short-term (on treatment) and long-term (off-treatment) benefits in adults and children and for HDM tablets (but not the solution). The guidelines advise a minimum of 3 years of treatment for long-term efficacy and that the selection of individual allergen products for treatment should be evidence-based.[14,15]

Prevention of Asthma

The Prevention of Asthma Trial (PAT) was a randomized controlled trial of a subcutaneous alum-based grass pollen vaccine in children aged 5 to 12 years who had seasonal rhinitis but no asthma. The primary endpoint was time to onset of an asthma diagnosis up to 5 years, 2 years following 3 years of immunotherapy. The treatment was successful in significantly reducing asthma prevalence (odds ratio 2.68, confidence interval [1.3 to 5.7]). Limitations were that the diagnosis of asthma was based on subjective criteria, and although randomized, the trial was non-blinded for ethical reservations concerning the need for 3 years of placebo injections in children.[23]

Recent EAACI guidelines,[14,15] based on such evidence, advise the use of allergen immunotherapy for birch/grass pollen rhinoconjunctivitis in children (both sublingual and subcutaneous routes) for (1) a sustained effect on symptoms and rescue medication after cessation of treatment and for (2) prevention of the onset of asthma onset for up to 2 years post-treatment and possibly longer, but more evidence is required.

Subcutaneous versus Sublingual Immunotherapy

Indirect comparison of symptom and medication scores for subcutaneous compared with sublingual immunotherapy from the recent meta-analysis (see Fig. 88.2) might suggest a trend in favor of the subcutaneous route.[17] However, the studies for subcutaneous immunotherapy are fewer in number and include older studies that may not have been performed with the same rigor as more recent studies of sublingual immunotherapy. Furthermore, there is considerable overlap in confidence intervals.

There are few head-to-head comparisons of the two modalities.[25] In one double-blind study of birch pollen immunotherapy, reductions in seasonal symptoms and rescue medication were numerically similar, although the study was underpowered to show a significant difference if one existed. In a recent double-blind placebo-controlled comparison in patients with moderate-severe seasonal grass pollen rhinoconjunctivitis, nasal allergen provocation was used as a surrogate clinical primary endpoint.[26] Both treatments were effective compared to placebo. The onset of efficacy for subcutaneous immunotherapy (after 1 year's treatment) was earlier than for the sublingual tablet route in reducing total nasal symptom scores from 0 to 10 hours after nasal challenge. Both treatments were effective at reducing allergic rhinitis symptoms at year 2, both following nasal allergen challenge and for symptoms and quality of life during the pollen season. However, 2 years of treatment was not effective in reducing allergic rhinitis symptoms at year 3, 1 year after treatment completion. This study confirms the need for 3 years of continuous grass pollen immunotherapy treatment to achieve long-term efficacy. Phase 3 head-to-head comparisons with seasonal combined symptom-medication scores as primary endpoint would be informative.

The indications for the subcutaneous and sublingual routes are essentially the same and require a prescription by specialists familiar with the diagnosis and treatment of allergic rhinitis and asthma, knowledge of the indications for immunotherapy, and the ability to recognize and treat early signs of anaphylaxis.[25] The decision depends on the availability of local resources and patient choice (Table 88.2). Both sublingual and subcutaneous allergen immunotherapies are effective in seasonal rhinitis and, with allergen extracts of proven value, are associated with long-term remission. For perennial rhinitis, there is higher quality evidence in favor of sublingual immunotherapy. Adherence to allergen immunotherapy is crucial for efficacy and requires close monitoring. The sublingual route may be more acceptable for children, as it avoids frequent injections. In the USA, subcutaneous immunotherapy is more prevalent, whereas, in Europe, practice is more heterogeneous but swinging in favor of the sublingual route. Overall, the patient is in equipoise—any advantage of subcutaneous treatment in terms of efficacy must be weighed against the convenience and safety of the sublingual route (Fig. 88.3).[25]

Safety

Subcutaneous allergen immunotherapy is an effective and safe treatment when performed in a specialist setting, according

TABLE 88.2 Comparison of Subcutaneous and Sublingual Immunotherapy

Subcutaneous	Sublingual
• Effective in seasonal rhinitis	• Effective in seasonal rhinitis
• Induces long-term remission	• Induces long-term remission
• Effective in perennial rhinitis	• Effective in perennial rhinitis
• Evidence base less in children	• More acceptable for children
• Local side effects (pain, swelling at injection site, not troublesome)	• Local side effects (itch, swelling in mouth, may be bothersome)
• Risk of anaphylaxis	• Severe systemic reactions very rare
• Administration in a specialist clinic	• Suitable for self-administration
• Adherence easily monitored	• Adherence may be a problem
• Direct comparisons with SLIT needed	• Direct comparisons with SCIT needed

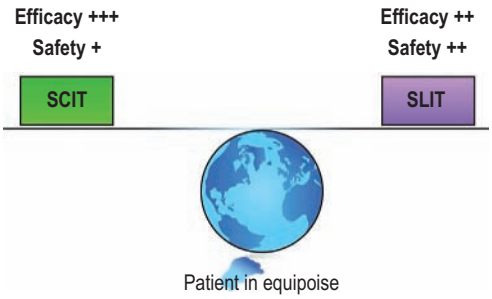


FIG. 88.3 Choice of subcutaneous v sublingual immunotherapy: a balance of efficacy, safety, and patient choice. (Reproduced with kind permission of *J Allergy Clin Immunol*.[25])

to international guidelines and with access to adrenaline and other resuscitative measures. Nonetheless, a major barrier to the wider adoption of subcutaneous immunotherapy has been the risk of systemic allergic side effects. In the UK, a Committee on Safety of medicine's report in 1986 identified 26 deaths due to "desensitization" in the preceding 30 years.

The majority of deaths occurred in a general practice setting. In the 19 deaths for whom the indication was known, 17 of these prescriptions were for bronchial asthma, the time of onset for most reactions being within 30 minutes, although for 2, it was up to 90 minutes. The committee highlighted the need for special caution in asthma that should be well-controlled and for an observation period of 2 hours following injections (subsequently reduced to 60 minutes). In a multi-center trial of alum-based grass pollen immunotherapy in the UK, 9 of 203 participants (4.3%) randomized to receive the top dose (20 µg major allergen Phleum p5) developed grade 3 systemic reactions with recovery, and none received adrenaline.[18] A 12-years survey from the USA reported fatal reactions in 1 per 2.5 million injections and an average of 3.4 deaths per year.[27] Of 17 fatal reactions, 15 occurred in patients whose asthma was not optimally controlled. Risk factors for systemic allergic reactions include dosing errors, delayed/absent use of adrenaline, a history of previous systemic reactions, use of grass pollen or cat allergen extracts, co-seasonal administration of extracts, and patient-related co-factors such as uncontrolled asthma, use of beta-blockers, upper respiratory infections, exercise, and fatigue.

Sublingual immunotherapy is safer than subcutaneous immunotherapy and suitable for home administration.[28] In a meta-analysis of 66 studies of sublingual immunotherapy that included 4000 participants and more than one million doses, one systemic reaction was observed every four treatment years and only one severe reaction for every 384 treatments.[29] There have been isolated case reports of systemic allergic reactions following sublingual immunotherapy, but no fatalities, and these tended to occur when treatment was not practiced according to published guidelines (use of unstandardized extracts, excessive allergen doses, and patients who had experienced a severe reaction during previous subcutaneous immunotherapy). Patients should be observed for at least 30 minutes after the first dose by staff able to recognize and treat anaphylaxis. Although rare, most systemic reactions to SLIT occur at home, so it is important to educate patients on how to recognize and treat reactions and when to seek medical help. Co-prescription of auto-injectable adrenaline devices is recommended in the USA, whereas this is not a routine recommendation in Europe unless there are risk factors, including a history of previous systemic reactions to immunotherapy. Local side effects such as itching and swelling in the mouth and throat irritation frequently occur in 40 to 75% of cases, although in general are mild, short-lived, last for minutes after administration, and resolve within 2 to 3 weeks. Local side effects may occasionally be more troublesome and/or persist for longer and, in clinical trials, result in discontinuation of sublingual immunotherapy in up to 4% to 8% of patients treated with sublingual tablets.

Both sublingual and subcutaneous allergen immunotherapies are effective in seasonal rhinitis and associated with long-term disease remission. For perennial rhinitis, both routes are effective with higher quality evidence in favor of sublingual immunotherapy.

Hymenoptera Venom Immunotherapy

Hymenoptera species include venomous wasps, hornets, and bees.[30] More than half the population report being stung, with adverse reactions ranging from large local reactions (>10 cms) to systemic reactions, including anaphylaxis and occasional fatalities. The rate of systemic reactions in Europe has been estimated between 0.3% and 7.5% in adults and 0.15% and 3.4% in children. Risk factors for anaphylaxis include a bee rather than a wasp sting, older age, underlying medical conditions, raised baseline serum tryptase, mastocytosis, and the history of severity and rapidity of onset of any previous systemic reaction following a sting. Diagnosis depends on the clinical history of a sting and confirmation of specific IgE sensitivity, in cases of doubt, including measurement of IgE to major allergens for bee (Api m 1) and for wasp (Vesp v 1 and 5). Large local reactions require either no treatment, the elevation of an affected limb and ice, or, if severe, oral corticosteroids for 1 to 3 days. Often, antibiotics or antihistamines are prescribed for large, local reac-

tions but are ineffective, unnecessary, and potentially harmful. Although local reactions occasionally precede subsequent systemic reactions (up to 15%), local reactions are not considered an indication for immunotherapy.[31]

Moderate-severe systemic reactors should adhere to avoidance advice and be supplied with two auto-injectable adrenaline pens and their technique of use described and rechecked at intervals, along with the expiry date of the device. Moderate-severe reactors (whose reactions exceed generalized skin reactions) should be offered subcutaneous immunotherapy, with a lowered threshold for those with a severely impaired quality of life due to fear of a subsequent sting. The threshold for immunotherapy is higher in children in view of generally milder reactions and lower risk compared with adults.

Venom immunotherapy confers complete protection from severe systemic reactions in 77 to 84% of cases for honeybee and 91% to 95% for wasp venom. Immunotherapy protocols are similar to those for inhalant allergens, with weekly updosing injections for 12 to 16 weeks followed by monthly injections for 1 year, extending to 6 to 8 weeks between maintenance injections in subsequent years—generally 3 years for wasp venom and 3 to 5 years for bee venom, depending on the risk of subsequent stings and access to medical care. Risk factors for systemic reactions (approximately 8% to 20%) during venom immunotherapy include bee venom, rapid updosing, and probably raised basal serum tryptase levels. Venom immunotherapy confers long-term tolerance although there are no long-term randomized controlled trials, so it is difficult to unravel tolerance due to long-term effects from natural tolerance due to declining IgE levels in the absence of subsequent stings.[30, 31]

Allergen Immunotherapy for Foods

Diagnosis of food allergy depends on the history combined with objective confirmation of IgE-sensitization to relevant foods by skin testing and/or serum-specific IgE. Specific IgE tests to recombinant allergens are often useful to distinguish true allergy to major protein antigens associated with severe reactions and distinguish them from cross-reacting allergens that may be either irrelevant or associated with pollen food syndromes. Oral provocation tests may be indicated when the diagnosis remains in doubt and/or to exclude food allergy to specific foods and enable their re-introduction into the diet. Standard treatment for food allergy involves accurate identification of offending foods and appropriate avoidance advice. Patients with food-induced anaphylaxis should additionally be supplied with two auto-injectable adrenaline pens with clear and repeated instructions on their use and replacement on expiry. In high-risk children with peanut allergy, anti-IgE therapy with omalizumab has been shown to be effective in raising the threshold in order to afford protection against accidental exposure, although this is expensive and would need to be continued long-term. Allergy to milk and egg in infants and children is usually self-limiting with age. Persistent allergy to egg and milk in adults is often serious and life-threatening, and attempts at oral immunotherapy are associated with a high risk of severe adverse events such that immunotherapy for egg and milk allergy is not recommended for use outside randomized controlled trials and confined to specialist centers.

A major area of research has been specific immunotherapy for peanuts. Peanut allergy is common in westernized countries where it has been estimated to affect 1.4% to 3% of the population and represents a common cause of food-induced anaphylaxis. It is a major risk factor for fatal reactions in asthmatics and is associated with expense and substantial psychosocial burden for affected individuals and their families. Most studies have focused on oral peanut immunotherapy in children, whereas sublingual and epicutaneous routes have also been tested.

PALISADE[32] was a phase 3 trial of oral peanut immunotherapy in 551 participants aged 4 to 55 years (mainly under 17) with a severe peanut allergy. Participants were selected by failing a double-blind oral food challenge at 100 mg peanut protein. They were randomized 3:1 to active or placebo treatment over a period of 24 weeks. The primary endpoint was the proportion who could ingest 600 mg peanut protein at the end of treatment. The results showed that in children and adolescents who received encapsulated peanut (AR101), approximately 2/3 achieved the primary endpoint compared with 1/25 of placebo-treated participants. Local and gastrointestinal side-effects were two-fold higher in AR1010-treated participants. Systemic allergic reactions occurred in 53/372 (14.2%), including one case of severe anaphylaxis compared to 4/124 (3.2%) in placebo-treated participants.

A recent meta-analysis (PACE) of nine randomized controlled trials of oral peanut immunotherapy[33] in over 1000 individuals showed that whereas the treatment was effective in achieving desensitization (risk ratio for passing a supervised oral peanut challenge of 12.42/1), the treatment was accompanied by a 3.12/1 risk ratio for anaphylaxis and 2.21/1 for receiving epinephrine compared to placebo. In an accompanying editorial, the question was raised whether trading treatment-related side-effects at home during treatment for allergic reactions to accidental exposures out of the house (i.e., in social situations) might be beneficial for some patients.

In contrast to inhalant allergen immunotherapy, oral immunotherapy for peanuts has not been associated with durable tolerance after discontinuation. The POISED study[34] evaluated the sustained effects of 2 years of oral peanut immunotherapy over 1 year after discontinuation. 120 participants aged 7 to 55 years with a severe peanut allergy were enrolled. A maintenance dose of 4-g peanut was achieved by 1 year in most participants. Whereas effective desensitization to 4-g peanut challenge was achieved, discontinuation, or even a reduction to 300 mg daily peanut, markedly increased the likelihood of return of clinical reactivity to peanut within months of discontinuation.

A phase 3 trial of epicutaneous immunotherapy was performed in 356 enrolled participants, median age 7 years, with peanut allergy but no history of anaphylaxis.[35] The treatment involved daily application to the skin of a peanut patch containing 250 mcg of peanut protein over 12 months. 35.3% on active treatment compared to 13.6% placebo-treated patients achieved the pre-determined peanut-eliciting dose (300 mg for most participants), although the trial did not achieve a component of the primary endpoint. Subsequent analyses and follow-up indicated persistent protection while the participants remained on treatment. In general, the treatment was well-tolerated, although local application site reactions were common for both active and placebo-treated groups (approximately 90%) and discontinuation rates were comparable (approximately 10%).

Taken together, these data suggest that oral peanut immunotherapy is effective but labor-intensive and associated with

frequent systemic allergic reactions, including increased rates of anaphylaxis and adrenaline usage during the treatment. With currently available strategies, long-term disease remission after discontinuation has not been achieved. A comparison between the oral and epicutaneous routes in terms of benefits and side-effects would be of interest. Data suggest that the earlier introduction and targeting of younger and less IgE-sensitized individuals is likely to be more effective. At present, however, immunotherapy for food allergy in general, and peanut allergy for which most data are available, should be confined to research and specialist centers and is not recommended for routine clinical use.

Mechanisms of Allergen Immunotherapy

A greater understanding of the underlying mechanisms of immunotherapy is important for developing biomarkers to assess disease severity, predict responders, and monitor response to treatment. Insights into mechanisms have provided a rationale for novel approaches to immunotherapy, including alternative routes, modified allergens, and "allergen+" strategies to improve efficacy, convenience for patients, and enable shorter, safer protocols.

Mechanisms of Allergic Rhinitis

Nasal allergen provocation has been used as a model to study mechanisms of allergic rhinitis and the influence of immunotherapy. The early nasal response at 0 to 60 minutes after challenge involves immediate sneezing followed by watery nasal discharge and eye symptoms. The early response occurs following allergen-crosslinking of IgE receptors on the surface of mast cells. A sequence of intracellular signaling events results in the immediate release of granule-associated mediators, including histamine and tryptase (the latter largely mast-cell-specific), and within minutes the release of newly formed mediators derived from membrane lipid including the sulfidopeptide leukotrienes C4 (LTC4), LTD4 and LTE4, platelet-activating factor, and, largely specific for mast cells, prostaglandin D2 (PGD2).

The late phase nasal response is associated with tissue eosinophilia and the recruitment and activation of basophils and Th2-type CD4-positive T cells. Innate lymphoid cells (ILCs) do not possess surface lineage markers and do not express T cell receptors, so they are therefore unable to recognize allergen. ILCs respond to epithelial cytokines such as thymic stromal lymphopoietin (TSLP) and IL-33. These cytokines also recruit and activate local dendritic cells ("DC2s") that promote preferential T-cell development in favor of a Th2 phenotype

An additional feature of allergic rhinitis is the activation of peripheral blood basophils and their recruitment and transepithelial migration during the pollen season. Whereas inhibition of nasal challenge is useful for proof of concept for immunotherapy and allergen dose-response studies, exposure challenges in an environmental allergen chamber are more like natural allergen exposure and have been shown to have similar underlying mechanisms.

Mechanisms of Allergen Immunotherapy

Allergen immunotherapy has been shown to inhibit the early nasal response with a corresponding inhibition of mast cell recruitment, activation, and decreases in histamine and leukotriene release as determined by measurements in collected nasal fluid. Immunotherapy inhibits the late nasal response and accompanying release of Th2 cytokines IL-4, 5, 9, and 13 and eosinophil chemotactic factors such as CCL11 (Eotaxin), CCL17, and CCL22. Local eosinophilia is reduced along with a decrease in local Th2-positive T cells. These inhibitory changes are mimicked during the pollen season. Immunotherapy additionally inhibited seasonal transepithelial migration of mast cells and basophils and CD1a-positive dendritic cells, decreased tissue eosinophilia, and inhibited local IgE-synthesis. These local reductions in Th2-dependent events are accompanied by increases in local regulatory T cells, including IL-10- and Transforming growth factor-beta (TGF-β)-producing peripheral Tregs. As well as downregulating Th2 T-cell responses, these cytokines are major switch factors in favor of IgG4 and IgA heavy chain switching, respectively, consistent with the observed local and systemic increases in allergen-specific IgG4 and IgA during immunotherapy. There are also local increases in FOXP3-expressing-CD25 high $CD3^+$ cells, presumed thymus-derived central T_{regs}. This induction of Tregs, detectable also in peripheral blood, occurs within 3 months, whereas successful immunotherapy is also accompanied by a more delayed-in-time increase in Th1 cells detectable in the skin, nose, and peripheral blood. The accompanying increase in interferon gamma$^+$ T cells suppresses IL-4 dependent IgE production and favors overall switching in favor of IgG isotypes, particularly IgG1 and IgG2, also observed during allergen immunotherapy (Fig. 88.4).[36]

Recent novel findings include inhibition of seasonal increases in peripheral blood ILC2s during the pollen season after immunotherapy and a change in ILC2 phenotype in favor of IL-10-producing ILC2s with typical regulatory properties. The induction of regulatory B cells as an alternative source of IL-10 following allergen immunotherapy is increasingly recognized.[36] Th2A cells represent a novel subset of Th2 T cells that increase during the pollen season and are characterized by low expression of the surface marker CD27. Th2A cells are inhibited following successful immunotherapy.

T follicular helper cells (Tfh) are located within the germinal centers of lymph nodes and are characterized by surface expression of CXCR5, intracellular transcription factor Bcl-6, and produce abundant IL-21. Through IL-21, Tfh cells are potent inducers, together with IL-4 of B-cell class switching to IgE. Immunotherapy has been shown to downregulate peripheral Tfh cells in favor of so-called T follicular regulatory (Tfr) cells that have a similar phenotype to Tfh cells, although additionally express FOXP3 and have inhibitory properties including inhibition of Tfh cells.[37] There is the emergence of a novel subset of IL-35-producing regulatory T cells. IL-35 has inhibitory properties like IL-10, induces regulatory B cells, and inhibits IL-4- and IL-21-stimulated human IgE synthesis *in vitro*. Peripheral IL-35 Treg numbers and IL-35 concentrations in allergen-stimulated *in vitro* T-cell culture supernatants are low in allergic rhinitis compared to normal healthy controls. Peripheral numbers of IL-35 Tregs and IL-35 production are increased during grass pollen immunotherapy and correlate with the accompanying suppression of nasal symptoms.[38]

Summary

A pattern is emerging where allergen immunotherapy for inhalant allergens is shown to potently inhibit Th2 T-cell and innate lymphoid cell activation and IgE-dependent mast cell

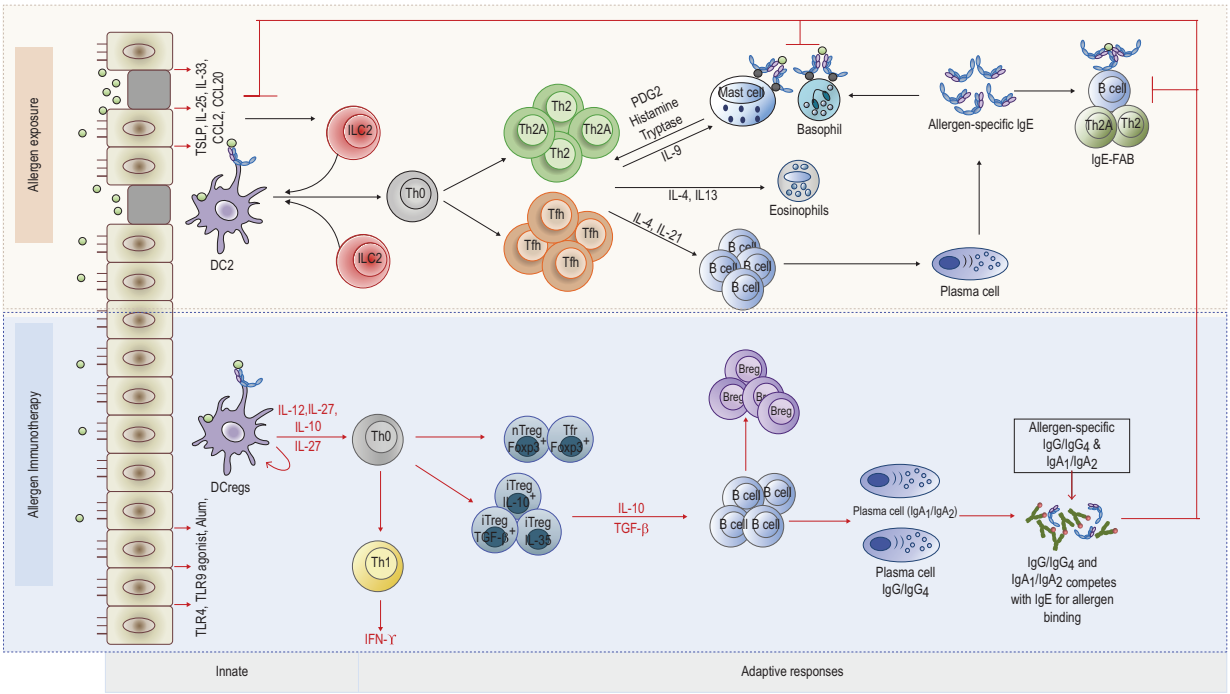


FIG. 88.4 Mechanisms of AIT. During the initial sensitization phase in patients with allergic rhinitis, low allergen exposure at the nasal mucosal surface results in activation of epithelial cells, which then activate DCs. DCs uptake and present antigens to naive T cells to induce allergic TH2 (Th2A) responses and IgE-facilitated antigen presentation. Subsequent allergen re-exposure leads to mast cell and basophil degranulation, causing classic early-phase reactions. Subsequent infiltration of other leukocytes leads to late-phase allergic inflammation. High-dose allergen exposure by immunotherapy restores DC function, which produces IL-12, IL-27, and IL-10 and promotes immune deviation from a TH2 to TH1 response and induction of Treg and Breg cells (including other B-cell subsets) that produce IgA, IgG, and IgG4 blocking antibodies. Suppressive activities of Treg cells, Breg cells, and IgG-blocking activity are indicated by red arrows. EC, Epithelial cells; TLR, Toll-like receptor. (Figure and legend reproduced with kind permission of *J Allergy Clin Immunol*.[36])

and basophil activation and tissue eosinophilia (see Fig. 88.4). This may occur because of an early induction of Treg responses, including peripheral IL-10-, TGF-beta, and IL-35-producing Tregs and a more delayed-in-time immune deviation in favor of allergen-specific Th1 responses at approximately 12 months. Different subsets of B regulatory cells are increasingly recognized as an alternative source of IL-10 following immunotherapy.[36] In addition to suppression of the Th2 response, "IgE-blocking" antibodies, particularly allergen-specific IgG4 and IgA, compete with IgE for the formation of allergen-IgE complexes. As a result, there is a decrease in both Fc€R1-dependent activation of mast cells and basophils and suppression of Fc€R2-dependent IgE-facilitated antigen-stimulated activation of Th2 T-cell development. Interestingly these "protective" IgE-inhibitory antibodies may also be detected locally in nasal fluid as well as in peripheral blood and correlate better with response to immunotherapy than simply measuring immunoreactive IgG4 and IgA levels.[39] A novel recent finding is that whereas for subcutaneous immunotherapy, the dominant blocking antibody is IgG4, for sublingual immunotherapy, the major IgE-inhibitory activity resides within IgA.[39] Thus, the two routes of immunotherapy (SCIT and SLIT) may act by distinct mechanisms, and in resistant cases, it may make sense to combine both routes—although this would first require confirmatory clinical trials.

THERAPEUTIC PRINCIPLES

- Allergen Immunotherapy is effective in seasonal and perennial allergic rhinoconjunctivitis
- Use of either SCIT or SLIT depends on the availability of resources and patient preference
- Three years of treatment with allergen extracts of proven value results in long-term efficacy
- HDM tablet immunotherapy for mite-induced asthma reduces asthma exacerbations
- Subcutaneous immunotherapy for Hymenoptera venom allergy is highly effective
- Oral peanut immunotherapy induces desensitization but not long-term tolerance
- Knowledge of mechanisms of immunotherapy should enable biomarkers to predict and/or monitor response and to develop more effective, safe, and more convenient strategies

Biomarkers of Response to Immunotherapy

At present, a history of symptoms on exposure to the relevant allergen and confirmation of IgE sensitization by skin test and/or measuring specific IgE are the two most reliable predictors of an individual's response to immunotherapy. If the history is doubtful, some centers recommend nasal or conjunctival provocation as a clinical surrogate to confirm end-organ response—although there is little information on threshold dose to confirm

clinical relevance and/or response to immunotherapy. Elevated IgG/IgG4 levels are indicative of exposure to the vaccine rather than a response to treatment, and there is evidence that functional IgE-blocking activity may more reliably predict response to treatment in clinical trials. These include inhibition by post-immunotherapy serum of allergen-IgE complex binding by CD23-expressing B cells as a surrogate of Th2 T-cell development. Second, inhibition *in vitro* of allergen-stimulated basophil histamine release. *Ex vivo* basophil activation may also be measured directly using whole blood allergen stimulation and CD63 measurement. Circulating allergen-responsive TH2A cells, ($CD27^{-}CCR4^{+}CRTH2^{+}CD45R0^{+}$ $CD4^{+}$ T cells) can be measured before/after immunotherapy. These assays require flow cytometry, which is not routinely available for clinic sampling' and real-time measurements. Recently, by use of whole blood RT-PCR, a pro-tolerogenic signature in peripheral dendritic cells that expressed high levels of C1Q and Stabilin 1 was recognized that correlated with immunotherapy response following sublingual grass pollen immunotherapy. A recent summary of potential biomarkers for immunotherapy was published by an EAACI task force report. Whether these assays may predict individual responses to immunotherapy as opposed to identifying significant associations at a group level in clinical trials remains to be determined.[40]

Novel Therapeutic Approaches

Subcutaneous allergen immunotherapy, although effective, is time-consuming and requires specialist supervision. There is a need for shorter, safer, more convenient immunotherapy products and protocols.[41] Sublingual treatment is safer and can be self-administered. Both routes require 4 to 6 months for efficacy in pre-co-seasonal regimens and a minimum of 3 years for long-term tolerance. One approach has been the chemical modification of allergens to reduce allergenicity—such allergoids may be effective but have no proven value over conventional extracts. Allergen extracts have been formulated with alum or tyrosine in order to delay absorption and have adjuvant properties. Alternative routes include epicutaneous "patch" application of allergen and intralymphatic injections directly into lymph nodes under ultrasound guidance.[41]

Peptide immunotherapy involves the use of shortened peptides derived from either crude standardized extracts or synthetic peptides. Peptide immunotherapy attempts to reduce allergenicity while preserving or increasing immunogenicity. Examples include Cat Fel d 1 peptides of 12 to 20 amino acid length that appeared effective and safe at phase 2 but have not been successful so far in phase 3 studies Similarly, medium-length peptides—either recombinant peptides or prepared by controlled hydrolysis of whole allergen extracts showed early promise but have not so far been successful at phase 3. The use of recombinant allergens, both for specific diagnosis and as products for immunotherapy, is an attractive approach as it involves a high degree of standardization and the potential to tailor products according to individual allergen sensitivities. One approach has involved a recombinant mixture of major grass allergens covalently linked to a highly immunogenic pre-S protein derived from the hepatitis C virus. The recombinant product selectively increases IgG antibody responses while reducing IgE and is currently in phase 3 trials.

In line with the experiments of Robert Cooke almost 100 years ago involving the passive transfer of immunity to ragweed by use of post-immunotherapy serum injected intradermally, an alternative approach is a passive immunotherapy by administration of recombinant antibodies of high affinity directed against major allergens. This is likely to be more effective where there is one dominant major allergen, as is the case for cat allergy (Fel d 1). In a recent study, a single injection of a combination of two high-affinity recombinant IgG4 antibodies directed against Fel d 1 (single dose 600 mg) was highly effective at suppressing the immediate response to nasal allergen challenge with whole-cat extract. In contrast to active immunization, the limitation of passive immunotherapy is that there is no possibility of durable long-term tolerance.

Another approach is the combination of allergen products with biologics that favorably modify the immune response to the allergen to enhance safety and/or favor induction of preferential tolerogenic responses. Anti-IgE (Xolair) combined with ragweed subcutaneous immunotherapy reduced systemic allergic reactions during a rush-updosing protocol by greater than 80%. Other examples include the combination of subcutaneous immunotherapy with Toll-like receptor agonists (TLR4 and TLR9 TLR2 and TLR7[42] agonists). The combination of allergen immunotherapy with monoclonal antibodies directed against Th2 cytokine pathways (for example, dupilumab targeting IL-4 receptor-alpha) or alternatively with antibodies directed against epithelial cytokines (tezepelumab, targeting TSLP). Such "allergen+" strategies may have the potential to improve efficacy and safety, and possibly enable a shorter, more efficacious course of immunotherapy and augment induction of long-term tolerance.

ON THE HORIZON

- Allergoids (chemical modified allergens)
- Allergens +Immunostimulants (TLR-4, TLR-7, TLR-9 agonists)
- Allergens +monoclonal antibodies (anti-IgE, anti-IL-4, anti-TSLP)
- Recombinant allergens (birch pollen, grass pollen, cat dander)
- Alternative routes (sublingual, intra-lymphatic, epicutaneous)
- Synthetic T-cell peptides (cat dander, house dust mites)
- Recombinant B-cell peptides (grass pollen)
- Medium-length hydrolyzed peptides (grass pollen)
- Passive immunotherapy (anti-Fel d 1 monoclonal antibodies)

REFERENCES

1. Bousquet J, Lockey R, Malling HJ. Allergen immunotherapy: therapeutic vaccines for allergic diseases. A WHO position paper. *J Allergy Clin Immunol*. 1998;102(4 Pt 1):58–62.
2. Noon L. Prophylactic inoculation against Hay Fever. *Lancet*. 1911;1:1572–1573.
3. Prausnitz C, Küstner H. Studien über die Überempfindlichkeit. *Zentralbl Bakteriol Parasitenkol Infektionskt*. 1921;86:160–169.
4. Johansson SGO, Bennich H. Immunological studies of an atypical myeloma protein. *J Immunol*. 1967;98:381–394.
5. Ishizaka K, Ishizaka T. Human reaginic antibodies and Immunoglobulin E. *J Allergy*. 1968;42:330–363.
6. Cooke RA, Barnard JH, Hebald S, Stull A, Cooke RA, et al. Serological evidence of immunity with coexisting sensitization in a type of human allergy (hay fever). *J Exp Med*. 1935;30(62):733–750.
7. Loveless MH. Immunological Studies of Pollinosis: I. The Presence of Two Antibodies Related to the Same Pollen-Antigen in the Serum of Treated Hay-Fever Patients. *J Immunol*. 1940;38:25–50.
8. Frankland AW, Augustin R. Prophylaxis of summer hay fever, a controlled trial comparing crude grass pollen extract with the isolated main protein component. *Lancet*. 1954;1:1055–1057.

9. Lowell F, Franklin W. A double blind study of the effectiveness and specificity of injection therapy in ragweed hay fever. *N Engl J Med*. 1965;273:675–679.
10. Norman PS, Lichtenstein LM, Tignall J. The clinical and immunologic specificity of immunotherapy. *J Allergy Clin Immunol*. 1978;61:370–377.
11. Warner JO, Price JF, Soothill JF, Hey EN. Controlled trial of hyposensitisation to Dermatophagoides pteronyssinus in children with asthma. *Lancet*. 1978;2(8096):912–915.
12. Bousquet J, Van Cauwenberge P, Khaltaev N, et al. Allergic rhinitis and its impact on asthma. *J Allergy Clin Immunol*. 2001;108(suppl 5):S147–S334.
13. Scadding GK, Kariyawasam HH, Scadding G, Mirakian R, Buckley RJ, Dixon T, et al. BSACI guideline for the diagnosis and management of allergic and non-allergic rhinitis. *Clin Exp Allergy*. 2017;47:856–889.
14. Roberts G, Pfaar O, Akdis CA, Ansotegui IJ, Durham SR, Gerth van Wijk R, et al. EAACI Guidelines on Allergen Immunotherapy: Allergic rhino conjunctivitis. *Allergy*. 2018;73:765–798.
15. Agache I, Lau S, Akdis CA, Smolinska S, Bonini M, Cavkaytar O, et al. EAACI Guidelines on Allergen Immunotherapy: House dust mite-driven allergic asthma. *Allergy*. 2019;74:855–873.
16. Tam HH, Calderon MA, Manikam L, Nankervis H, Núñez IG, Williams HC, et al. Specific allergen immunotherapy for the treatment of atopic eczema: a Cochrane systematic review. *Allergy*. 2016;71:1345–1356.
17. Dhami S, Nurmatov U, Arasi S, Khan T, Asaria M, Zaman H, et al. Allergen immunotherapy for allergic rhino conjunctivitis: A systematic review and meta-analysis. *Allergy*. 2017;72:1597–1631.
18. Frew AJ, Powell RJ, Corrigan CJ, Durham SR. Efficacy and safety of specific immunotherapy with SQ allergen extract in treatment-resistant seasonal allergic rhino conjunctivitis. *J Allergy Clin Immunol*. 2006;117:319–325.
19. Demoly P, Emminger W, Rehm D, Backer V, Tommerup L, Kleine-Tebbe J. Effective treatment of house dust mite-induced allergic rhinitis with 2 doses of the SQ HDM SLIT-tablet: Results from a randomized, double-blind, placebo-controlled phase III trial. *J Allergy Clin Immunol*. 2016;137:444–451.
20. Dhami S, Kakourou A, Asamoah F, Agache I, Lau S, Jutel M, et al. Allergen immunotherapy for allergic asthma: A systematic review and meta-analysis. *Allergy*. 2017;72:1825–1848.
21. Virchow JC, Backer V, Kuna P, Prieto L, Nolte H, Villesen HH, et al. Efficacy of a House Dust Mite Sublingual Allergen Immunotherapy Tablet in Adults with Allergic Asthma: A Randomized Clinical Trial. *JAMA*. 2016;315:1715–1725.
22. Duff Hogan A, Bernstein JA. GINA updated 2019: Landmark changes recommended for asthma management. *Ann Allergy Asthma Immunol*. 2020;124:311–313.
23. Penagos M, Eifan AO, Durham SR, Scadding GW. Duration of Allergen Immunotherapy for Long-Term Efficacy in Allergic Rhinoconjunctivitis. *Curr Treat Options Allergy*. 2018;5:275–290.
24. Durham SR, Walker SM, Varga EM, Jacobson MR, O'Brien F, Noble W. Long-term clinical efficacy of grass-pollen immunotherapy. *N Engl J Med*. 1999;341:468–475.
25. Durham SR, Penagos M. Sublingual or subcutaneous immunotherapy for allergic rhinitis? *J Allergy Clin Immunol*. 2016;137:339–349.
26. Scadding GW, Calderon MA, Shamji MH, Eifan AO, Penagos M, Dumitru F, et al. Immune Tolerance Network GRASS Study Team. Effect of 2 Years of Treatment with Sublingual Grass Pollen Immunotherapy on Nasal Response to Allergen Challenge at 3 Years Among Patients with Moderate to Severe Seasonal Allergic Rhinitis: The GRASS Randomized Clinical Trial. *JAMA*. 2017;317:615–625.
27. Bernstein DL, Wanner M, Borish L, Liss GM. Twelve-year survey of fatal reactions to allergen injections and skin testing: 1990-2001. *J Allergy Clin Immunol*. 2004;113:1129–1136.
28. Canonica GW, Bousquet J, Casale T, Lockey RF, Baena-Cagnani CE, et al. Sub-lingual immunotherapy: World Allergy Organization Position Paper 2009. *Allergy*. 2009;64(Suppl. 91):1–59.
29. Cox LS, Larenas-Linnemann D, Nolte H, Weldon D, Finegold I, Nelson HS. Sublingual immunotherapy: a comprehensive review. *J Allergy Clin Immunol*. 2006;117:1021–1035.
30. Sahiner UM, Durham SR. Hymenoptera Venom Allergy: How Does Venom Immunotherapy Prevent Anaphylaxis from Bee and Wasp Stings? *Front Immunol*. 2019;21:10 1959.
31. Sturm GJ, Varga EM, Roberts G, Mosbech H, Bilò MB, Akdis CA, et al. EAACI guidelines on allergen immunotherapy: Hymenoptera venom allergy. *Allergy*. 2018;73:744–764.
32. Vickery BP, Vereda A, Casale TB, Beyer K, du Toit G, Hourihane JO, et al. AR101 Oral Immunotherapy for Peanut Allergy (PALISADE). *N Engl J Med*. 2018;379:1991–2001.
33. Chu DK, Wood RA, French S, Fiocchi A, Jordana M, Waserman S, et al. Oral immunotherapy for peanut allergy (PACE): a systematic review and meta-analysis of efficacy and safety. *Lancet*. 2019;393:2222–2232.
34. Chinthrajah RS, Purington N, Andorf S, Long A, O'Laughlin K. Lyu nali SC et al., Sustained outcomes in oral immunotherapy for peanut allergy (POISED study): a large, randomised, double-blind, placebo-controlled, phase 2 study. *Lancet*. 2019;394:1437–1449.
35. Fleischer DM, Greenhawt M, Sussman G, Bégin P, Nowak-Wegrzyn A, Petroni D, et al. Effect of Epicutaneous Immunotherapy vs Placebo on Reaction to Peanut Protein Ingestion Among Children with Peanut Allergy: The PEPITES Randomized Clinical Trial. *JAMA*. 2019;321:946–955.
36. Shamji MH, Durham SR. Mechanisms of allergen immunotherapy for inhaled allergens and predictive biomarkers. *J Allergy Clin Immunol*. 2017;140:485–498.
37. Sharif H, Acharya S, Dhondalay GKR, Varricchi G, Krasner-Macleod S, Laisuan W, et al. Altered chromatin landscape in circulating T follicular helper and regulatory cells following grass pollen subcutaneous and sublingual immunotherapy. *J Allergy Clin Immunol*. 2020 Nov 6 https://doi.org/10.1016/j.jaci.2020.10.035. S0091–6749(20)31561-X.
38. Shamji MH, Layhadi JA, Achkova D, Kouser L, Perera-Webb A, Couto-Francisco NC, et al. Role of IL-35 in sublingual allergen immunotherapy. *J Allergy Clin Immunol*. 2019;143:1131–1142.
39. Shamji MH, Kappen J, Abubakar-Waziri H, Zhang J, Steveling E, Watchman S, et al. Nasal allergen-neutralizing IgG4 antibodies block IgE-mediated responses: Novel biomarker of subcutaneous grass pollen immunotherapy. *J Allergy Clin Immunol*. 2019;143:1067–1076.
40. Shamji MH, Kappen JH, Akdis M, Jensen-Jarolim E, Knol EF, Kleine-Tebbe J, et al. Biomarkers for monitoring clinical efficacy of allergen immunotherapy for allergic rhinoconjunctivitis and allergic asthma: an EAACI Position Paper. *Allergy*. 2017;72:1156–1173.
41. Gunawardana NC, Durham SR. New approaches to allergen immunotherapy. *Ann Allergy Asthma Immunol*. 2018;121:293–305.
42. Kirtland ME, Tsitoura DC, Durham SR, Shamji MH. Toll-Like Receptor Agonists as Adjuvants for Allergen Immunotherapy. *Front Immunol*. 2020;11:599083.

89

Solid Organ Transplantation: Rejection, Immunosuppression, and Tolerance

Elinor C. Mannon, Kathryn J. Wood, and Roslyn B. Mannon

The clinical era of transplantation began on December 23, 1954, when Dr. Joseph Murray and colleagues performed the first successful renal transplant on the genetically identical Herrick twins. Solid organ transplantation (SOT) has since transformed the management of end-stage organ failure, as a lifesaving technique but also improving the quality of life. The development of powerful immunosuppressive regimens and cutting-edge biologic agents represents an elegant proof of concept, translating seminal work from the laboratory bench to the patient's bedside. However, there are many challenges to overcome. In addition to the major insufficiency in organ supply for the many patients waiting for transplantation, there are immunologic challenges including cellular and antibody-mediated rejection that require all recipients to be maintained on a regimen of immunosuppressive therapy for the life span of the organ. Such global immunosuppression is associated with toxicities including infection, malignancy, and cardiovascular risks that may result in fatal events. Judging the adequacy of immunosuppression is a balance, and the tools to measure sufficiency and our understanding of that process are elusive and complicated by donor–host genetics. In this chapter, we review mechanisms of host immune response to an organ transplant, discuss current and new agents for immunosuppression, insights into immunemodulation in terms of transplant tolerance, and the potential for pig organs to mitigate the demand on human donors, and emphasize the contributions of clinical trials and recent studies.

THE IMMUNE RESPONSE IN ORGAN TRANSPLANTATION

The immune system must have the ability to distinguish "self" from "non-self" or "altered-self" to avoid damaging the host. Any immune response that is generated must also be proportional to the threat; thus, antigens encountered in the context of inflammation will prime T cells and evoke a more aggressive immune response. While essential to survival, this is a major barrier to successful transplantation. Specific signals of "danger" may evoke an immune response, are nonspecific, and are managed by the innate immune response (Fig. 89.1). In contrast, allorecognition (that is, the recognition of a specific protein or antigen on a donor tissue that is different from the host or recipient) is specific and results in rejection. In animal models, transplantation between genetically identical individuals (isograft, syngeneic graft) does not result in rejection, while transplantation between disparate donors and hosts of the same species (allograft) results in an aggressive immune response. The transplantation of organs across species (xenograft) is receiving more focus, as barriers such as zoonosis and gene editing may facilitate early acceptance; this is discussed later.

Innate Immunity and Ischemia Reperfusion Injury

Deceased donor brain death leads to systemic inflammation, leading to hemodynamic responses and subsequent organ injury, whereas donation after cardiac death is associated with warm ischemic injury. At organ retrieval, organs are perfused in an attempt to mitigate the reactive oxygen species and aberrant metabolism associated with brain death.[1] Perfusion solutions have been developed to reduce biochemical injury, and coupled with cold storage, are used to reduce the injury associated with reperfusion following implantation. More recently, ex vivo approaches for organ preservation have emphasized metabolic reprogramming to improve organ function prior to implantation of multiple organs.[2] Moreover, normothermic (i.e., physiologic body temperature) oxygenated perfusion is under study to better "resuscitate" organs through a more physiologic environment. These approaches are crucial to optimizing the limited donor organ pool available.

Innate immunity plays a key role in ischemic injury, resulting in release of damage-associated molecular patterns (DAMPs) like heat shock proteins (HSPs), nucleic acids, and high-mobility group box-1 (HMGB1) protein. These molecules are recognized by invariant pattern-recognition receptors (PRRs) of the innate immune system; for example, Toll-like receptors (TLRs) (see Chapter 3). Release of inflammatory mediators, such as interleukin-1 (IL-1) and IL-6 chemokines, results in adhesion molecule upregulation that expands the immune response. This further triggers activation of macrophages and dendritic cells (DCs), resulting in greater antigen-presenting capacity and entry to a cytocidal state. Endogenous signals can also activate the complement cascade, which promotes DC maturation and subsequently their ability to activate T cells. Thus, the innate immune system, activated by local tissue injury, promotes the initiation of adaptive immune responses when there are antigenic differences between the donor and the recipient. Activation of the adaptive immune system results in a series of effector mechanisms, both cell- and antibody-mediated, which lead to further graft injury (see Fig. 89.1).

Clinical Implications

Implantation of an allograft with severe damage may result in primary non-function. In the kidney, lack of function necessitating dialysis treatments in the recipient is known as delayed allograft function (DGF),[3] occurring in about 30% of all transplanted kidneys. Such poorly functioning allografts may eventually recover

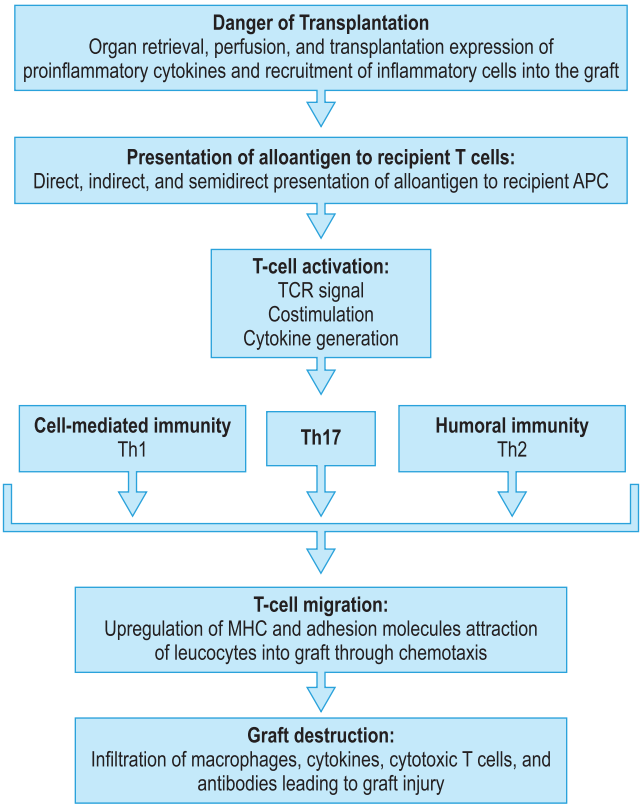


FIG. 89.1 Mechanisms Leading to Graft Injury. After the trauma of surgery ("danger of transplantation"), there is an interplay of innate and adaptive immune responses resulting in eventual graft destruction.

but are associated with inferior transplant outcomes, including more frequent rejection episodes and worse baseline function. While preclinical models of ischemia reperfusion have suggested potential clinical targets for therapy, the results of recent clinical trials with novel agents have been disappointing. The difficulty in translating therapy may relate to the complexity of organ allocation as well as clinical trial design. Understanding ischemia-related dysfunction remains an unmet need in SOT.

Activation of the Adaptive Immune Response

Recognition by T cells of differences between donor and recipient major and/or minor histocompatibility antigens is central to the adaptive immune response. The extensive polymorphism of major histocompatibility complex (MHC) genes (see Chapter 5) makes complete matching of unrelated individuals rare, and, almost inevitably, transplantation of cells or organs between genetically unrelated individuals occurs across an MHC mismatch. Knowledge of MHC structure and molecules has contributed significantly to our understanding of how rejection is triggered and facilitated the development of immunosuppressive drugs. While these differences may be overcome initially by the potency of immunosuppressive therapy, mismatch in class II DR and DQ antigens is associated with graft failure.[4]

Signal 1: Recognition of Alloantigen

Class I MHC molecules are cell surface glycoproteins expressed on most nucleated cells and are recognized by CD8 T-cell receptors (TCRs). Class II MHC molecules are not expressed by every cell in the body; rather, they are found on DCs, B lymphocytes, macrophages, and, in humans, endothelial cells. MHC class II molecules are recognized by T cells bearing the CD4 TCR. Expression of both class I and II MHC molecules can be induced during inflammation of ischemia or rejection, particularly by interferon-gamma (IFN-γ). The inflammatory response triggered in the allograft by retrieval and implantation of the organ initiates not only the migration of donor-derived passenger leukocytes but also their maturation into functional APC expressing high levels of donor MHC molecules.

Antigen processing within antigen-presenting cells (APCs) results in the production of peptides that can bind in these grooves, producing MHC–peptide complexes that are recognized by T cells (see Chapter 6). These peptides may be of self-origin or derived from foreign molecules (e.g., from an allograft after transplantation or from a virus after an infection). In general, peptides derived from molecules within the cell are processed and loaded into MHC class I molecules, whereas extracellular molecules present outside the cell are processed into peptides that load into class II molecules. However, cross-presentation can also occur.

As discussed in Chapter 4, the TCR is composed of two chains that confer MHC–peptide specificity and is associated with a complex of polypeptides referred to collectively as CD3. When the TCR of host naïve or memory T cells engages its specific antigen, CD3 delivers intracellular signals to the T cell (Chapter 10). This is the first step in T-cell activation commonly referred to as "signal 1" and is believed to take place in secondary lymphoid organs rather than in the transplanted graft itself. In clinical transplantation, treatment with calcineurin inhibitors (CNIs) to block signal 1 and/or the use of T cell–depleting agents has been highly successful in the prevention and reversal of rejection episodes.

Antigen presentation may occur in one of three pathways (Fig. 89.2). Intact allogeneic MHC molecules presented by donor-derived passenger leukocytes to host T cells is known as the *direct pathway* of allorecognition and this is the dominant pathway through which the immune response to the graft is initiated. T cells responding via direct antigen presentation constitute a vast majority of the alloreactive immune repertoire, estimated at 10% of T cells.[5] Over time, when donor APCs within the graft are depleted, the *indirect pathway* of allorecognition dominates. Here, the recipient APCs largely consisting of DCs and B cells process and present peptides derived from allogeneic MHC molecules shed from the graft (both soluble MHC molecules and apoptotic cells), as well as minor histocompatibility antigens. Experimental models suggest that indirect presentation of donor antigens may play a greater role in rejection than direct presentation overall, as it is a continuous process as long as the graft remains in situ. Finally, a third pathway referred to as the *semidirect pathway* of antigen presentation exists in which donor MHC proteins are transferred intact to recipient APCs (through membrane transfer or the exosomal route), enabling them to present allogeneic MHC–peptide complexes to host T cells. This MHC transfer is temperature and energy dependent and requires close cell-to-cell contact. Both MHC class I and class II may be transferred, although class II MHC appears to be transferred more efficiently.

Signal 2: Costimulation

T-lymphocyte activation also requires signals delivered by the interaction of several costimulatory receptors and their ligands, known collectively as "signal 2" (Fig. 89.3). During T-cell activation, the TCR–CD3 complex and costimulatory molecules are brought together in the cell membrane to form the im-

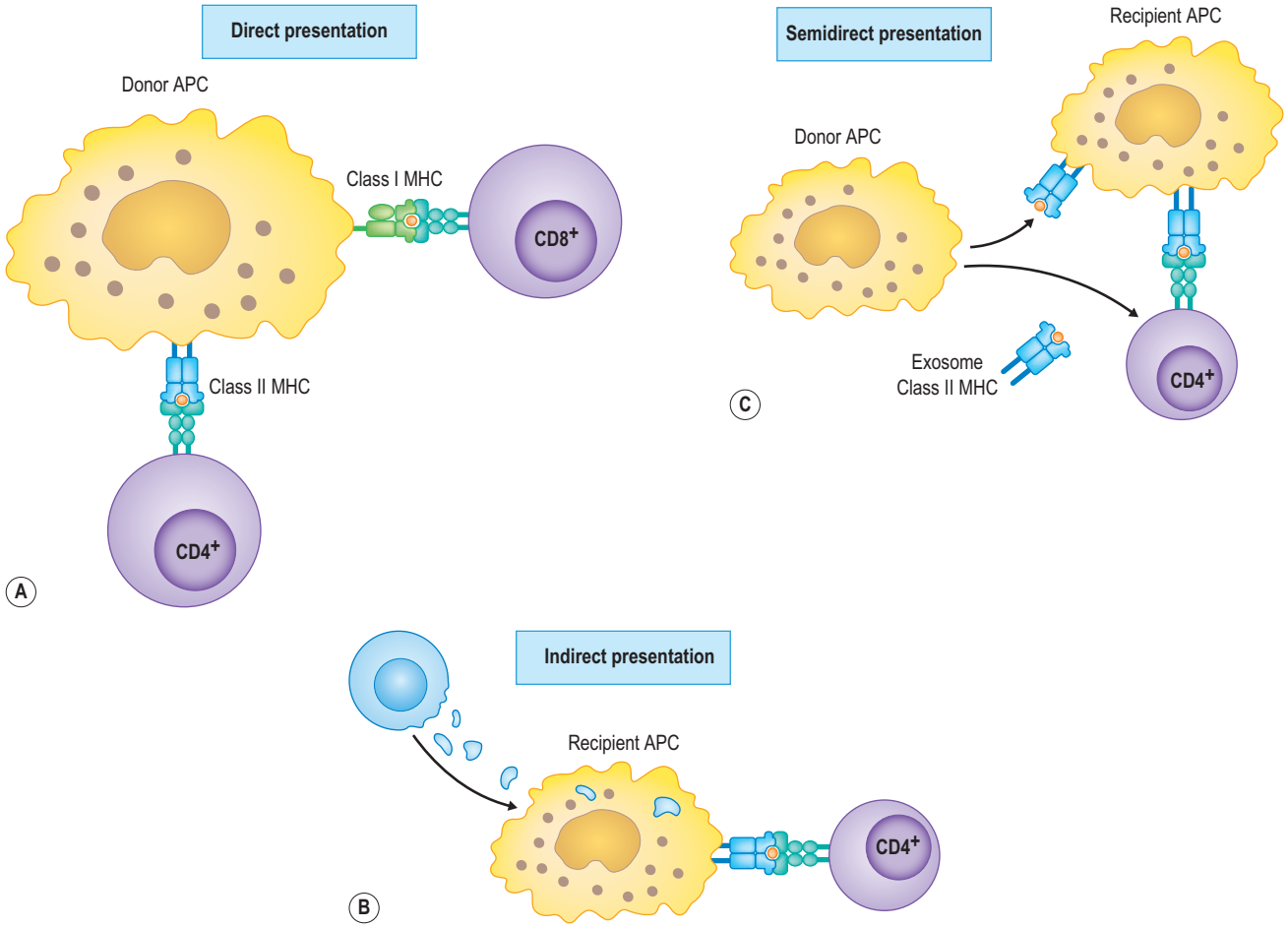


FIG. 89.2 Antigen Presentation. (A) Direct presentation: passenger donor antigen-presenting cell (*APC*) presents alloantigen to recipient T cells in lymphoid tissue. (B) Indirect presentation: alloantigen from donor cells is processed and presented by recipient APC via major histocompatibility complex (*MHC*) class II to recipient $CD4^+$ cell. (C) Semidirect presentation: donor MHC class I and class II may be transferred to the surface of recipient APCs, enabling presentation of alloantigen to recipient T cells.

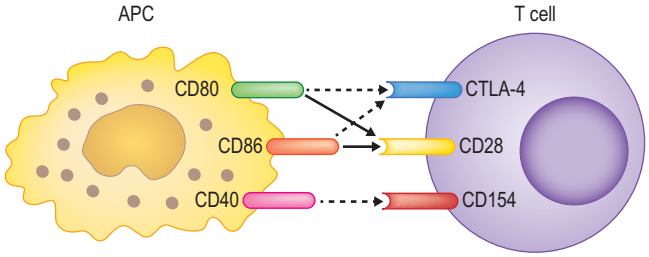


FIG. 89.3 Costimulatory Signal 2. Costimulation follows binding to major histocompatibility complex (MHC) class I and class II on antigen-presenting cells (*APCs*) and involves signal transduction via intracellular proteins and increased interaction affinity via the binding of several cell surface proteins. Several costimulatory molecule pairs have been identified, both activating and inhibitory.

munologic synapse. Signal 1 is specific for the antigen involved, whereas costimulation is not antigen-specific.

Costimulatory receptors fall into two major families; the B7 family (e.g., CD28 and CD152) and the tumor necrosis factor (TNF) family (e.g., CD154 and CD70) (see Chapter 10). Although several costimulatory molecule pairs have been identified,[6] the CD28 and CD154 pathways have been the most clearly defined in transplantation. CD80 and CD86 ligands on the surface of DCs and other cells can bind to CD28 on T cells, resulting in the activation of additional signal–transduction pathways within the T cell. This has a number of effects: lowering the threshold for T-cell activation; increasing glucose metabolism, as well as cytokine and chemokine expression, including IL-2 production; reducing T-cell death through apoptosis; and expanding the number of T cells that respond through proliferation. CD80 and CD86 are also linked to an inhibitory receptor, CD152 or cytotoxic T-lymphocyte antigen-4 (CTLA-4), which inhibits T-cell activation possibly by competing with CD28 to bind with CD80 or CD86. CTLA-4 has a 10-fold higher receptor affinity compared with CD28 and is rapidly upregulated following T-cell activation. Additionally, CTLA-4 activates tryptophan catabolism in DCs, resulting in the inhibition of proliferation and promoting apoptosis of the responding T cells. This pathway is the mechanism behind belatacept, CTLA4-Ig, a biologic used in transplantation to mitigate rejection (as discussed later in this chapter).

CD40, a member of the TNF receptor family, is expressed on all APCs and binds to CD154 (CD40L), which is present on activated CD4 cells, a subset of CD8 cells and Natural Killer (NK) cells. CD40 stimulation causes triggering signals for antibody production and induces MHC expression on APCs, thus amplifying antigen presentation. In preclinical models, treatment with

anti-CD154 promoted long-term survival of heart allografts in mice and of kidney allografts in non-human primates (NHPs).[7] The clinical efficacy of costimulatory blockade is discussed further in the Immunosuppression section.

Signal 3: T-Cell Proliferation and Differentiation

Following signals 1 and 2, a complex process known as "signal 3" begins. This signal includes the induction of three signal–transduction pathways: the calcium–calcineurin pathway, the Ras- and Rac-mitogen–activated protein kinase pathways, and the protein kinase C nuclear factor κB (NF-κB) pathway (see Chapter 10). These pathways interact with inositol triphosphate (IP3) and diacylglycerol (DAG), formed from the hydrolysis of phosphatidylinositol 4,5-bisphosphonate, to activate 3 transcription factors: nuclear factor of activated T cells (NFAT), activating protein-1 (AP-1), and NF-κB, respectively. These transcription factors cause the expression of many genes leading to upregulation of growth factors and cytokines; in particular, IL-2, a key T-cell growth factor, and CD25 (IL-2Rα). Growth signals are then delivered through the phosphoinositide-3-kinase (PI3K) and mammalian target of rapamycin (mTOR) pathways to promote cell cycle progression, and the clonal expansion and differentiation of activated T cells are initiated leading to their effector functions. These mechanisms are targeted by various immunosuppressive agents discussed in the immunosuppression section and seen in Table 89.1.

The pattern of cytokine production through the interaction of the combination of signaling processes determines the nature of the response, in which either cell-mediated or antibody-mediated immunity is seen to dominate (see Chapter 11). T-helper 1 (Th1) cells are T cells known to promote a cell-mediated response and are associated with the production of IFN-γ.

TABLE 89.1 Immunosuppressive Therapies in Transplantation: Maintenance Induction and Adjuvant Treatments

Drug	Mechanisms	Adverse Effects
Maintenance Agents		
Corticosteroids	Induces phospholipase A_2 inhibitory proteins, inhibits arachidonic acid synthesis, inhibits prostaglandins and leukotrienes	Diabetes, delayed wound healing, peptic ulcers, psychosis, osteoporosis, infection, blurred vision, fluid retention, weight gain, acne, constipation
Azathioprine	Inhibits purine and DNA synthesis, inhibits cell proliferation	Bone marrow depression, opportunistic infection, macrocytosis, liver toxicity
Mycophenolate mofetil	Inhibits inosine-monophosphate dehydrogenase, inhibits purine synthesis. and blocks cell proliferation	Gastrointestinal symptoms, bone marrow depression, opportunistic infection, in particular CMV and BK nephropathy
Cyclosporine	Binds to cyclophilin, inhibits calcineurin phosphatase, blocks NFAT dephosphorylation, blocks IL-2 transcription and T-cell activation	Hypertension, hyperlipidemia, nephrotoxicity, hepatotoxicity, pancreatitis, peptic ulcers, thrombotic microangiopathy, opportunistic infection, neurotoxicity, tremor, gingival hyperplasia, hirsutism
Tacrolimus	Binds to FKBP12, inhibits calcineurin phosphatase and blocks T-cell activation	Posttransplantation diabetes mellitus, nephrotoxicity, thrombotic microangiopathy, neurotoxicity
Rapamycin Everolimus	Binds to FKBP12, inhibits mTOR, and blocks IL-2–driven cell proliferation	Delayed graft function, delayed wound healing, mouth ulcers, pneumonitis, increased proteinuria, peripheral edema, hyperlipidemia
Belatacept	Binds CD80/86 with higher affinity than CD28 blocking signal 2 in T cells	Headache, anemia PTLD, opportunistic infection
Induction Agents		
Antithymocyte globulin	Polyclonal effects not well characterized; immunosuppressive efficacy attributed to T-cell depletion through apoptosis, antibody-dependent cytolysis, and complement-dependent lysis	Polyclonal effects: cytokine release syndrome, serum sickness, leukopenia, thrombocytopenia. De novo tumors and opportunistic infection: CMV and HSV.
Alemtuzumab	Binds to CD52 antigen, (expressed on 95% of peripheral blood lymphocytes, NK cells, macrophages, and thymocytes) Results in profound lymphopenia	Opportunistic infections: *Candida*, CMV
Basiliximab	Binds to IL-2R with similar affinity as IL-2, thereby inhibiting IL-2–driven T-cell proliferation Abrogates Signal 3	Occasional hypersensitivity reactions, inadequate immunosuppression in immunologically high-risk recipients
Adjunctive Treatments		
Intravenous Immune Globulin	Inhibition of antibody production, inhibition of B-cell differentiation, inhibition of production of interleukin-6 and tumor necrosis factor-α, induction of B-cell apoptosis, inhibits complement activation, saturates FcRn to accelerate the breakdown of endogenous alloantibody IgG, anti-idiotypic antibody blocking alloantibody function	Hemolysis, headache, acute renal failure with sucrose-containing preparations, thromboembolic events
Rituximab	Binds CD20 with cellular destruction by ADCC	Infusion side effects (headache, nausea), hypogammaglobulinemia
Eculizumab	Binds to the terminal complement component 5 limiting the production of C5b and membrane-activating complexes	Meningococcal infections, *Neisseria* infections
Proteasome inhibitors	Bind the 26 S proteasome inhibiting proper ubiquitination and protein degradation leading to cell dysfunction and apoptosis	Neuropathy, diarrhea, headache, cytopenia

ADCC, antibody-dependent cytotoxicity; *CMV,* cytomegalovirus; *DHFR,* dihydrofolate reductase; *HSV,* herpes simplex virus; *IL-2,* interleukin-2; *IL-2R,* interleukin-2 receptor; *mTOR,* mammalian target of rapamycin; *NFAT,* nuclear factor of activated T cells; *PTLD,* post-transplant lymphoproliferative disease.

T cells promoting a humoral response (i.e., Th2 cells) are associated with the generation of IL-4, −5, and −6. Additionally, a Th17 population that has been identified is characterized by the production of IL-17 and promotion of the infiltration of neutrophils. Th22 cells, which express IL-13, IL-22, and TNF-α, have also been described. A subset of CD4 cells called *regulatory T cells* (Treg) can also be induced following antigen exposure in the periphery (pTregs; see Chapter 13). These cells secrete IL-10 or transforming growth factor-beta (TGF-β) and have suppressive or regulatory functions against effector cells and APCs. Skewing the immune response in favor of Treg is being investigated as a potential strategy to improve long-term graft survival or through adoptive transfer.[8] The balance of these responses results in either graft injury or the induction of tolerance.

Following exposure to an antigen, antigen-specific memory T and B cells are generated. These memory cells are then able to produce a more rapid and intense immune reaction if the antigen is encountered on a second occasion because they have a lower activation threshold and are less dependent on costimulation. Transplant recipients, particularly older patients or those who have had previous antigen exposure through previous transplantation, blood transfusion, or pregnancy, may therefore have specific anti-donor memory cells. Memory-type responses may also occur due to antigen receptor cross-reactivity known as *heterologous immunity*. Moreover, T-effector memory cells are resistant to lymphocyte depletion therapies and $CD28^+CD8^+CD45RA^+CCR7^-$ effector memory cells are resistant to costimulatory blockade.[9]

HOW ARE GRAFTS DESTROYED: THE HOST EFFECTOR IMMUNE RESPONSE

In organ transplantation, the type of transplanted tissue, location of implanted organ, and immune status of the recipient at time of transplantation may modify the immune response. Although initiation of rejection in a naïve recipient is principally T-cell dependent, many components of the immune system contribute to the subsequent destruction of the transplanted tissue. Graft destruction may be alloantigen-specific, or may be due to bystander tissue destruction.

Acute T Cell–Mediated Rejection

As described above, following the innate response to organ acquisition, implantation, and reperfusion, the inflammatory environment within the graft promotes an adaptive cellular response to the graft itself. Naïve cytotoxic T cells, activated by CD4 cells clustering with APCs, migrate to the graft, where they recognize allogeneic class I MHC molecules. These cytotoxic cells release key molecules such as perforin and granzyme B, upregulate surface Fas ligand, and secrete soluble mediators, such as TNF-α. Perforins insert into the target cell membrane to form pores, allowing granzyme to enter the cell, causing proteolysis and activation of the apoptotic caspase cascade. Moreover, Fas ligand binds to Fas on target cells, similarly inducing apoptosis. While the clinical risk of acute T cell–mediated rejection (TCMR) occurs in the first months of transplantation, TCMR may be detected at any time following transplantation, when there is insufficient immunosuppression.

In animal models, a nonspecific delayed-type hypersensitivity response may occur, usually mediated by CD4 cells that are attracted to the graft, involving the release of multiple proinflammatory cytokines, including IL-1, IFN-γ, and TNF-α. This leads to the further recruitment and activation of leukocytes, increasing graft cell permeability and vascular smooth muscle tone, affecting graft function, and contributing to both acute and/or sustained rejection. In the latter process, CD4 alloreactive T cells responding to donor-derived peptides bound to recipient MHC class II molecules have also been correlated with chronic allograft dysfunction. In human kidney transplant recipients (KTRs), this phenomenon is called chronic active T cell–mediated rejection (CA-TCMR) and has specific histologic criteria (Table 89.2).[10] CA-TCMR is associated with prior episodes of TCMR and later development of allograft fibrosis and tubular atrophy.[11,12] This entity reflects under immunosuppression, but appropriate clinical intervention remains uncertain.

As acute allograft rejection is initiated by the recognition of polymorphic donor MHC molecules by recipient T cells, it may follow that transplantation of MHC incompatible tissues will elicit a strong T cell–dependent immune response to donor tissues, based on extent of mismatches as demonstrated in rodent models. These differences are often not apparent in clinical transplantation in the context of modern immunosuppression. Moreover, rejection may still occur between MHC-matched siblings due to T-cell recognition of minor histocompatibility antigen differences.

From work in animal models and findings in human transplant recipients, so-called "chronic rejection" was a label used to describe functionally failing organ transplants with associated inflammatory cell infiltrates coupled with alloantibody interaction of the vascular endothelium. This interaction resulted in smooth muscle cell activation and proliferation of the arterial medial wall ("vasculopathy"). This entity can be seen in heart and kidney allografts. In the latter, interstitial fibrosis and tubular atrophy are also characteristic of the failing allograft. Indirect allorecognition has been implicated in the development of "chronic rejection," though this is primarily derived from animal models (reviewed in Siu JHY, et al.[13]).

Clinical Implications

Acute graft rejection is suspected when there is a sudden deterioration in allograft function and is confirmed by allograft biopsy. As shown in Table 89.2, there are semi-quantitative measures for the intensity and location of inflammatory response, developed for all the organs.[10] An example of TCMR is shown **in** Fig. 89.4. The extent of rejection grade is associated with the number of cells within the tubular epithelial cell as well as the extent of inflammation in the graft. Vascular invasion is considered more severe and raises the grade of rejection. Typically, the severity of an episode of rejection and the level of graft dysfunction are key factors in terms of responsiveness to therapy. Finally, the frequency of rejection in the first year after transplant varies depending on the organ, ranging from 8% in the kidneys up to 39% in the small bowel,[14] reflecting the immunogenicity of the allograft and effectiveness of immunosuppressive therapy.

Detection of graft dysfunction varies based on organ, and while lab findings are used for kidney, liver, and pancreas transplants, heart transplants undergo surveillance biopsy to monitor for rejection. In kidney transplantation, the finding of rejection in a biopsy without graft dysfunction is called "subclinical" rejection. However, in a recent survey of US transplant centers, less than 25% performed surveillance biopsies even in selected patients, citing the low yield of actionable information

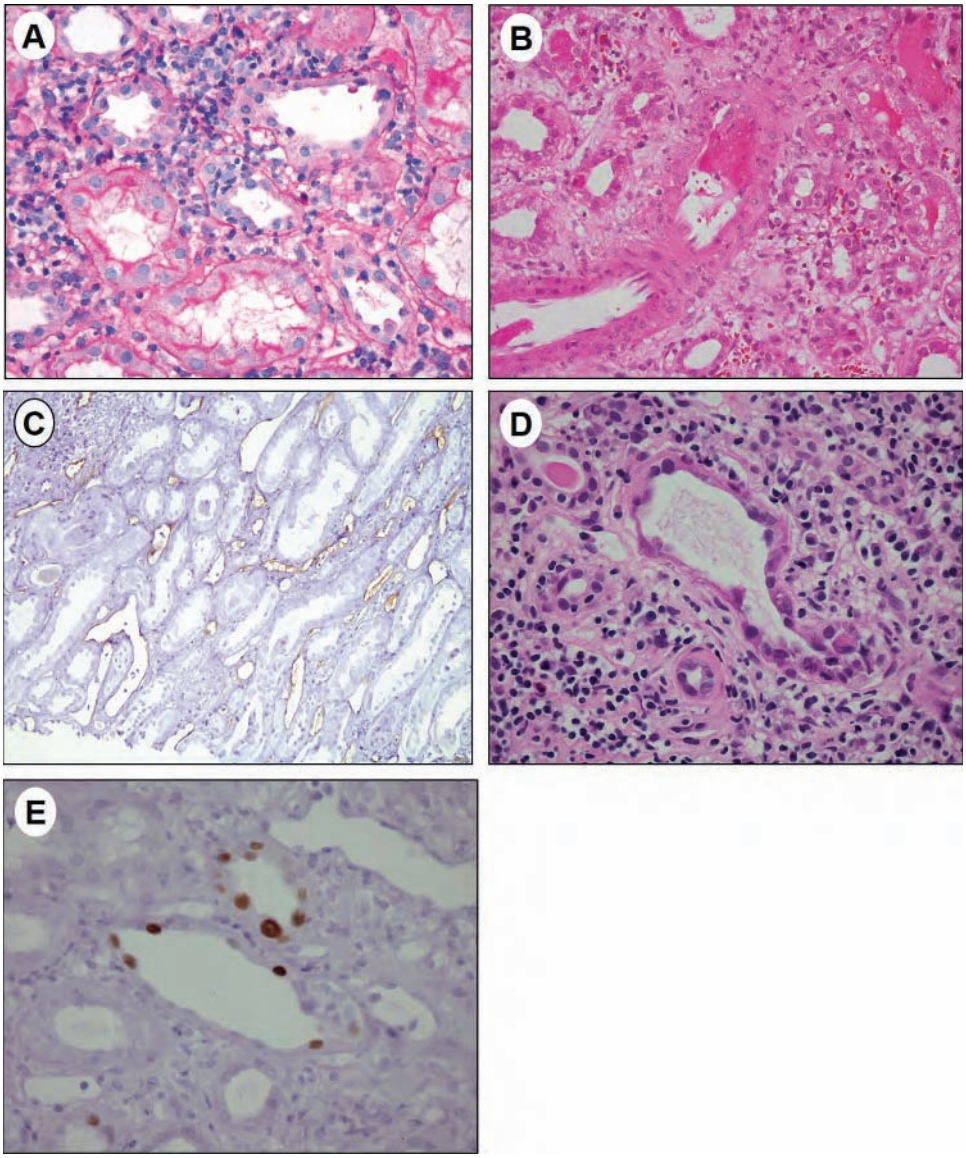


FIG. 89.4 Kidney Biopsies Demonstrating Abnormal Histology. (A) This is an example of Banff 1b T cell–mediated rejection (TCMR). Note that there is substantial interstitial inflammation (i3) but the extent of tubular infiltration by monocytes (tubulitis, t) is critical to define TCMR. (B) Antibody-mediated rejection (AbMR). The key findings are peritubular capillaritis (ptc) accompanying vasculitis (v) and thrombosis. There is also tubulitis present. (C) Immunostaining for complement component C4d, with positive staining of peritubular capillaries. (D) BK polyomavirus nephropathy. There is intense interstitial inflammation (i) and tubulitis with obliteration of the renal tubules; these findings may also be seen in TCMR. However, within the tubular epithelium are viral cytopathic changes including nuclear inclusions in tubular epithelial cells, enlarged irregular nuclei, and chromatin smudging, as well as detached tubular cells with denuded patches of basement membrane. (E) Immunostaining for SV40 T antigen demonstrating positive nuclei in the renal tubule epithelium. This finding is diagnostic for viral infection with the SV40 family of viruses that include BK.

as the primary reason for nonperformance. Moreover, the early detection of subclinical rejection identifies patients at increased risk for allograft failure who might benefit from early interventions that increase immunosuppression exposure. However, the long-term consequences of detecting and treating subclinical inflammation remain controversial, with some arguing that there are insufficient long-term benefits to justify the practice.

With the growing use of immune checkpoint inhibitors (CPIs) in advanced cancer treatment (see Chapter 80), there is increasing recognition of immune-related adverse events that occur in 70%–90% of treated patients.[15] Such events include skin disorders, acute kidney injury with acute interstitial nephritis, gastrointestinal (GI) disorders, hepatotoxicity, and endocrinopathies like thyroiditis and diabetes. The activation of T-cell antitumor activity may also adversely affect allografts in transplant recipients. Further data are needed to better assess the risks to transplanted organs and possible negative impact on patient survival beyond the impact of the tumor itself.

Antibody-Mediated Rejection

Alloantigen-specific antibodies, or alloantibodies, are produced after alloantigen-driven B-cell activation in the presence of T-cell help, which may occur during cellular rejection or following a blood transfusion. In addition to DCs, B cells themselves may

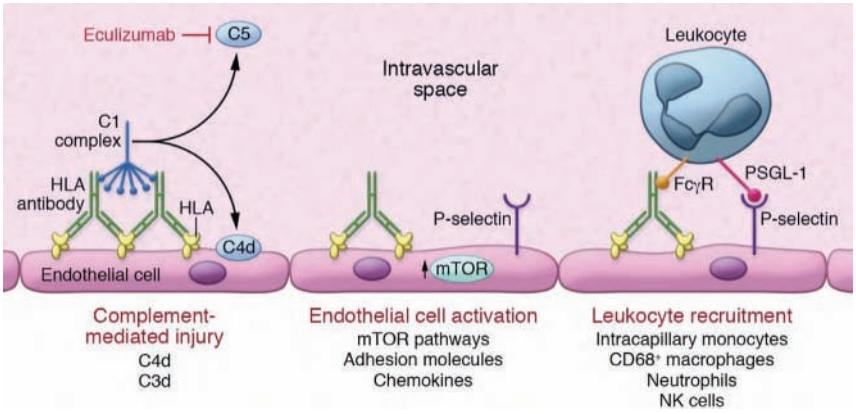


FIG. 89.5 Mechanisms of Antibody-Mediated Injury on the Vascular Endothelium. Once alloantibody is bound, activation may occur via complement–dependent pathway. Typically, non-HLA anti-donor antibodies activate through complement-independent pathways. Finally, bound antibody exposes the Fc portion of the antibody, which may bind and attract leukocytes to the endothelium, adding to the immune-mediated injury by innate immune cells. (Adapted from Valenzuela N, Reed EF. *J Clin Invest* 2017;127(7):2492. Fig. 89.2A.).

act as APCs. MHC class II molecules are presented to and bind immunoglobulins (Igs) on the surface of B cells, enabling the B cell to internalize the alloantigen and process it into peptides that are presented at the cell surface within MHC class II molecules (see Chapter 6). These activated T cells produce cytokines that activate B cells, enabling them to differentiate into alloantibody-producing plasma cells. When allospecific antibodies encounter their specific antigen, antibody-mediated rejection occurs.

There are three mechanisms by which antibody mediates tissue injury in antibody-mediated rejection (AbMR; Fig. 89.5).[16] First, bound antibody activates endothelial cells within the graft, resulting in the expression of adhesion molecules, cytokines, and chemokines, as well as the synthesis of tissue factor. This process results in increased immune cell trafficking to the site, enhancing the immune response. Second, antibody binding may trigger complement activation, which can result in cell lysis and graft damage directly or indirectly by the binding of complement components to the graft and the recruitment of macrophages and neutrophils. Finally, the exposed Fc non-antigen-specific portion of the antibody molecule may serve as a receptor for other innate immune cells, including NK cells and monocytes. This encourages NK cells and macrophages to kill any target cells with antibodies bound to their surface. This is a nonspecific process termed *antibody-dependent cellular cytotoxicity* (ADCC) and can contribute to graft rejection (see Chapter 12).

Antibodies to Human Leukocyte Antigens (HLA) antibodies may be present in all hosts. The diversity and extent of these antibodies are referred to as allosensitization. This occurs as a result of blood transfusions due to contaminating passenger leukocytes in the blood product, pregnancy, and prior organ transplant. While we typically associate alloantibodies forming after sensitizing events, their generation in the context of host infection and heterologous immunity has important clinical implications.[17]

When donor-specific anti-HLA antibodies (DSA) exist in a recipient prior to transplantation, a dramatic response known as hyperacute rejection may occur upon perfusion of the transplanted graft. Coagulation and complement cascades are activated, resulting in extensive thrombosis and graft infarction within minutes. Hyperacute rejection is rarely seen in clinical practice anymore because of the advances in screening for HLA antibodies prior to transplantation (see the following section).

Non-HLA antibodies are increasingly recognized as associated and perhaps causative of antibody-mediated injury.[18] The antigens recognized include the major histocompatibility complex class I chain-related gene A (*MICA*) or B (*MICB*), or structural proteins in the donor organ such as collagen IV in basement membranes, vimentin, or angiotensin 1 receptor. Anti-endothelial-cell antibodies to targets such as ICAM-4 and endoglin have also been identified in pre-kidney transplant sera and associated with post-transplant development of HLA DSA and AbMR, and similarly correlated to poor outcomes in heart transplantation. However, there is no consensus on the routine monitoring for detection of such antibodies, which are clinically assayed when HLA DSA are not detected in the setting of graft dysfunction and AbMR.

Clinical Implications

In the past, the assessment of HLA antibodies, and specifically those that are donor specific, utilized an in vitro ADCC cytotoxicity assay that required donor cells (lymphocytes) and the serum from the intended recipient. DSA are now measured by sensitive flow cytometry–based techniques using HLA-loaded beads and recipient serum, allowing for a more complete profile than older serologic analyses.[19] These sensitive assays are more efficient in busy laboratories but are subject to lab-to-lab differences in their performance, and hence are difficult to standardize across all transplant centers.[20] However, within a particular histocompatibility laboratory, methods and thresholds have been established to define clinically apparent antibodies. In clinical practice, laboratories perform a "virtual" cross-match in which detailed molecular typing of the donor and a detailed antibody profile of the recipient render a determination of HLA compatibility.[19] This improves the time to transplantation and also supports proper organ allocation to avoid those recipients with existing DSAs.

Diagnosis of AbMR requires an allograft biopsy. Specific criteria are shown in Table 89.2, but the recognition of injury was finally galvanized by the ability to detect complement activation within the allograft based on immunostaining of C4d in peritubular capillaries, the site of antibody–endothelial interaction (see Fig. 89.4). The dependence on this criterion has also led to debate about C4d-negative biopsies with otherwise typical findings for AbMR, and also the presence of C4d in the absence of injury findings in the context of

TABLE 89.2 Banff Diagnostic Criteria for Renal Allografts

Category 1: Normal or nonspecific changes

Category 2: Antibody-mediated changes

Active AbMR (requires all 3 criteria)

1. Histologic evidence of acute tissue injury:
   - Microvascular inflammation (g > 0 and/or ptc > 0)
   - Intimal or transmural arteritis (v > 0)[b]
   - ATM
2. Evidence of current/recent antibody interaction with vascular endothelium, including one or more of the following:
   - C4d staining in ptc
   - Moderate microvascular inflammation ([g + ptc] ≥2)
   - Gene transcripts/classifiers of ABMR
3. Serologic evidence of donor-specific antibodies (DSA).

Chronic active ABMR (requires all 3 criteria)

1. Morphologic changes of chronic tissue injury
   - Transplant glomerulopathy (cg > 0)
   - Severe peritubular capillary basement membrane multilayering
   - Arterial intimal fibrosis
2. Same as above
3. Same as above

Chronic Inactive ABMR

1. Same as above
2. Absence of criterion 2 above
3. Prior documented active or CA ABMR and/or DSA

Category 3: Borderline (suspicious for acute TCMR

Foci of tubulitis (t1, t2, or t3) with mild interstitial inflammation **(i1)**, OR mild tubulitis (t1) with moderate to severe interstitial inflammation (i2 or i3)

Category 4: T-cell–mediated rejection (TCMR)

Active TCMR

Grade IA: Interstitial inflammation (i2 or i3) with moderate tubulitis

Grade IB: Interstitial inflammation (i2 or i3) with severe tubulitis (t3)

Grade IIA: Mild to moderate intimal arteritis (v1), with/without interstitial inflammation

Grade IIB: Severe intimal arteritis (v2), with/without interstitial inflammation

Grade III: Transmural arteritis and/or arterial fibrinoid necrosis (v3), with/without interstitial inflammation

Chronic active TCMR (requires all 3 criteria)

Grade IA–Grade IB: Interstitial inflammation involving >25% of sclerotic cortex AND > 25% of total cortical parenchyma (ti2 or ti3) with moderate (t2 or **t-IFTA2**) or severe (t3 or **t-IFTA3**) tubulitis

Grade II: Chronic allograft arteriopathy

Category 5: Polyomavirus nephropathy

Class I: pvl 1 and ci 0–1

Class II: pvl 1 and ci 2–3, or pvl 2 and ci 0–3, or pvl 3 and ci 0–1

Class III: pvl 3 and ci 2–3

Histological criteria have been developed and are regularly reviewed to aid diagnosis of the cause of chronic allograft dysfunction. Determining the cause of dysfunction aids decision making with regard to pathology and management.

AbMR, Antibody-mediated rejection; *CA*, chronic active; *ptc*, peritubular capillaritis; *v*, vasculitis.

For full details, see Loupy A, Haas M, Roufosse C, et al. The Banff 2019 Kidney Meeting Report (I): Updates on and clarification of criteria for T cell- and antibody-mediated rejection. *Am J Transplant*. 2020;20(9):2318–31.

so-called accommodation.[21] This has led to revisions of standardization for AbMR biopsy diagnosis in the kidney and includes the use of gene transcripts associated with AbMR.[10]

With the increasing recognition of de novo development of DSA following transplantation contributing to late allograft failure, in part due to insufficient immunosuppressive therapy for a particular host, there is a growing consensus of the necessity for monitoring for DSA after transplantation to mitigate the contribution of antibody-mediated injury in graft failure.[22]

Late Allograft Failure

Both nonimmune and immune injury contribute to late allograft loss.[23] In the former, key contributors include advanced donor age and cell senescence, delayed graft function (as noted in the previous section), and calcineurin-mediated nephrotoxicity. These injuries involve cell signaling, promoting tissue remodeling, and interstitial fibrosis and tubular atrophy. With regard to immune-mediated injury, the notion of chronic rejection or chronic allograft nephropathy has been replaced with more refined entities, based on the analysis of large cohort studies that have pointed to the contribution of DSA anti-HLA antibodies based on data in large clinical cohorts.[24,25] Such studies and others suggest the presence of immune activation either by intentional nonadherence or immune responses that have evaded the immunosuppressive state over time. This phenomenon is seen in the lung allograft, manifested as chronic lung allograft dysfunction (CLAD),[26] and in the heart transplant, where it is known as cardiac allograft vasculopathy.[27] In both instances, there is injury and inflammation affecting the airways and vasculature, respectively.

Viral-Mediated Kidney Allograft Injury

In the last two decades, a critically important contributor to kidney transplant injury and graft failure as well as to native kidney injury in other solid organ transplants has become known as BK polyomavirus nephropathy. BK virus, a DNA virus related to JC polyomavirus and SV-40, reactivates in the host, eliciting an immune response that histologically can mimic ACR with an inflammatory, interstitial inflammatory response with tubulitis, as well as viral inclusions in tubular epithelial cells[28] associated with kidney dysfunction. This is not a systemic disease, with focal infection in the tissues of the urinary tract. The virus gains entry into urothelium via an N-linked glycoprotein containing a (2,3)-linked sialic acid receptor followed by caveolae-mediated endocytosis. Following internalization into the cell, the virus migrates through the cytoplasm and gains entry into the nucleus, where viral transcription, replication, and particle assembly take place. Innate immune responses, in part via Toll-like receptor 3 (TLR3) that senses viral double-stranded RNA as part of the viral replication, result in induction of IL-6 and IL-8. Small peptide defensins and IFN-γ inhibit viral replication. Viral-specific T-cell responses, activated following DC presentation of viral antigens, are important to control viral infection, whereas antiviral antibodies do not appear to correlate with viral clearance.

Clinical Implications

It is the extent of immunosuppression that is associated with viral reactivation and suppression of normal antiviral cellular responses. As such, clinicians use a proactive strategy to monitor for BK viral DNA in urine and/or serum by polymerase chain reaction (PCR). Once detected, reduction in immunosuppression is undertaken. With worsening kidney function, diagnosis is made by allograft biopsy, which reveals interstitial inflammation, tubulitis, and immunohistochemical staining for SV-40 T-cell antigen, so called polyomavirus replication/load level (pvl) (see Fig. 89.4). With no known effective antiviral therapy, immunosuppression may be reduced further and empiric treatments have included IVIG, conversion to cyclosporine, and corticosteroid therapy. Typically, there is progressive inflammation and fibrosis, and ultimately allograft failure.

IMMUNOSUPPRESSIVE MANAGEMENT

The potency of current immunosuppressive strategies has been associated with dramatic improvements in short-term graft outcomes due to reduced rates of TCMR. But as noted previously, long-term graft survival remains a challenge. Therapy is aimed at T cell–mediated immune responses. The treatment paradigm is induction therapy, followed by maintenance therapy. Induction may be lymphocyte-depleting or not, with the former quite popular due to the potency of this strategy of reducing the bulk of T lymphocytes (monoclonal or polyclonal antibody) at the time of implantation of the allograft. Based on the discussion of allorecognition and T-cell activation, maintenance therapy is a combination of agents to interfere with these processes (Fig. 89.6) and includes corticosteroids, anti-metabolite (typically mycophenolic acid [MPA]), and calcineurin inhibitor (typically tacrolimus). Table 89.1 lists the mechanisms of these agents and, importantly, their off-target effects, with overall concerns of opportunistic infection and malignancy. To achieve the proper level of immunosuppression, clinicians monitor 12-hour trough levels of calcineurin inhibitor (CNI) as well as MPA levels. Note that certain solid organ transplants such as the liver don't typically include induction therapy, and maintenance therapy may include two agents only. In contrast, highly immunogenic simultaneous kidney–pancreas transplants utilize "quadruple therapy" with T-cell depletion as induction. Typically over the first 3 to 6 months maintenance therapy is reduced to lower baseline levels as the risk of TCMR is reduced.

Signal 1: Blockade of Antigen Recognition

Activation of the rejection response to an allograft hinges on the recognition of antigen by the host immune system. Targeting

CLINICAL PEARLS

When graft dysfunction is detected, a biopsy is performed and therapy is based on those findings:

T cell–mediated rejection (i.e., acute cellular rejection or TCMR)
- Increase in baseline therapy doses
- High-dose steroids
- Anti-thymocyte globulin
- Alemtuzumab

Antibody-mediated rejection (AbMR)
- Plasmapheresis
- Intravenous immunoglobulin
- Rituximab

BK virus nephropathy (BKPVN)
- Immunosuppression reduction (Anti-metabolite, CNI)
- Serial monitoring of serum and/or urine viral load by PCR

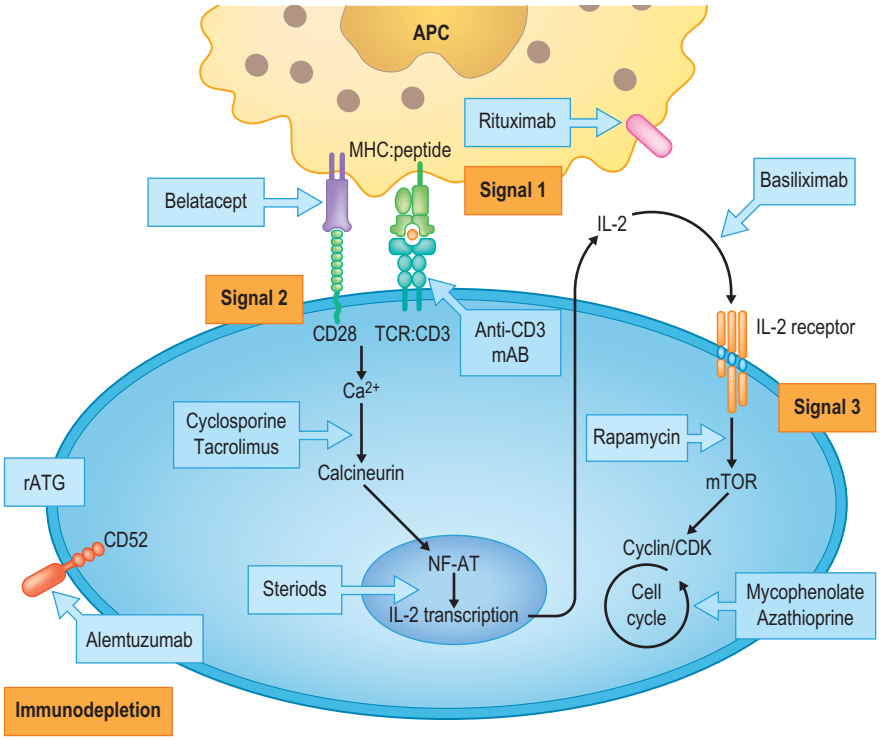


FIG. 89.6 Immunosuppressive Drugs and Their Targets. Signal 1 results from major histocompatibility complex *(MHC)*: antigen recognition through the T-cell receptor *(TCR)–CD3* complex, a process blocked by anti-CD3 mAbs and indirectly by rituximab. Signal 2 results in costimulation, a process that can be blocked by belatacept. Costimulation results in full activation of the TCR–CD3 complex, initiating signal transduction. Signal 3: downstream signaling pathways result in calcineurin activation, a stage that can be inhibited by tacrolimus and cyclosporine. Activated calcineurin dephosphorylates nuclear factor of activated T cells *(NFAT)*, allowing interleukin-2 (*IL-2*) transcription, a process that can be inhibited by steroids. IL-2 receptor stimulation, which is blocked by basiliximab, activates the mammalian target of rapamycin *(mTOR)* signaling cascade, which can be inhibited by sirolimus. This pathway induces the T cell to enter the cell cycle and proliferate, which in turn can be blocked by mycophenolate and azathioprine. Rabbit antithymocyte globulin (*rATG*) exerts polyclonal effects, and alemtuzumab binds to CD52, both resulting in immunodepletion.

THERAPEUTIC PRINCIPLES

- Patients receive induction immunosuppression

Non-lymphocyte depleting mAb or
Lymphocyte-depleting antibody or
Corticosteroids

- Maintenance therapy begins on the day of transplantation

Calcineurin inhibitor (CNI)
Anti-metabolite (mycophenolate mofetil)
Corticosteroids (or avoidance depending on induction)

- Special situations such as HLA-incompatible transplants: highly sensitized patients may also receive rituximab and/or intravenous immune globulin
- Therapeutic drug level monitoring for CNI and sometimes mycophenolic acid (MPA)

signal 1 via monoclonal antibodies has been used in both transplantation and autoimmunity and includes the already mentioned lymphocyte-depleting antibodies.

Rabbit Antithymocyte Globulin

Current therapeutic strategies are based on induction therapies that concentrate on profound immune-cell depletion at the time of transplantation, when immune activation is most intense and prior clinical uses included horse antilymphocyte globulin and anti-CD3 antibody. In current practice, the most commonly used agent is rabbit antithymocyte globulin (rATG), a lymphocyte-depleting polyclonal IgG preparation with specificity toward human thymocytes. It primarily binds to peripheral blood lymphocytes as well as to those present in the lymphoid organs, including lymph nodes, spleen, and thymus, as demonstrated by in vivo studies in NHPs. The agent's polyclonal nature enables it to display specificity toward a wide variety of molecules expressed on the surface of T cells, B cells, DCs, NK cells, and endothelial cells, including those involved in T-cell activation, proliferation, apoptosis, signal transduction, cell adhesion, and trafficking. The precise mechanism of action underlying the immunosuppressive efficacy of rATG in transplant recipients is unclear, although it has been primarily attributed to T-cell depletion. In vitro studies suggest that rATG modulates the expression of various lymphocyte surface antigens, resulting in apoptosis, antibody-dependent cytolysis, or complement-dependent lysis. Initially, this agent was primarily used for treatment of TCMR (and had approval for that purpose by the US Food and Drug Administration [FDA]), but evolving data for induction therapy have demonstrated non-inferiority with non-lymphocyte depletion induction in terms of incidence of biopsy-proven acute rejection, graft loss, or death at 6 and 12 months after transplantation, leading to FDA approval.[29] Treatment may be monitored by flow cytometry of total CD3 cells present in the recipient. Over 3 to 6 months, repopulation occurs, but with overall reduced total absolute lymphocyte count even at 1 year[30] with thymopoiesis and homeostatic proliferation contributing to immune reconstitution and expansion $CD4^+CD25^+$ Forkhead box $P3^+$ ($FOXP3^+$), CTLA-4^+, and glucocorticoid-induced TNF receptor (GITR+)$^+$ Treg from human peripheral blood lymphocytes. These Treg subsets are key to immune response modulation.

While definitive superiority compared to non-depletional induction therapy has not been demonstrated,[31] rATG has facilitated the avoidance of corticosteroids that were utilized widely in the prior decade and still utilized for immune "low-risk" patients in about 25% of KTRs. However, a systematic review has demonstrated higher rates of acute rejection but no difference in patient and graft survival at 5 years, without clear benefit or detriment beyond that.[32]

Alemtuzumab (Campath-1H)

Alemtuzumab is a humanized rat IgG2b directed against the CD52 antigen, which is expressed on 95% of peripheral blood lymphocytes, NK cells, macrophages, and thymocytes,[33] affecting nearly all mononuclear cells. The profound and long-lasting lymphopenia produced after the administration of a single dose of 30 mg of alemtuzumab is likely explained by such abundance on monocyte cell surfaces. Examination of the peripheral blood lymphocytes from recipients after alemtuzumab induction has identified a subset of T cells, predominantly CD4 central memory cells that survive despite alemtuzumab induction and appear largely resistant to depletion; these memory T cells express lower CD52 levels compared with naïve T cells. CD52 is not present on granulocytes, platelets, erythrocytes, or hematopoietic stem cells (HSCs). After binding to CD52, alemtuzumab causes cell death through several mechanisms: complement-mediated cytolysis, antibody-mediated cytotoxicity, and apoptosis. With a plasma half-life of approximately 12 days, its clinical effects are far more persistent, with greater than 99% lymphocyte depletion after a single dose, and lymph-node depletion taking up to 3 to 5 days compared with less than 1 hour seen in the peripheral lymphocytes.[34]

In this setting of severe lymphopenia, homeostatic proliferation affects the recovery of cell subpopulations, with rapid return of naïve and memory T cells that may trigger rejection.[35] CD8 cells appear to recover within 6 months, but CD4 cells may not reach pretransplant levels by 1 year, if at all. NK cells are almost unaffected and decrease only transiently (a population of $CD52^-$ NK cells has also been identified); monocyte and B-cell recovery can be seen at 3 and 12 months, respectively; T-cell levels recover to only 50% of baseline at 36 months.[34] Thus, there are lasting impacts on the recipient's peripheral blood cellular makeup, although long-term impacts have yet to show definitive damage.

While first used safely in transplantation as an induction agent in 1998 alone or with cyclosporine, other small clinical studies suggested the potency of lymphocyte depletion and ability to limit maintenance therapy with limited rejection and successful short-term graft survivals. Two randomized controlled trials in kidney transplantation have assessed the efficacy and safety of alemtuzumab compared to rATG[36,37] demonstrating alemtuzumab as an effective induction agent in both high- and low-immune-risk patients. Furthermore, it facilitates reduced exposure to CNI and mycophenolate, as well as steroid avoidance. Critically, late-term rejections have been observed with such strategies,[38] and some groups have noted AbMR as a common rejection phenotype, perhaps due to the minimization of maintenance therapy that facilitates (inadvertently) development of *de novo* DSA. Combinations of maintenance therapies including mTOR inhibitors (mTORi) and deoxyspergualin have been attempted to improve long-term engraftment. However, with the lack of pharmacokinetic data and additional trials, there is no FDA or European Medicines Agency (EMA) approval for its use in transplant and limited clinical availability unless off-label use.

Anti-CD20 Monoclonal Antibody (Rituximab)

Rituximab is an anti-CD20, chimeric mAb that eliminates most B cells from the circulation. Originally used to treat B-cell

lymphoproliferative diseases in patients other than transplant recipients as well as to treat post-transplant lymphoproliferative disease (PTLD), it is also now used in SOT as a treatment of antibody-mediated rejection and to "desensitize" patients who are receiving HLA-incompatible or ABO-incompatible transplants. However, depletion of antibody-producing cells may be incomplete because rituximab cannot target CD20-negative plasmablasts and plasma cells, which are the cells producing antibodies. The actions of rituximab in AbMR are to effectively deplete APC (e.g., B cells), limiting indirect pathway T-cell activation resulting in a sustained immune response. Recent studies of CCR5-deficient mice receiving kidney allografts demonstrated that early treatment with anti-CD20 at the time of transplantation effectively blocked the typical accelerated AbMR of control allografts, but when given at day 5, and afterwards, was unable to control the accelerated rejection.[39] While there may be a role for targeting B cells preemptively in SOT in highly sensitized individuals, additional strategies to downregulate the B-cell response are under evaluation.

Signal 2: Costimulatory Blockade: CTLA4Ig and Anti-CD40L

As already noted, a positive costimulatory signal is needed to promote T-cell activation; in its absence, partially activated T cells become hyporesponsive (anergic) or die via apoptosis.

The primary pathways of blockade for clinical implementation are CD28:CD80/86 and CD40:CD154. It has been hypothesized that the inhibition of full T-cell activation by costimulatory blockade rather than total T-cell depletion might more selectively target effector T cells and spare beneficial Treg while avoiding the many adverse effects of nonspecific immunodepletion. This has been based on preclinical studies and not been borne out in human transplantation.[7]

CTLA-4 (CD152) is an inducible T-cell surface antigen that, when bound to CD80/86 receptor ligands (B7 molecules) on APC, delivers inhibitory signals to the activated T cell. Belatacept (LEA29Y) is a fusion protein that combines a mutated version of the extracellular binding domain of CTLA-4 with the Fc portion of IgG1, with specificity for CD80/86 expressed on APC. Ligation of CD80/86 by CD28 (a surface antigen constitutively expressed on T cells) usually lowers the activation threshold of T cells (Fig. 89.6). Belatacept has a higher affinity and slower dissociation rate from human B7 molecules (i.e., CD80/86) compared with CD28, resulting in inhibition of the costimulation required for effective T-cell activation.

Belatacept has been studied in KTRs and is FDA-approved for use as prophylaxis for rejection. In the phase 2 trial KTRs,[40] belatacept treatment was associated with a low rate of rejection, similar to that in control treatment of cyclosporin A, mycophenolate, and corticosteroids. Moreover, graft function was significantly better with belatacept treatment. In two larger randomized trials,[41,42] treatment with belatacept was associated with higher rates of acute rejection in the first year of transplant compared to control treatment with cyclosporine, but with significantly better renal function and improved cardiovascular and metabolic profiles. These beneficial effects have been demonstrated out to 7 years post-transplant.[43] The higher rate of rejection with belatacept therapy has been borne out in other studies attempting to optimize its use. While these rejections are reversible and have not had demonstrable allograft loss, they make patient management impractical. Studies to understand this unexpected consequence in humans have noted the loss of CD28 on effector memory CD8 and CD4 T cells, effectively making these cells resistant to costimulatory blockade by CTLA4Ig.[44] In the latter case, they may be found in rejection biopsies and appear to acquire CD57, a marker of terminal differentiation and a ligand for P and L selectin. Their presence prior to transplant may identify patients resistant to treatment. Studies are ongoing to understand this risk and identify practical ways to implement these drugs. In clinical practice, they may also be useful in conversion when CNIs are not tolerated.[45]

Promising data in NHP models of kidney transplantation demonstrated beneficial effects of long-term allograft survival when anti-CD154 antibody was either combined with CTLA4Ig or as monotherapy.[46] However, human trials were halted due to unexpected life-threatening pulmonary emboli, with demonstration that activated platelets express and shed soluble CD154, promoting rejection, but which are also associated with the thrombosis seen in humans.[47] Removing the Fc portion of the antibody not only reduces this thrombosis but also promotes tolerance in murine models of allogeneic bone marrow and skin transplantation. For the most part, clinical attempts to use anti-CD154 antibody continue to be on hold[6] while the focus has turned to a non-activating anti-CD40 monoclonal antibody (bleselumab) to block the CD40–CD154 pathway, which showed promise in prolonging kidney and liver allograft survival in NHPs. A phase 2 trial in KTRs demonstrated three- to fourfold higher rates of biopsy-proven rejection[48] ending any potential in kidney transplants for the present. These disappointing results demonstrate the difficulty in translating immunologic findings in murine models, which are highly inbred and kept in isolation from pathogens, as compared to more complex findings in NHPs and ultimately to man.

Signal 3: Blockade of Proliferation and Differentiation

Anti-IL-2R (CD25) Monoclonal Antibody (Basiliximab)

As already discussed in Chapters 10 and 11, activated T cells produce IL-2, which binds to the IL-2 receptor high-affinity transmembrane protein subunit α (CD25), leading to signal transduction and T-cell activation. Anti-IL-2R (anti-CD25) mAb specifically target activated T cells but do not cause significant lymphocyte depletion and are not associated with major adverse effects compared with lymphocyte-depleting agents. However, other T-cell subtypes, including Treg, also express CD25, and therefore the use of these agents may impact some of the natural mechanisms of immunoregulation. Basiliximab, a chimeric mAb, binds the IL-2R with similar affinity as IL-2, thereby effectively competing with IL-2 and subsequently inhibiting IL-2–driven T-cell proliferation (Fig. 89.6).

It has been used in renal transplantation in low-immune-risk recipients and supported by consensus guidelines.[49] First transplants, well-matched organs, as well as living donors are considered "low" risk. However, in high-immune-risk situations, such as with the occurrence of detectable anti-donor HLA antibodies, re-transplantation, and the presence of high risk for acute kidney injury/DGF post-transplantation, rATG is favored.[31] With the selective and short-term immunosuppressive effect of basiliximab, confined to the highly immunogenic period immediately after transplantation, this class of drug has been under investigation to substitute for corticosteroids in early steroid withdrawal or steroid-free regimens. However, the debate

continues about the implementation of basiliximab in favor of rATG,[50] and it has also been utilized in the other solid organs.

Corticosteroids

Corticosteroids have complex immunosuppressive as well as anti-inflammatory effects (see Chapter 83). They act principally by binding to cytoplasmic glucocorticoid receptors, although at higher doses they can exhibit receptor-independent effects as well. The steroid–receptor complex translocates to the nucleus, where through DNA binding and by targeting transcription factors such as AP-1 and NF-κB affects a variety of cytokines including IL-1, IL-2, IL-3, and IL-6, TNF-α, IFN-γ; and a number of chemokines. By inhibiting cyclooxygenase, corticosteroids are also able to reduce the production of inflammatory mediators, such as leukotrienes, thromboxanes, and prostaglandins. Corticosteroids have been the mainstay of maintenance immunosuppression regimens for several decades (see Table 89.1, see Fig. 89.6). However, because of off-target effects, including its diabetogenic properties and induction of bone loss, skin changes, and other negative metabolic profiles, steroid avoidance has been utilized in many kidney transplant programs with significant success when coupled with antibody induction, at least in the short term.[37,51]

Antiproliferative Agents

With the discovery of azathioprine in 1957, anti-metabolites have been important maintenance immunosuppressive agents that interfere with DNA synthesis and prevent cell-cycle progression (see Fig. 89.6). In the context of allograft rejection, this impairs the clonal expansion of alloreactive T cells. While not used as commonly as mycophenolate, azathioprine is hepatically metabolized to the purine analogue 6-mercaptopurine and incorporated into DNA. By inhibiting purine nucleotide synthesis (and therefore DNA and RNA synthesis), it reduces gene transcription and prevents cell cycle progression. The effects of azathioprine are not lymphocyte-specific, resulting in bone marrow suppression.

In current clinical practice, mycophenolate mofetil (MMF) is the primary anti-metabolite used. MMF is metabolized in the liver to the active MPA. MPA is a noncompetitive, reversible inhibitor of inosine monophosphate dehydrogenase (IMPDH), the enzyme that catalyzes purine synthesis in the de novo pathway by conversion of inosine monophosphate to guanosine monophosphate. As the salvage pathway of purine synthesis from guanine is less active in lymphocytes, they are relatively dependent on the de novo pathway of purine synthesis compared with other cell types. As a result, the effects of MMF are more lymphocyte-specific than azathioprine, and less myelosuppressive. MMF therapy is utilized in all solid organs transplanted and is considered a standard of care following FDA approval in 1995 and EMA approval in 1996 for prophylaxis for transplant rejection.[52]

Calcineurin Inhibitors

Since their clinical introduction in the early 1980s with cyclosporin A (CsA) and tacrolimus (Tac) in early 1990s, this class of agent has had a substantial impact on transplantation, with a significant reduction in acute rejection rates and improved graft survival. About 95% of renal transplant recipients currently receive CNI.[14] Both CsA and Tac bind cytoplasmic immunophilins (cyclophilin and FK506-binding protein 12 [FKBP12], respectively) to form complexes that can inhibit the calcium-dependent phosphatase calcineurin, a rate-limiting enzyme in the TCR signal–transduction pathway (see Fig. 89.6). By preventing translocation of the transcription factor NFAT to the nucleus, calcineurin inhibition impairs upregulation of many molecules important for T-cell proliferation and the generation of an effective immune response, including the cytokines IL-2, IL-4, TNF-α, and IFN-γ, and costimulatory molecules, such as CD154. However, CNIs are associated with nephrotoxicity, both for native kidneys when used in other solid organs and in kidney allografts. Indeed, there have been many attempts to avoid and/or minimize their use, but this has been challenging with the occurrence of improved kidney function, albeit with a strong risk of rejection.[53] As already noted, the development of *de novo* DSA and late antibody-mediated rejection are consequences of patient- or physician-implemented minimization. The "Symphony" study, a large, randomized control study of ~1700 KTRs compared four maintenance regimens including reduction in tacrolimus target dosing and basiliximab induction. When compared to a CsA-based regimen or mTORi regimen, Tac treatment was associated with a significantly lower rejection rate (12% in the first year with a 95% 1-year graft survival and better renal function) than the other regimens, and this rate was maintained out to 3 years post-transplantation.[54] In the last few years, long-acting Tac preparations have been developed, allowing once-daily dosing to enhance adherence and reduce both pill burden and side effects such as tremor and nausea.[55]

Mammalian Target of Rapamycin Inhibitors

Sirolimus (rapamycin) and everolimus bind to the same immunophilin as tacrolimus (FKBP12), although the complexes formed do not interact with calcineurin but instead bind to the regulatory kinase mTOR, which has a critical role in cytokine receptor signal transduction (Fig. 89.6). mTOR activates the ribosomal enzyme p70 S6 kinase and blocks an inhibitory protein 4E-BP1, both of which are required for translation of proteins necessary for progression from the G_1 (growth) phase to the S (DNA synthesis) phase of the cell cycle. Therefore, inhibition of this pathway in T cells blocks the action of cytokines, such as IL-2, IL-4, and IL-15, preventing cell-cycle progression and clonal expansion. mTORi have also been tested in a variety of tumors with modest impact but are under investigation as a "sensitizer" for chemotherapy. In some vascularized cancers, the tumor suppressor gene *PTEN*, which is also a negative regulator of mTOR, loses function, leading to ongoing mTOR activity and tumorigenesis.

Unfortunately, this class has a substantial side-effect profile, including disabling extremity edema, pneumonitis, and dyslipidemia, as well as proteinuria. A recent study using everolimus in place of MPA with CNI demonstrated outcomes that were not inferior to standard immunosuppressive therapy.[56] While these agents are not used commonly in de novo maintenance therapy, they have been studied intently to avoid or minimize CNI. Conversion at 1 to 6 months post-transplant to sirolimus was associated with significantly improved renal function and similar rejection frequency as CsA-based therapy.[57] Another large randomized controlled trial used conversion to everolimus at 4.5 months, but with higher rejection rates, although graft function remained superior on everolimus, regardless of

rejection.[58,59] Late conversion has had mixed results. Practically speaking, for kidney transplantation, these agents are reserved as alternatives when other drugs have side effects and for use in kidney-transplant recipients who have repeated squamous cell cancers of the skin and who are at high risk for cancer recurrence post-transplantation, to mitigate the development of new tumors.[60] Finally, in heart transplantation, these agents are employed to mitigate the development of coronary vasculopathy.[61]

An interesting feature of mTORi therapy is their promotion of Treg generation and survival, as demonstrated in animal models of transplantation with prolonged graft survival.[62] In kidney and liver recipients, de novo mTORi use has also been associated with enhanced Treg development,[63,64] suggesting that mTORi are attractive agents to potentially promote tolerance, discussed subsequently in this chapter.

Desensitization and Antibody-Mediated Rejection Treatments

A major challenge, as noted above, is the development of AbMR. While we avoid preformed DSA prior to transplantation, some patients have such broad sensitization to HLA that it isn't possible to find a donor that is HLA-compatible. In order to facilitate a transplant, clinicians resort to *desensitization* therapy. While protocols vary and the approaches differ, the agents used here primarily are plasmapheresis to reduce total antibody load coupled with intravenous immune globulin (IVIG). IVIG treatment has broad effects, including anti-idiotypic antibodies to block response on endothelium, as a general anti-inflammatory binding cytokines and chemokines, and, importantly, blocking FcRn and facilitating excretion of anti-HLA antibody by the kidney (see Table 89.1). Rituximab is also used. Unfortunately, neither is FDA-approved, but they are used in humans based on substantial preclinical and mechanistic data and by consensus.[65]

Another approach to mitigate AbMR in this situation is to disrupt complement activation. Eculizumab, a humanized monoclonal antibody against complement C5 (Fig. 89.5), inhibiting the cleavage into C5a and C5b and thus MAC formation (see Chapter 3), has been utilized to mitigate the injury of AbMR. While the data for its use are limited, two studies in de novo showed that its use in HLA-incompatible transplantation were associated with low rates of AbMR, particularly in deceased-donor transplants,[66] but less significantly in living donor transplants[67] where the control treatment had very low rejection rates as well. Interestingly, eculizumab has no effect on the development of chronic AbMR in sensitized patients with high levels of DSA at the time of transplantation. Hence, eculizumab use is infrequent and typically reserved for resistant AbMR not responding to standard treatments. C1 esterase inhibitors are currently under study in AbMR treatment and mechanistically may be more effective in blocking complement injury more proximally.[68]

Plasma cells, the factories for antibody production, are an important potential target for treatment in AbMR. Several approaches are under investigation that seem promising.[69] Proteasome inhibitors bortezomib and carfilzomib, FDA-approved for treatment of multiple myeloma, function by inhibiting the 26 S proteasome, disrupting ubiquitination and protein degradation, ultimately disrupting the cell cycle and leading to cell apoptosis (see Table 89.1). Both bortezomib and carfilzomib have been successful in desensitizing KTRs but with rebound of antibody levels weeks after treatment. This allows engraftment without early rejection, but there is a return of DSA associated with later rejection. They have also been used in lung and heart allograft desensitization, in which bortezomib may be more effective. When used for late AbMR, bortezomib was not successful in reducing DSA levels or improving renal function, and was associated with significantly reduced DSA levels but could not mitigate late injury and was associated with hematologic, cardiac, and neurologic side effects, limiting its use.[70] Finally, promising preclinical work using novel combinations of proteasome inhibitor with anti-CD40 or with belatacept has been applied to patients, in sensitized heart[71] and kidney transplant candidates.[72] These studies demonstrate the ability to reduce HLA antibodies in the host, facilitate the transplant, and mitigate the development of AbMR when used in combination targeting both plasma cells and T-follicular helper cell/B-cell costimulation and activation that leads to AbMR. Other approaches that affect plasma cells include anti-CD38 antibody (daratumumab, isatuximab), which in preliminary studies decreases DSA in heart allograft rejection and targets IL-6, a proinflammatory cytokine produced by plasma cells either by neutralizing antibody or by blocking the IL-6 receptor.[69]

As another means to aggressively remove HLA antibodies in highly sensitized patients to facilitate transplantation, recent reports have utilized a novel agent IdeS (imlifidase). IdeS is an endopeptidase derived from *Streptococcus pyogenes* that has specificity for human IgG, and when infused intravenously results in the rapid cleavage of immunoglobulin. There have been several small studies in KTRs who were not transplantable due to high levels and broad specificity of HLA antibodies and were treated within 24 hours of transplantation. Treatment was associated with a dramatic drop in HLA antibody load, allowing transplantation[73,74] although AbMR occurred in 10 of 25 patients with only 1 graft loss. This approach was used for deceased donors as well as living donors because therapy can be given so proximally to the transplant and antibody loss is so comprehensive. Larger studies are underway to determine if this is a generalizable approach.

TOLERANCE

With organ grafts succumbing over time to graft loss due to immune-mediated injury, and the overall toxicity of chronic immunosuppression, numerous investigators have attempted to develop treatments that lead to donor specific hyporesponsiveness (i.e., tolerance) while maintaining an intact immune response for pathogens. In 1953, Billingham and co-workers first introduced the term *transplantation tolerance*, with the report that inoculation of fetal mice with lymphoid cells from an allogeneic adult donor mouse of a different strain led to later acceptance of skin grafts from the original skin graft donors. This work resulted in the Nobel Prize for Medicine in 1960.

The mechanisms of tolerance include deletion, anergy, immune regulation, clonal exhaustion, and ignorance and are discussed in detail in Chapter 10. The clinical approaches have leveraged our understanding of these fundamental mechanisms.[75] When an SOT recipient exhibits a well-functioning graft and lacks histologic signs of rejection after receiving no immunosuppression for at least 1 year, they have achieved *operational tolerance* (OT).[76] Importantly, these patients must also be capable of responding to other non–transplantation-related immune challenges, including infections and vaccinations. Over recent years, experimental models have shown that it is possible to exploit the processes by which immune homeostasis and tolerance

to self-antigens are maintained to induce tolerance to alloantigen. The optimal outcome for patients after transplantation would be the harnessing of these mechanisms to induce specific immunologic unresponsiveness or tolerance to the graft, thus avoiding the adverse side effects associated with current immunosuppressive regimens (see Table 89.1).

While ideal, OT in humans is extremely difficult to induce intentionally and has only been identified in a subset of transplant recipients. While rodent models have had success, translation into NHPs and humans has been more difficult. Another challenge is the lack of definitive laboratory parameters that give a clear indication of whether a particular recipient is tolerant of his or her graft. Furthermore, OT appears to be organ dependent. For example, recipients of a liver transplant have an advantage in developing OT because of the immune-privileged status of the liver. In humans, several studies demonstrate that a permanent and stable immunosuppressive-free state can be safely attempted and sometimes achieved in patients who have received a liver transplant.[77] However, OT has not been reported after intestinal, islet, or whole-organ pancreas transplantation, whereas two exceptional cases of OT have been described after lung and heart transplantations.[78] There have been more reports of late with cases in renal transplantation, based on wide nationwide surveys.[79,80] Important questions in studying these populations are the biologic signatures or biomarkers of OT.[79] The potential of these biomarkers goes beyond simply identifying potential individuals that may be "susceptible" to OT for immunosuppressive withdrawal. Some are assays assessing antigen-specific responses but are challenging, requiring abundant cellular material from both donor and recipient. A molecular approach seems to have the most interest, with a goal of a gene expression profile in peripheral blood with non–antigen-specific transcripts that correlate to outcome. Such a molecular profile or signature overcomes the complexity of a system with practically infinite individual cell combinations.

KEY CONCEPTS

Barriers to Transplantation Tolerance

- T-cell memory
  - Presensitization—direct exposure to alloantigen (e.g., pregnancy or blood transfusion).
  - Heterologous immunity—cross-reactivity in the T-cell repertoire among antiviral, antibacterial, environmental, and transplantation antigens.
  - Homeostatic proliferation—induced by lymphocyte-depleting antibodies (i.e., alemtuzumab).
  - Memory T cells generated by the above mechanisms—can result in rapidly formed effector immune responses upon rechallenge. These T cells are less sensitive to T-cell-depleting antibodies and costimulatory blockade. and thus may be more resistant to some tolerance induction strategies.
- B-cell response
  - Recipients treated with lymphocyte-depleting antibodies display a general increase in the naïve B-cell population.
  - There is a prevalent development of alloantibody in recipients treated with depleting antibody therapy.
  - Much focus of tolerogenic strategies is on the T-cell response. However, recent data suggest that the humoral immune system may play a more significant role than previously thought, possibly contributing to more long-term outcomes. Further work in this area continues.
- Lack of tolerance signature
  - An episode of acute rejection can severely affect graft survival in most transplanted organs.
  - There is absence of validated biomarkers of tolerance or predictors of rejection.
  - It is clinically difficult to often justify high-risk tolerizing strategies in patients who would otherwise do moderately well with standard immunosuppression.

Clinical Protocols in Tolerance Induction

Molecule-Based Protocols

One of the earliest attempts at OT was the administration of presumed tolerogenic induction therapy, followed by immunosuppressive withdrawal. As noted above in the immunosuppression section, the initial use of alemtuzumab in human recipients of kidney transplants with reduction in maintenance immunosuppression is defined as "prope," or near tolerance.[81,82] Complete withdrawal of immunosuppression has not been successful and even with modification of maintenance therapy to include mTORi, late AbMR has been noted. At best, reduction of baseline immunosuppression has been achieved with extremely close follow-up, making them impractical in kidney transplant programs that are high-volume and involve sensitized recipients. Moreover, it is now apparent that leukocyte depletion is not accompanied by a permanent and complete deletion of alloaggressive donor-reactive cells, and the establishment of a regulatory network is required to maintain tolerance. In contrast, numerous studies in liver transplant recipients have occurred with an overall tolerance rate of about 20% to 30% of the study population.[77] These results are dependent on the tolerogenic nature of the liver, the age of recipient, maintenance immunosuppression, and cause of liver disease, whether metabolic or immunologic/infectious.[76]

Full Chimerism

The more robust experimental strategies for the induction of tolerance to foreign antigen utilize the mechanisms of central deletion to eliminate T-cell clones with specificity for the foreign antigens in question, thereby preventing them from entering the periphery. The establishment of hematopoietic chimerism, through bone marrow transplantation, can reliably achieve this. Stable engraftment of donor HSCs results in repopulation of the recipient thymus with donor-type thymic DCs, and T cells with anti-donor specificity are deleted by negative selection.

Full chimerism requires the ablation of the recipient's immune system with high-dose radiation and/or chemotherapy. Alternatively, nonablative conditioning regimens may be used, followed by the infusion of donor's marrow to colonize the recipient completely. This phenomenon paves the way for the onset of tolerance in the case of a subsequent SOT from the same donor.[83] As proof of concept, numerous patients have undergone successful bone marrow transplant (BMT) for hematologic malignancy indications and have subsequently been successfully transplanted with a kidney from the same donor, without the requirement for increased immunosuppression.[84] It is important to highlight that in all these cases, the use of BMT was justified on the basis of the need for treatment of hematologic malignancies.

An alternative approach to achieving full chimerism in HLA-*mis*matched unrelated stem cell/renal transplant human recipients involves infusing a cell product enriched for tolerogenic graft facilitating cells (FCs) as well as HSCs and T cells ("FCRx") rather than just bone marrow graft alone.[85,86] These bone mar-

row–derived FCs, which are $CD8^+$ but do not express a TCR, potently enhance engraftment of allogeneic HSCs in conditioned recipients. FCs are composed predominantly of a plasmacytoid precursor DC subpopulation, induce the generation of antigen-specific Treg in vitro and in vivo, and have been found to effectively prevent graft-versus-host disease (GvHD) in mice. Initial studies have demonstrated macrochimerism with no incidence of GvHD or engraftment syndrome in a small number of patients, and were immunosuppression-free—that is, these patients were clinically operationally tolerant. In vitro studies indicated that chimeric donor lymphocytes were tolerized to the recipient, with a significant increase in the CD4 Treg/T-effector cell population ratio observed in these patients when compared with those who had lost chimerism. This approach to establish high levels of donor multilineage chimerism has exciting therapeutic implications for disorders for which HSC transplantation can provide a "functional cure," including inherited metabolic disorders, hemoglobinopathies, and autoimmune disease, as well as in SOT. Longer-term follow-up data on these initial patients as well as data from a phase 3 clinical trial are anticipated.

Mixed Chimerism

Another promising approach is the induction of mixed hematopoietic chimerism, which can be achieved in experimental models and clinical settings using nonmyeloablative conditioning regimens with far less toxic induction therapy.[87] Examples include either a combination of depleting anti-CD4 and anti-CD8 antibodies together with mild, nonmyeloablative total body irradiation or costimulatory blockade with anti-CD154 and/or CTLA4Ig. When these induction protocols are followed by BMT, the result is mixed chimerism, meaning the continued survival of both donor and recipient hematopoietic progenitor cells. In mouse and NHP models that have undergone these therapies, there is durable tolerance to donor-type allografts, which have a much lower incidence of GvHD compared with full chimeras; similar results have been demonstrated in early studies in human recipients as well.

Regulatory T Cells

Treg have the capability of suppressing immune responses and are thus a focus in transplantation to mitigate alloantigen responses and/or induce tolerance. As discussed in Chapter 13, Treg may exist as thymic-derived naturally occurring $CD25^+CD4^+$ cells (tTreg) or induced Treg (iTreg) that are either differentiated from $CD25^-CD4^+$ nonregulatory cells or expanded from $CD25^+CD4^+$ cells in response to antigen. Both tTreg and iTreg have been demonstrated to play important roles in transplant tolerance.[88] Expression of the transcription factor FOXP3 is essential for the development and function of Treg. However, tTreg and iTreg differ in origin, antigen experience, methylation patterns of FOXP3, and suppressive mechanisms. Unlike mice, FOXP3 in humans is also expressed transiently by activated non-regulatory T cells that also upregulate CD25 expression. Thus, not all $CD25^+FOXP3^+CD4^+$ cells will be genuine human Treg, and so isolation strategies based on $CD25^{hi}CD4^+$ are likely to be imperfect. In humans, $CD127^{lo}CD25^+CD4^+$ T cells are characterized by a higher level of FOXP3 expression and a more pronounced suppressive capacity.[88,89]

Strategies exist for both in vivo or ex vivo generation and/or expansion of Treg. The most common ex vivo approach is based on stimulation with anti-CD3 and anti-CD28 beads in the presence of high concentrations of IL-2, which stimulates proliferation sometimes with rapamycin (mTORi) present.[90,91] Importantly, ex vivo–expanded $CD25^{hi}CD4^+$ and $CD127^{lo}CD25^+CD4^+$ Treg are effective at inhibiting vasculopathy seen in late allograft rejection in a humanized mouse model, whereas $CD127^{lo}CD25^+CD4^+$ T cells are five times more efficient than those not expressing a low level of CD127.[62] In humans, infusion was well tolerated and in some cases reduced intragraft inflammation. Despite the generation of sufficient numbers of Treg for cellular therapy, this mode of expansion is *antigen nonspecific*, without any enrichment steps for the cells of interest.

Donor-antigen-specific Treg are more appealing for clinical application in transplantation, as they may be more potent on a cell-per-cell basis at controlling allograft rejection. Here, exposure to antigen may either expand tTreg or induce the generation of iTreg from cells that do not originally possess regulatory activity. Results from the ONE study consortium of six nonrandomized trials across multiple transplant centers included 104 patients receiving Treg, DCs, or macrophages with an overall rejection rate of 16%, as compared to 12% in standard therapy patients. Forty percent of treated subjects tolerated weaning to tacrolimus alone,[92] demonstrating the safety of regulatory-cell adoptive transfer. Treg therapy in liver transplant recipients resulted in reduced anti-donor T-cell responses post-transplantation.[93] However, questions remain regarding such therapy, including stability and plasticity of such cells, migration patterns, and optimal maintenance immunosuppressive therapy.

Finally, typical immunosuppressive therapy may affect Treg function in vivo. CNIs, in particular cyclosporine, are detrimental to Treg expansion and growth, whereas the mTORi rapamycin appears to be beneficial to Treg in terms of both in vivo generation and function in mouse models and in vitro cultures of human Treg.

There are other cell-based therapies being investigated for their tolerogenic potentials. It should be stressed that these studies have been conducted in limited numbers of patients and include mesenchymal stromal cells (MSCs)[94] and regulatory macrophages to facilitate immunosuppression minimization as well as regulate inflammation in other inflammatory conditions.

Biomarkers of Rejection or Tolerance

A major challenge in the field of transplantation is the ability to determine the sufficiency of immunosuppression in each transplant recipient and provide a more personalized approach to care. Current methods for care include monitoring trough drug levels and assessment for viral infections. Even so, recipients may still develop infection or rejection even when they appear to be adherent to their regimen. Moreover, there is new recognition that inflammation in the transplant may occur prior to functional changes, and this "subclinical" inflammation leads to late graft loss. As functional changes reflect clinical inflammation, surveillance biopsies may be employed, both in the kidney and heart allografts, to detect inflammation. These are costly, invasive, and have complications. Hence, the ability to noninvasively determine the immune status of the recipient, predict and/or diagnose rejection, and detect tolerance would allow for individualized immunosuppressive therapy.[95]

Technical advances in multiparameter flow cytometry, antigen-specific lymphocyte assays, and genome-wide analyses have led to the development of powerful and more standardized

techniques to characterize alloimmune responses. Gene expression profiles in peripheral blood have been a critical focus for detecting transplant tolerance[80] or detecting rejection,[96] with the latter achieving commercialization. Gene expression in allograft biopsies has made significant progress creating a "molecular microscope" as an adjunct to the histologic features that suffer from inter-observer variability and limited sample.[97] In plasma, the detection of donor-derived cell-free DNA correlates with organ rejection in heart and kidney allografts[98] and is commercialized for use in the latter.

While these are exciting developments for patient care, and some of these tools are commercialized, there is a need for validation in larger cohorts.[99] None of the referenced assays are FDA-approved for clinical use. Moreover, biomarker qualification for regulatory drug development is an entirely different process, but a recent meeting with leading experts provided some significant guidance for advancing transplant therapeutics.[100]

NEW FRONTIERS IN ORGAN SUPPLY FOR TRANSPLANTATION: XENOTRANSPLANTATION

There remains a critical shortage of solid organs for those patients who would benefit from transplant, and many individuals die waiting to be transplanted. To address this challenge, new strategies in mitigating the shortage are being investigated, including using human immunodeficiency virus (HIV)-positive donor organs into HIV-positive recipients, and using hepatitis C–infected organs that would otherwise be discarded and transplanting into hepatitis C naïve recipients followed by antiviral treatment.[101] An alternative strategy is using animal organs that are genetically manipulated to limit adaptive immune responses for human transplantation, a strategy called xenotransplantation.

Xenotransplantation began in the early 1960s, with studies of kidneys transplanted from NHP into human patients with end-stage kidney disease being met with limited success. In 1985, headlines were made when a baboon heart was transplanted into an infant with a severe congenital heart defect, with fatal rejection. With little success using NHP–human xenotransplantation and concerns of zoonosis, advancements stalled. Current research is focused on enhancing the suitability of pig-organ transplantation for human recipients, primarily through the genetic modification of pigs using NHPs as the recipients most closely matching humans.

The critical barrier for engraftment is the presence of preformed natural antibodies against donor antigens expressed on grafts from other species. For clinical purposes, the most important have been antibodies to galactose-α−1,3-galactose (α-gal). These preformed antibodies are responsible for instances of hyperacute rejection and graft failure in the early xenotransplant experiments involving pig solid organs from genetically *un*modified animals.[102] Further research in pigs, NHPs, and humans has revealed a number of additional immune proteins that play important roles in species–specific host immunologic responses. To enhance the longevity of the pig xenografts, investigators have genetically modified the donor pig in two ways: by reducing complement activation and by deleting GAL. In the former, protective constructs can be introduced into the genome, such as those of human complement regulatory proteins (e.g., hCD59), to dampen the human host complement response to the xenograft. In the latter, deletion of GAL, which is now accomplished more easily than homologous recombination using CRISPR-Cas9 technology, can prevent a destructive host response.[103] Coupled with treatment with anti-CD154 antibody and standard immunosuppressive agents, survival can be extended beyond a few days to months in heart and kidney xenotransplants. Given the progress in this area, it has been speculated that a clinical trial in xenotransplantation is anticipated within the next few years.[104] This will require optimization of immunosuppressive therapy and assessment of the presence of human recipient anti-HLA antibodies, which may cross-react with swine leukocyte antigens to design the compatible pig organ.

ON THE HORIZON

- Clinical introduction of antigen-specific regulatory T-cell therapy for tolerance induction
- Implementation of biomarker monitoring of allografts for subclinical immune-mediated injury
- Development of pig organs for the first pig-to-man kidney xenotransplant
- Identification of safe and effective treatments for antibody-mediated rejection

REFERENCES

1. Petrenko A, Carnevale M, Somov A, et al. Organ preservation into the 2020s: the era of dynamic intervention. *Transfus Med Hemother*. 2019;46(3):151–172.
2. Resch T, Cardini B, Oberhuber R, et al. Transplanting marginal organs in the era of modern machine perfusion and advanced organ monitoring. *Front Immunol*. 2020;11:631.
3. Mannon RB. Delayed graft function: the AKI of kidney transplantation. *Nephron*. 2018;140(2):94–98.
4. Wiebe C, Kosmoliaptsis V, Pochinco D, et al. HLA-DR/DQ molecular mismatch: a prognostic biomarker for primary alloimmunity. *Am J Transplant*. 2019;19(6):1708–1719.
5. Afzali B, Lombardi G, Lechler RI. Pathways of major histocompatibility complex allorecognition. *Curr Opin Organ Transplant*. 2008;13(4):438–444.
6. van der Zwan M, Hesselink DA, van den Hoogen MWF, Baan CC. Costimulation blockade in kidney transplant recipients. *Drugs*. 2020;80(1):33–46.
7. Schroder PM, Fitch ZW, Schmitz R, et al. The past, present, and future of costimulation blockade in organ transplantation. *Curr Opin Organ Transplant*. 2019;24(4):391–401.
8. Romano M, Tung SL, Smyth LA, Lombardi G. Treg therapy in transplantation: a general overview. *Transpl Int*. 2017;30(8):745–753.
9. Mathews DV, Wakwe WC, Kim SC, et al. Belatacept-resistant rejection is associated with CD28+ memory CD8 T cells. *Am J Transplant*. 2017;17(9):2285–2299.
10. Loupy A, Haas M, Roufosse C, et al. The Banff 2019 Kidney Meeting Report (I): updates on and clarification of criteria for T cell- and antibody-mediated rejection. *Am J Transplant*. 2020;20(9):2318–2331.
11. Lefaucheur C, Gosset C, Rabant M, et al. T cell–mediated rejection is a major determinant of inflammation in scarred areas in kidney allografts. *Am J Transplant*. 2018;18(2):377–390.
12. Nankivell BJ, Shingde M, Keung KL, et al. The causes, significance and consequences of inflammatory fibrosis in kidney transplantation: the Banff i-IFTA lesion. *Am J Transplant*. 2018;18(2):364–376.
13. Siu JHY, Surendrakumar V, Richards JA, Pettigrew GJ. T cell allorecognition pathways in solid organ transplantation. *Front Immunol*. 2018;9:2548.
14. OPTN/SRTR 2018 annual data report: introduction. *Am J Transplant*. 2020;20(s1):11–19.
15. Brahmer JR, Lacchetti C, Schneider BJ, et al. Management of immune-related adverse events in patients treated with immune checkpoint inhibitor therapy: American Society of Clinical Oncology Clinical Practice Guideline. *J Clin Oncol*. 2018;36(17):1714–1768.

16. Valenzuela NM, Reed EF. Antibody-mediated rejection across solid organ transplants: manifestations, mechanisms, and therapies. *J Clin Invest*. 2017;127(7):2492–2504.
17. van den Heuvel H, Heidt S, Roelen DL, Claas FH. T-cell alloreactivity and transplantation outcome: a budding role for heterologous immunity? *Curr Opin Organ Transplant*. 2015;20(4):454–460.
18. Hickey MJ, Valenzuela NM, Reed EF. Alloantibody generation and effector function following sensitization to human leukocyte antigen. *Front Immunol.* 2016;7:30.
19. Morris AB, Sullivan HC, Krummey SM, et al. Out with the old, in with the new: virtual versus physical crossmatching in the modern era. *HLA*. 2019;94(6):471–481.
20. Reed EF, Rao P, Zhang Z, et al. Comprehensive assessment and standardization of solid phase multiplex-bead arrays for the detection of antibodies to HLA. *Am J Transplant*. 2013;13(7):1859–1870.
21. Garcia de Mattos Barbosa M, Cascalho M, Platt JL. Accommodation in ABO-incompatible organ transplants. *Xenotransplantation.* 2018;25(3):e12418.
22. Tambur AR, Campbell P, Claas FH, et al. Sensitization in transplantation: assessment of risk (STAR) 2017 Working Group Meeting Report. *Am J Transplant*. 2018;18(7):1604–1614.
23. Seifert ME, Mannon RB. Modernization of chronic allograft injury research: better biomarkers, better studies, better outcomes. *Clin Transpl*. 2015;31:211–225.
24. Gaston RS, Cecka JM, Kasiske BL, et al. Evidence for antibody-mediated injury as a major determinant of late kidney allograft failure. *Transplantation.* 2010;90(1):68–74.
25. Viglietti D, Loupy A, Vernerey D, et al. Value of donor-specific anti-hla antibody monitoring and characterization for risk stratification of kidney allograft loss. *J Am Soc Nephrol*. 2017;28(2):702–715.
26. Levy L, Huszti E, Renaud-Picard B, et al. Risk assessment of chronic lung allograft dysfunction phenotypes: validation and proposed refinement of the 2019 International Society for Heart and Lung Transplantation classification system. *J Heart Lung Transplant*. 2020;39(8):761–770.
27. Lee F, Nair V, Chih S. Cardiac allograft vasculopathy: insights on pathogenesis and therapy. *Clin Transplant.* 2020;34(3):e13794.
28. Nickeleit V, Singh HK, Randhawa P, et al. The banff working group classification of definitive polyomavirus nephropathy: morphologic definitions and clinical correlations. *J Am Soc Nephrol.* 2018;29(2):680.
29. Alloway RR, Woodle ES, Abramowicz D, et al. Rabbit anti-thymocyte globulin for the prevention of acute rejection in kidney transplantation. *Am J Transplant*. 2019;19(8):2252–2261.
30. Gurkan S, Luan Y, Dhillon N, et al. Immune reconstitution following rabbit antithymocyte globulin. *Am J Transplant*. 2010;10(9):2132–2141.
31. Brennan DC, Daller JA, Lake KD, et al. Rabbit antithymocyte globulin versus basiliximab in renal transplantation. *N Engl J Med*. 2006;355(19):1967–1977.
32. Haller MC, Royuela A, Nagler EV, et al. Steroid avoidance or withdrawal for kidney transplant recipients. *Cochrane Database Syst Rev*. 2016;8:Cd005632.
33. Morris PJ, Russell NK. Alemtuzumab (Campath-1H): a systematic review in organ transplantation. *Transplantation*. 2006;81(10):1361–1367.
34. Kirk AD, Hale DA, Mannon RB, et al. Results from a human renal allograft tolerance trial evaluating the humanized CD52-specific monoclonal antibody alemtuzumab (CAMPATH-1H). *Transplantation.* 2003;76(1):120–129.
35. Wu Z, Bensinger SJ, Zhang J, et al. Homeostatic proliferation is a barrier to transplantation tolerance. *Nat Med.* 2004;10(1):87–92.
36. Group CSC, Haynes R, Harden P, et al. Alemtuzumab-based induction treatment versus basiliximab-based induction treatment in kidney transplantation (the 3C Study): a randomised trial. *Lancet.* 2014;384(9955):1684–1690.
37. Hanaway MJ, Woodle ES, Mulgaonkar S, et al. Alemtuzumab induction in renal transplantation. *N Engl J Med*. 2011;364(20):1909–1919.
38. Watson CJ, Bradley JA, Friend PJ, et al. Alemtuzumab (CAMPATH 1H) induction therapy in cadaveric kidney transplantation—efficacy and safety at five years. *Am J Transplant*. 2005;5(6):1347–1353.
39. Abe T, Ishii D, Gorbacheva V, et al. Anti-huCD20 antibody therapy for antibody-mediated rejection of renal allografts in a mouse model. *Am J Transplant*. 2015;15(5):1192–1204.
40. Vincenti F, Larsen C, Durrbach A, et al. Costimulation blockade with belatacept in renal transplantation. *N Engl J Med*. 2005;353(8):770–781.
41. Larsen CP, Grinyo J, Medina-Pestana J, et al. Belatacept-based regimens versus a cyclosporine A-based regimen in kidney transplant recipients: 2-year results from the BENEFIT and BENEFIT-EXT studies. *Transplantation.* 2010;90(12):1528–1535.
42. Vincenti F, Charpentier B, Vanrenterghem Y, et al. A phase III study of belatacept-based immunosuppression regimens versus cyclosporine in renal transplant recipients (BENEFIT study). *Am J Transplant*. 2010;10(3):535–546.
43. Vincenti F, Rostaing L, Grinyo J, et al. Belatacept and long-term outcomes in kidney transplantation. *N Engl J Med*. 2016;374(4):333–343.
44. Murakami N, Riella LVCD4+. CD28-negative cells: armed and dangerous. *Am J Transplant*. 2016;16(4):1045–1046.
45. Grinyó JM, Del Carmen Rial M, Alberu J, et al. Safety and efficacy outcomes 3 years after switching to belatacept from a calcineurin inhibitor in kidney transplant recipients: results from a phase 2 randomized trial. *Am J Kidney Dis*. 2017;69(5):587–594.
46. Kirk AD, Burkly LC, Batty DS, et al. Treatment with humanized monoclonal antibody against CD154 prevents acute renal allograft rejection in nonhuman primates. *Nat Med.* 1999;5(6):686–693.
47. Xu H, Zhang X, Mannon RB, Kirk AD. Platelet-derived or soluble CD154 induces vascularized allograft rejection independent of cell-bound CD154. *J Clin Invest*. 2006;116(3):769–774.
48. Harland RC, Klintmalm G, Jensik S, et al. Efficacy and safety of bleselumab in kidney transplant recipients: a phase 2, randomized, open-label, noninferiority study. *Am J Transplant*. 2020;20(1):159–171.
49. Kidney Disease: Improving Global Outcomes Transplant Work G. KDIGO clinical practice guideline for the care of kidney transplant recipients. Am J Transplant. 2009;9 Suppl 3:S1–155.
50. Hellemans R, Bosmans JL, Abramowicz D. Induction therapy for kidney transplant recipients: do we still need anti-IL2 receptor monoclonal antibodies? *Am J Transplant*. 2017;17(1):22–27.
51. Swanson SJ, Hale DA, Mannon RB, et al. Kidney transplantation with rabbit antithymocyte globulin induction and sirolimus monotherapy. *Lancet.* 2002;360(9346):1662–1664.
52. Dalal P, Grafals M, Chhabra D, Gallon L. Mycophenolate mofetil: safety and efficacy in the prophylaxis of acute kidney transplantation rejection. *Ther Clin Risk Manag*. 2009;5(1):139–149.
53. Sawinski D, Trofe-Clark J, Leas B, et al. Calcineurin inhibitor minimization, conversion, withdrawal, and avoidance strategies in renal transplantation: a systematic review and meta-analysis. *Am J Transplant*. 2016;16(7):2117–2138.
54. Ekberg H, Bernasconi C, Tedesco-Silva H, et al. Calcineurin inhibitor minimization in the Symphony study: observational results 3 years after transplantation. *Am J Transplant*. 2009;9(8):1876–1885.
55. Budde K, Bunnapradist S, Grinyo JM, et al. Novel once-daily extended-release tacrolimus (LCPT) versus twice-daily tacrolimus in de novo kidney transplants: one-year results of Phase III, double-blind, randomized trial. *Am J Transplant*. 2014;14(12):2796–2806.
56. Pascual J, Berger SP, Witzke O, et al. Everolimus with reduced calcineurin inhibitor exposure in renal transplantation. *J Am Soc Nephrol*. 2018;29(7):1979–1991.
57. Weir MR, Mulgaonkar S, Chan L, et al. Mycophenolate mofetil–based immunosuppression with sirolimus in renal transplantation: a randomized, controlled Spare-the-Nephron trial. *Kidney Int*. 2011;79(8):897–907.
58. Budde K, Becker T, Arns W, et al. Everolimus-based, calcineurin-inhibitor-free regimen in recipients of de-novo kidney transplants: an open-label, randomised, controlled trial. *Lancet.* 2011;377(9768):837–847.
59. Budde K, Lehner F, Sommerer C, et al. Five-year outcomes in kidney transplant patients converted from cyclosporine to everolimus: the randomized ZEUS study. *Am J Transplant*. 2015;15(1):119–128.
60. Wolf S, Hoffmann VS, Habicht A, et al. Effects of mTOR-Is on malignancy and survival following renal transplantation: a systematic review and meta-analysis of randomized trials with a minimum follow-up of 24 months. *PloS one*. 2018;13(4). e0194975-e.

61. Ueyama H, Kuno T, Takagi H, et al. Maintenance immunosuppression in heart transplantation: insights from network meta-analysis of various immunosuppression regimens. *Heart Fail Rev*. 2020. https://doi.org/10.1007/s10741-020-09967-3.
62. Nadig SN, Wieckiewicz J, Wu DC, et al. In vivo prevention of transplant arteriosclerosis by ex vivo-expanded human regulatory T cells. *Nat Med*. 2010;16(7):809–813.
63. Salama AD, Najafian N, Clarkson MR, et al. Regulatory CD25$^+$ T cells in human kidney transplant recipients. *J Am Soc Nephrol*. 2003;14(6):1643–1651.
64. Akimova T, Kamath BM, Goebel JW, et al. Differing effects of rapamycin or calcineurin inhibitor on T-regulatory cells in pediatric liver and kidney transplant recipients. *Am J Transplant*. 2012;12(12):3449–3461.
65. Schinstock CA, Mannon RB, Budde K, et al. Recommended treatment for antibody-mediated rejection after kidney transplantation: the 2019 expert consensus from the Transplantion Society Working Group. *Transplantation*. 2020;104(5):911–922.
66. Glotz D, Russ G, Rostaing L, et al. Safety and efficacy of eculizumab for the prevention of antibody-mediated rejection after deceased-donor kidney transplantation in patients with preformed donor-specific antibodies. *Am J Transplant*. 2019;19(10):2865–2875.
67. Marks WH, Mamode N, Montgomery RA, et al. Safety and efficacy of eculizumab in the prevention of antibody-mediated rejection in living-donor kidney transplant recipients requiring desensitization therapy: a randomized trial. *Am J Transplant*. 2019;19(10):2876–2888.
68. Montgomery RA, Orandi BJ, Racusen L, et al. Plasma-derived C1 esterase inhibitor for acute antibody-mediated rejection following kidney transplantation: results of a randomized double-blind placebo-controlled pilot study. *Am J Transplant*. 2016;16(12):3468–3478.
69. Agarwal D, Allman D, Naji A. Novel therapeutic opportunities afforded by plasma cell biology in transplantation. *Am J Transplant*. 2020;20(8):1984–1991.
70. Eskandary F, Regele H, Baumann L, et al. A randomized trial of bortezomib in late antibody-mediated kidney transplant rejection. *J Am Soc Nephrol*. 2018;29(2):591–605.
71. Alishetti S, Farr M, Jennings D, et al. Desensitizing highly sensitized heart transplant candidates with the combination of belatacept and proteasome inhibition. *Am J Transplant*. 2020;20(12):3620–3630.
72. Jain D, Rajab A, Young JS, et al. Reversing donor-specific antibody responses and antibody-mediated rejection with bortezomib and belatacept in mice and kidney transplant recipients. *Am J Transplant*. 2020;20(10):2675–2685.
73. Lonze BE, Tatapudi VS, Weldon EP, et al. IdeS (Imlifidase): a novel agent that cleaves human IgG and permits successful kidney transplantation across high-strength donor-specific antibody. *Ann Surg*. 2018;268(3):488–496.
74. Jordan SC, Lorant T, Choi J. IgG endopeptidase in highly sensitized patients undergoing transplantation. *N Engl J Med*. 2017;377(17):1693–1694.
75. Leventhal JR, Mathew JM. Outstanding questions in transplantation: tolerance. *Am J Transplant*. 2020;20(2):348–354.
76. Feng S, Bucuvalas J. Tolerance after liver transplantation: where are we? *Liver Transplant*. 2017;23(12):1601–1614.
77. Levitsky J, Feng S. Tolerance in clinical liver transplantation. *Hum Immunol*. 2018;79(5):283–287.
78. Madariaga MLL, Kreisel D, Madsen JC. Organ-specific differences in achieving tolerance. *Curr Opin Organ Transplant*. 2015;20(4):392–399.
79. Massart A, Ghisdal L, Abramowicz M, Abramowicz D. Operational tolerance in kidney transplantation and associated biomarkers. *Clin Exp Immunol*. 2017;189(2):138–157.
80. Newell KA, Adams AB, Turka LA. Biomarkers of operational tolerance following kidney transplantation—The immune tolerance network studies of spontaneously tolerant kidney transplant recipients. *Hum Immunol*. 2018;79(5):380–387.
81. Calne R, Friend P, Moffatt S, et al. Prope tolerance, perioperative campath 1H, and low-dose cyclosporin monotherapy in renal allograft recipients. *Lancet*. 1998;351(9117):1701–1702.
82. Calne R, Moffatt SD, Friend PJ, et al. Campath IH allows low-dose cyclosporine monotherapy in 31 cadaveric renal allograft recipients. *Transplantation*. 1999;68(10):1613–1616.
83. Oura T, Cosimi AB, Kawai T. Chimerism-based tolerance in organ transplantation: preclinical and clinical studies. *Clin Exp Immunol*. 2017;189(2):190–196.
84. Eder M, Schwarz C, Kammer M, et al. Allograft and patient survival after sequential HSCT and kidney transplantation from the same donor–a multicenter analysis. *Am J Transplant*. 2019;19(2):475–487.
85. Leventhal JR, Elliott MJ, Yolcu ES, et al. Immune reconstitution/immunocompetence in recipients of kidney plus hematopoietic stem/facilitating cell transplants. *Transplantation*. 2015;99(2):288–298.
86. Leventhal JR, Miller J, Mathew JM, et al. Updated follow-up of a tolerance protocol in HLA-identical renal transplant pairs given donor hematopoietic stem cells. *Hum Immunol*. 2018;79(5):277–282.
87. Sasaki H, Oura T, Spitzer TR, et al. Preclinical and clinical studies for transplant tolerance via the mixed chimerism approach. *Hum Immunol*. 2018;79(5):258–265.
88. Juvet SC, Whatcott AG, Bushell AR, Wood KJ. Harnessing regulatory T cells for clinical use in transplantation: the end of the beginning. *Am J Transplant*. 2014;14(4):750–763.
89. Putnam AL, Brusko TM, Lee MR, et al. Expansion of human regulatory T-cells from patients with type 1 diabetes. *Diabetes*. 2009;58(3):652–662.
90. Tang Q, Vincenti F. Transplant trials with Tregs: perils and promises. *J Clin Invest*. 2017;127(7):2505–2512.
91. Hu M, Wang YM, Wang Y, et al. Regulatory T cells in kidney disease and transplantation. *Kidney Int*. 2016;90(3):502–514.
92. Sawitzki B, Harden PN, Reinke P, et al. Regulatory cell therapy in kidney transplantation (The ONE Study): a harmonised design and analysis of seven non-randomised, single-arm, phase 1/2A trials. *Lancet*. 2020;395(10237):1627–1639.
93. Sánchez-Fueyo A, Whitehouse G, Grageda N, et al. Applicability, safety, and biological activity of regulatory T cell therapy in liver transplantation. *Am J Transplant*. 2020;20(4):1125–1136.
94. Casiraghi F, Perico N, Remuzzi G. Mesenchymal stromal cells for tolerance induction in organ transplantation. *Hum Immunol*. 2018;79(5):304–313.
95. Eikmans M, Gielis EM, Ledeganck KJ, et al. Non-invasive biomarkers of acute rejection in kidney transplantation: novel targets and strategies. *Front Med (Lausanne)*. 2018;5:358.
96. Friedewald JJ, Kurian SM, Heilman RL, et al. Development and clinical validity of a novel blood-based molecular biomarker for subclinical acute rejection following kidney transplant. *Am J Transplant*. 2019;19(1):98–109.
97. Halloran PF, Reeve J, Akalin E, et al. Real time central assessment of kidney transplant indication biopsies by microarrays: the INTERCOMEX study. *Am J Transplant*. 2017;17(11):2851–2862.
98. Bloom RD, Bromberg JS, Poggio ED, et al. Cell-free DNA and active rejection in kidney allografts. *J Am Soc Nephrol*. 2017;28(7):2221–2232.
99. Naesens M, Anglicheau D. Precision transplant medicine: biomarkers to the rescue. *J Am Soc Nephrol*. 2018;29(1):24–34.
100. Mannon RB, Morris RE, Abecassis M, et al. Use of biomarkers to improve immunosuppressive drug development and outcomes in renal organ transplantation: a meeting report. *Am J Transplant*. 2020;20(6):1495–1502.
101. Polanco NP, Goldberg D. Transplanting livers from "HCV-positive" donors to HCV-negative recipients: increased experience but many unanswered questions. *Am J Gastroenterol*. 2020;115(7):1022–1023.
102. Cooper DK, Ezzelarab MB, Hara H, et al. The pathobiology of pig-to-primate xenotransplantation: a historical review. *Xenotransplantation*. 2016;23(2):83–105.
103. Ladowski J, Martens G, Estrada J, et al. The desirable donor pig to eliminate all xenoreactive antigens. *Xenotransplantation*. 2019;26(4):e12504.
104. Cooper DKC, Hara H, Iwase H, et al. Clinical pig kidney xenotransplantation: how close are we? *J Am Soc Nephrol*. 2020;31(1):12–21.

90

Allogeneic Transplantation for Immunodeficiency

Sung-Yun Pai

Following the discovery of the major human leukocyte antigens (HLA) in 1958, hematopoietic stem cell transplantation (HSCT) became a widespread treatment for a variety of congenital and acquired disorders. The cure of an infant with severe combined immunodeficiency (SCID) in 1968 was the first successful experience in HSCT, marking the beginning of a new era in medicine. Shortly thereafter, success was achieved also with HSCT in a child with Wiskott-Aldrich syndrome (WAS).

For many years, the successful use of HSCT in severe primary immune deficiency (PID) was largely restricted to transplantation from HLA-matched sibling donors (MSD), as HSCT from mismatched related family donors (MMRDs), usually haploidentical parents, was followed by severe complications, graft-versus-host disease (GvHD) in particular. In the late 1970s, it was demonstrated in animal models that removal of mature T lymphocytes from the graft obtained from mismatched marrow allowed successful reconstitution upon injection into lethally irradiated recipient animals. This important achievement opened the way to the broader use of HSCT in severe forms of PID. Transplantation of hematopoietic stem cells (HSCs) from adult volunteer matched unrelated donors (MUD), and of umbilical cord blood (UCB), has been increasingly used in individuals with PID. Overall, since 1968, over 2000 transplantations have been performed in patients with PID, most of them in children with SCID.[1,2] The increasing number of transplantations over the years reflects an increased awareness of PID, the broader availability of diagnostic tools (including newborn screening for SCID), improved outcomes as a result of advances in supportive and critical care before and after HSCT, increasingly improved strategies to prevent GvHD, and the greater availability of MUD and UCB for transplantation.

HEMATOPOIETIC STEM CELL TRANSPLANTATION: GENERAL CONSIDERATIONS

Sources of Hematopoietic Stem Cells for Transplantation

HSCs with distinct properties can be retrieved from bone marrow, peripheral blood, or UCB (Table 90.1) (Chapter 2). Bone marrow is the most traditional source and is obtained by multiple small volume (5 to 10 mL) aspirations from along the iliac crests, usually while the donor is under general anesthesia. The volume of marrow obtained is typically 1 L for adults or 10 to 20 mL/kg of the recipient weight. Blood group matching for ABO antigens is not required for HSCT, as mature red blood cells (RBCs) or anti-ABO antibodies can be removed by RBC depletion and plasma depletion, respectively. In the case of HLA-identical transplantation, marrow stem cells are then injected intravenously without further manipulation into the recipient. In the case of mismatched transplantation, bone marrow cells have historically been T-cell depleted *ex vivo*, or *in vivo* administration of cyclophosphamide has been employed to prevent GvHD.

HSCs can also be retrieved from peripheral blood following *in vivo* administration of agents that cause mobilization of stem cells into the peripheral circulation, including granulocyte-colony-stimulating factor (G-CSF) and/or plerixafor. These cytokines are given to the donor typically over 5 days before the harvest of peripheral blood HSCs.

Finally, UCB is another rich source of HSCs. At birth, cord blood is collected in a heparinized medium and stored in liquid nitrogen, with small aliquots preserved for HLA typing. Whenever sufficient compatibility is identified between a patient and stored cord blood, the latter is thawed and injected into the recipient without further manipulation. More recently, *in vitro* expansion of cord blood stem cells, and transplantations with multiple cord blood units, have been used to overcome the limitation of the small volume of cord blood available in the sample.

Donor Selection and Manipulation of the Graft

KEY CONCEPTS

Human Leukocyte Antigen Testing and Matching

- The human leukocyte antigen (HLA) genes, which encode class I and class II major histocompatibility complexes (MHCs), are located in closely linked fashion on chromosome 6. They are inherited as blocks, termed haplotypes, one from each parent.
- HLA genes are inherited in Mendelian fashion. Hence the likelihood of a sibling being fully matched is ~1 in 4. Siblings have a 50% chance of being haploidentical or sharing one haplotype, and a 25% chance of not matching at all. Crossovers within the HLA locus lead to different degrees of matching between siblings.
- HLA class I genes used commonly for typing are HLA-A, HLA-B, and HLA-C, while HLA class II genes are HLA-DRB1, HLA-DQB1, and HLA-DPB1.
- Improved HLA typing has changed the definition of a full match:
  - 6/6 match: matched at both HLA-A, HLA-B, and HLA-DRB1 genes.
  - 8/8 match: matched 6/6 plus both HLA-C genes.
  - 10/10 match: matched 8/8 plus both HLA-DQB1 genes.
  - 12/12 match: matched 10/10 plus both HLA-DPB1 genes.
- Early methods of HLA typing were serological, using antibodies from multiparous women. Current methods employ panels of sequence-specific oligonucleotides or DNA sequencing.
- The greater the degree of HLA mismatch, the higher the chance of both graft rejection and GvHD.

TABLE 90.1 Sources of Hematopoietic Stem Cells for Transplantation

	Bone Marrow	Peripheral Blood Stem Cells	Umbilical Cord Blood
General properties	Used since 1950s Obtained from a postnatal donor	Used since 1980s Obtained from a postnatal donor	Used since 1990s Obtained from newborn umbilical cord
Collection	Bone marrow harvest under general anesthesia	Collection by leukapheresis after administering subcutaneous mobilization agents for 5–6 days	Collection at the time of birth, then cryopreserved
Donor types	Matched related Matched or mismatched unrelated (8–10 of 10) Mismatched related (haploidentical)	Matched related Matched or mismatched unrelated (8–10 of 10) Mismatched related (haploidentical)	Matched related Matched or mismatched unrelated (4–6 of 6)
Infection risk	May transmit bloodborne infections	May transmit bloodborne infections	Practically no risk of transmitting bloodborne infections
Cellular properties	Will contain antigen-specific memory T cells Potent driver of GVHD	Will contain antigen-specific memory T cells Potent driver of GVHD, contains ~1 log more T cells than bone marrow Higher dose of $CD34^+$ cells compared to bone marrow	Little to no antigen-specific memory T cells Less GVHD at a comparable degree of matching ~1 log less nucleated cells and $CD34^+$ cells/kg compared to other sources
Typical minimal cell dose	2×10^8 nucleated cells/kg 2×10^6 $CD34^+$ cells/kg	2×10^8 nucleated cells/kg 2×10^6 $CD34^+$ cells/kg	2.5×10^7 nucleated cells/kg at time of cryopreservation
Median time to neutrophil engraftment	18 days	13 days	23 days
Median time to platelet engraftment	26 days	18 days	60+ days
Additional considerations	May harvest donor repeatedly if needed Volunteer unrelated donor must be willing and pass medical clearance	May harvest donor repeatedly if needed Volunteer unrelated donor must be willing and pass medical clearance	Repeat transplantation from donor is not possible Cells are available immediately frozen Cell dose is usually limiting for adults, may be overcome by using two different donor units Donor is not screened, and could have underlying hematologic or immunologic abnormalities

Hematopoietic Stem Cell Transplantation From a Related Human Leukocyte Antigen–Identical Donor

The use of unfractionated stem cells from an HLA-identical sibling offers the best chance of rapid engraftment and immune reconstitution. In such cases, the HLA-identity between recipient and donor minimizes the risk of GvHD. Furthermore, the mature T cells contained in the graft provide the first line of immune reconstitution after transplantation, as they may expand and lead to a rapid increase in the number of circulating T lymphocytes as early as 2 weeks after HSCT. Engraftment of these mature T cells transplanted with the bone marrow of a matched related donor (MRD) is particularly important in infants with SCID, as it is early evidence of immune reconstitution in a severely immunocompromised host.

Hematopoietic Stem Cell Transplantation From a Haploidentical Donor

Unfortunately, the option of related HLA-identical HSCT is limited to a minority of patients. HSCT from a haploidentical parent was historically the next considered option, particularly in infants with SCID. The rationale for haploidentical HSCT is based on the ability of donor-derived stem cells to repopulate the recipient's vestigial thymus and give rise to fully mature T lymphocytes. Indeed, this is a life-saving procedure that has been successfully applied to several hundreds of infants with SCID.[1,2] However, donor T lymphocytes that would otherwise cause severe GvHD must be eliminated. Several methods are available to attain T-cell depletion.

In the past, the most frequent method was soybean lectin agglutination and E-rosetting. With this method, soybean lectin induced agglutination of the majority of mature marrow cells, which were removed by sedimentation. Further depletion of T lymphocytes was achieved by rosetting with sheep erythrocytes (E-rosetting technique) and density gradient centrifugation. Importantly, T-cell depletion by soybean lectin agglutination and E-rosetting maintains all immature marrow cells in the final preparation. T-cell depletion can also be achieved by incubation of bone marrow with monoclonal antibodies (mAbs) to T lymphocytes plus complement. Campath-1 G, Leu 1, and other mAbs have been used for this purpose, but the degree of T-cell depletion that is achieved with these agents is less complete than with the soybean lectin and E-rosetting, and therefore a higher incidence of GvHD has been reported.

A major advance in graft processing occurred with the development of devices that used antibodies and magnetic beads to positively or negatively select cells of interest. T-cell depletion by graft manipulation is now most frequently achieved either by (1) positive selection of $CD34^+$ cells, enriching for HSC and very immature progenitors, or (2) negative selection of T cells. For $CD34^+$ selection, the CliniMACS device (Miltenyi) employs mAb against CD34 that are directly conjugated to magnetic beads; after binding the anti-CD34 magnetic bead/cell conjugate, release of the magnetic force releases the cells. The CliniMACS device, in general, leads to more efficient and consistent T-cell depletion, often to $<10^4$ $CD3^+$ T cells/kg of recipient

weight. This device is widely used beyond its labeled indication of adults with acute myelogenous leukemia (AML) undergoing MSD transplant, through a humanitarian device exemption. This extremely robust method of T-cell depletion also removes $CD34^-$ progenitors, and other cells (especially stromal marrow cells) that can facilitate stem cell engraftment.

Negative selection of T cells has also been used to preserve progenitors that support short-term engraftment, early post-transplant, and other cells, such as natural killer (NK) cells, that facilitate engraftment and/or improve control of disease in patients with hematologic malignancy. In addition to depletion of $CD3^+$ T cells, more recently, depletion of cells bearing αβ T-cell receptors (TCR) has been increasingly used with success in many PID, including SCID.[3] Agents such as the anti-B cell antibody CD19 can be added to deplete B cells, thereby reducing transmission of Epstein-Barr virus (EBV).

The T-cell depletion methods described above can be very effective at preventing GvHD but come with consequences, including increase in the risk of graft rejection, slowing immune reconstitution, and leaving patients vulnerable to severe viral infections post-transplant. Post-transplant cyclophosphamide (PTCy) has emerged recently as a technique to deplete donor T cells *in vivo* after infusion of unmanipulated haploidentical marrow or peripheral blood stem cells (PBSC). PTCy is typically given on days +3 and +4 at 50 mg/kg. In principle, mature T cells from the donor that are contained within the graft and are directed against recipient alloantigens will vigorously proliferate and be most susceptible to the action of PTCy. The increasing use of PTCy has been fueled by its low cost, particularly in resource-poor areas where the depletion technology and machinery are not available.

Hematopoietic Stem Cell Transplantation From Matched Unrelated Donors

HSCT from MUD has been increasingly used to treat severe PID since 1977. Transplantation from MUD has been facilitated by the increasing number of volunteer donors included in registries worldwide and advances in the techniques for HLA typing, which permit the identification of an optimal MUD and reduction in the risk of GvHD. As of February 2021, more than 38 million donor volunteers and UCB units were included in the World Marrow Donor Association registry. At present, it takes only a few weeks to identify a MUD. However, the probability of finding a suitable donor is lower for ethnic or racial groups that have extreme polymorphism of haplotypes, such as those of African ancestry, or are from populations poorly represented among volunteer donors.

Hematopoietic Stem Cell Transplantation Using Unmanipulated Cord Blood

As opposed to MUD HSCT, which requires identification, willingness, and medical clearance of an adult volunteer, stored cord blood is readily available as a source of stem cells for transplantation. In addition, the risk of GvHD at any given degree of HLA matching is lower when using cord blood, compared with MUD HSCT, so that greater HLA disparity with the recipient can be tolerated. However, the number of cells in a unit is still a major limitation of cord blood. Low cell dose is not usually a problem for HSCT performed in infants because of the low weight of the recipient. Indeed, unrelated UCB transplantation has been successfully used in hundreds of patients with severe PID.[4] In practice, an unrelated HSC donor should be simultaneously sought in cord blood banks and bone marrow donor registries for patients lacking an HLA-identical sibling HSC donor. The option of performing UCB transplants should be based on urgency of the transplantation, the cell dose required, and the number of HLA disparities. Outcomes of transplant for nonmalignant disease for overall survival, graft rejection, and acute GvHD are better with UCB with mismatches at two or more HLA alleles than with MUD with a full match or one allele mismatch.[4]

Principles of Conditioning and Effect on Hematopoietic and Immune Reconstitution

Before HSCT for all malignant and nonmalignant disorders other than SCID, administration of agents to prevent immune-mediated graft rejection and promote engraftment of donor HSC are required (Table 90.2). This pre-HSCT treatment, or 'conditioning,' prevents immunological rejection using agents that target T cells and/or NK cells, and provides 'space' for donor HSC to engraft using agents that damage or kill recipient HSC. Some agents achieve more than one purpose, such as high-dose total body irradiation (1200 to 1400 cGy), which in the case of acute lymphoblastic leukemia has long been favored to eliminate leukemia from sanctuary sites, such as the testes or the central nervous system.

The backbone of conditioning for PIDs other than SCID was historically high-dose busulfan as a stem-cell-directed agent and cyclophosphamide as an immune-cell-directed agent (Bu/Cy). Other alkylating agents that target HSC include melphalan, thiotepa, and treosulfan, the latter approved only in Europe as of 2021. Fludarabine, a nucleoside analog, has now largely replaced cyclophosphamide and generally leads to less organ toxicity than regimens with more than one alkylating agent. Antibodies that deplete immune cells, termed serotherapy, such as anti-thymocyte globulin or alemtuzumab, and low-dose total body irradiation (TBI) (200 to 400 cGy) are also effective for depleting T cells. The likelihood of graft rejection is increased when there is greater HLA disparity (related vs. unrelated donors, matched vs. mismatched donors), pre-sensitization of the host to donor antigens (*e.g.*, through blood transfusions), and when the graft has been depleted of donor T cells. It must be noted, however, that serotherapy agents, due to their long half-life, do not act solely on recipient T cells; if not fully consumed by the time of HSC infusion, these agents will also act on T cells contained within the donor graft. In contrast, TBI has effects only on recipient cells.

Bu/Cy is a typical myeloablative conditioning (MAC) regimen, in that hematopoiesis will not recover unless stem cells are infused. Individualized pharmacokinetic monitoring and adjustment of busulfan dose to achieve a desired area-under-the-curve (AUC) exposure is now standard clinical practice. Reduced toxicity and reduced intensity conditioning (RIC) regimens may employ a lower AUC target of busulfan or other alkylating

TABLE 90.2 Agents Used for Conditioning

Myeloablation	Lymphoablation
Busulfan (may use pharmacokinetic adjustment to achieve a given exposure)	Cyclophosphamide
Melphalan	Fludarabine
Treosulfan	Anti-thymocyte globulin
High-dose total body irradiation (1200–1400 cGy)	Alemtuzumab
Thiotepa	Pentostatin
	Low- or high-dose total body irradiation (as little as 200–400 cGy)

TABLE 90.3 Classification of Common Conditioning Regimens

Myeloablative (MAC)	Busulfan (16 mg/kg or adjust to cAUC 50–70) Cyclophosphamide (200 mg/kg) TBI 1200–1400 cGy Cyclophosphamide or fludarabine or other agents
Reduced toxicity (RTC)	Busulfan (adjust to cAUC 80–90) Fludarabine (120–180 mg/m²) Treosulfan Cyclophosphamide Treosulfan Fludarabine
Reduced intensity (RIC)	Busulfan (adjust to cAUC 45–65) Fludarabine (120–180 mg/m²) +/– anti-thymocyte globulin or alemtuzumab Fludarabine (120–180 mg/m²) Melphalan (100–140 mg/m²) Anti-thymocyte globulin or alemtuzumab
Immunosuppression alone (Minimal intensity)	Fludarabine Cyclophosphamide Anti-thymocyte globulin or alemtuzumab TBI 200 cGy Fludarabine Anti-thymocyte globulin or alemtuzumab

AUC, Area-under-the-curve; *TBI*, total body irradiation.

TABLE 90.4 Measurement of Busulfan Exposure in Different Units

Target Cumulative Exposure (cAUC)	Schedule	Target Range Per Dose: μM × min	Target Range Per Dose: Css ng × min/L
80–90 mg × h/L	Every 6 h × 4 days	~1200–1400	~850–950
	Every 12 h × 4 days	~2400–2800	~850–950
	Daily × 4 days	~4900–5500	~850–950
45–65 mg × h/L	Every 6 h × 4 days	~700–1000	~450–700
	Every 12 h × 4 days	~1400–2000	~450–700
	Daily × 4 days	~2800–4000	~450–700
20–30 mg × h/L	Daily × 2 days	~2400–3600	~400–650

AUC, Area-under-the-curve.

agents thought to be milder than busulfan. RIC regimens are expected, at least some of the time, to result in a mixture of donor and recipient cells, termed mixed chimerism, and autologous recovery after RIC is possible without stem cell infusion. Typical regimens, associated AUC targets, and range of exposure per dose in different units are given in Tables 90.3 and 90.4.

The profound absence of T-cell development and function in patients with SCID makes them unable to reject an allogeneic graft, allowing HSCT to be performed in the absence of conditioning to eliminate host T cells and NK cells. Successful reconstitution of T cells depends on the engraftment of HSC and progenitors in the thymus, leading to sustained neogenesis of T donor-derived cells. After MAC or RIC, donor-derived HSCs engraft in the bone marrow, typically replacing all hematopoietic cells with donor-derived cells; a state termed full-donor chimerism (Fig. 90.1). In contrast, HSCT without conditioning in SCID patients leads to minimal to no HSC engraftment in the bone marrow. While T cells generated in the thymus are donor-derived, all other cells arising from HSC in the bone marrow (B cells, NK cells, neutrophils, monocytes, RBCs, platelets) are recipient-derived; a state termed split chimerism (see Fig. 90.1).

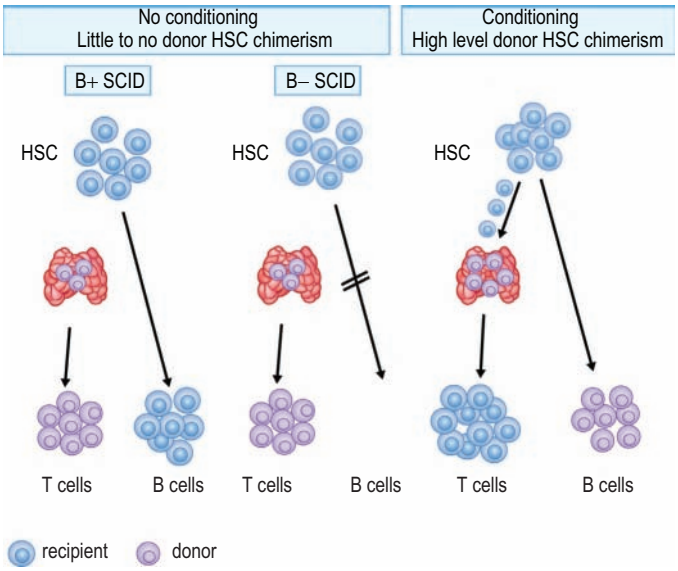


FIG. 90.1 T-cell and B-cell donor chimerism after hematopoietic stem cell *(HSC)* transplantation with or without conditioning in patients with severe combined immunodeficiency *(SCID)*. (With permission from Fig. 1 from Pai S-Y. Treatment of primary immunodeficiency with allogeneic transplant and gene therapy. *Hematology*. 2019;2019[1]:457–465.)

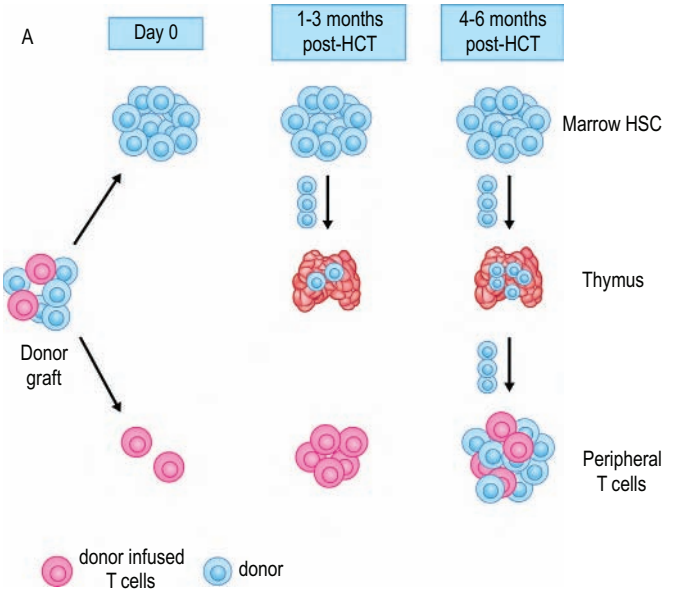


FIG. 90.2 Kinetics of extrathymic versus thymic reconstitution of donor-derived T cells post-hematopoietic cell transplantation *(HCT)*. *HSC*, Hematopoietic stem cell. (With permission from Fig. 2A from Pai S-Y. Treatment of primary immunodeficiency with allogeneic transplant and gene therapy. *Hematology*. 2019;2019[1]:457–465.)

The kinetics of T-cell reconstitution proceeds in two phases, a thymus-independent phase, followed by a thymus-dependent phase (Fig. 90.2). When the graft is first infused, mature T cells

contained in the graft expand, driven by the lymphopenia of the host and cytokines such as IL-7. This occurs over the first 1 to 3 months after HSCT, and it is these donor-derived allogeneic T cells that may spark GvHD. Donor-derived progenitors developing in the thymus meanwhile give rise to naïve T cells, which are selected to be tolerant of the host thymic epithelium. This process takes much longer, 4 to 6 months, and may be impeded due to age, or medication-related damage to thymic epithelium.

KEY CONCEPTS

Principles of Conditioning

- Currently, agents used to target recipient HSCs are either alkylating agents or irradiation (see Table 90.2). As such, these agents cause DNA damage to non-hematopoietic tissues, leading to acute toxicities such as mucositis and veno-occlusive disease of the liver.
- Potential long-term effects of alkylating agents and irradiation are also related to DNA damage and include decreased final height, hypothyroidism, delayed puberty due to hypogonadism, infertility (low sperm count in boys, premature ovarian failure in girls), learning problems, and secondary malignancies (typically myelodysplastic syndrome or soft tissue or bony tumors).
- Cyclophosphamide, an alkylating agent, was the most commonly used immunosuppressive agent for the prevention of rejection. More recently, several other agents, that are not alkylating agents, may be used, which should have fewer long-term toxicities.
- The composition of the conditioning regimen, in general, must be adjusted to the donor type because of the role of donor T cells in promoting engraftment. The immunologic activity of the donor T cells against recipient T cells and recipient marrow facilitates engraftment. If the graft is T-cell depleted, as is commonly done for MMRD haploidentical HSCT, more immunosuppression must be included to prevent recipient T cells from rejecting the graft.
- Agents such as anti-thymocyte globulin (ATG) and alemtuzumab are antibodies that target T cells and/or other lymphocytes. These agents have a half-life of weeks, with variable clearance depending on the lymphocyte content in the host. Even when given before transplant, these agents may be circulating in sufficient quantity to have effects on donor T cells infused with the graft. The dosing and timing are thus critical to control the effect of these agents as intended.

Complications of Hematopoietic Stem Cell Transplantation

A variety of complications can compromise the success of HSCT. Incompatibility between donor and recipient can lead to graft rejection by the host immune system or GvHD caused by alloreactivity of donor-derived lymphocytes to the recipient's cells. Conditioning regimens can cause the toxicity of several organs. The frequency and severity of these complications depend on the type of donor, the type and intensity of the conditioning regimen, specific considerations related to the underlying disorder, and to the clinical status of the recipient before transplantation.

Acute Graft-Versus-Host Disease

Acute GvHD (aGvHD) is the result of alloreactivity of donor-derived T lymphocytes against the recipient's antigens and is one of the most severe complications of HSCT. It may occur as early as 1 week after HSCT and is potentially fatal. Clinical manifestations of a GvHD include maculopapular skin rash (that tends to be confluent), diarrhea, and liver abnormalities (hepatomegaly, elevated liver enzymes, increased levels of conjugated bilirubin).[5] The disease may progress to severe skin manifestations with exfoliative dermatitis and significant liver and gut damage (with intractable watery or bloody diarrhea, protein-losing enteropathy, and abdominal pain). Bone marrow aplasia and high susceptibility to infections are also often observed in severe aGvHD.

TABLE 90.5 Staging and Grading of Acute Graft-Versus-Host Disease

Stage	Skin	Liver (Bilirubin)	Gastrointestinal System (Stool Output)
0	None	<2 mg/dL	Adult: <500 mL/day Child: <10 mL/kg/day
1	Rash <25% BSA	2–3 mg/dL	Adult: 500–999 mL/day Child: 10–19.9 mL/kg/day or persistent nausea, vomiting, or anorexia, with a positive upper GI biopsy
2	Rash 25%–50% BSA	3–6 mg/dL	Adult: 1000–1500 mL/day Child: 20–30 mL/kg/day
3	Rash >50% BSA (generalized erythroderma)	6.1–15 mg/dL	Adult: >1500 mL/day Child: >30 mL/kg/day
4	Generalized erythroderma with bullous formation	>15 mg/dL	Severe abdominal pain with or without ileus

Grade	
Grade 0	No stage 1–4 of any organ
Grade 1:	Stage 1 or 2 skin involvement; no liver or gut involvement
Grade 2:	Stage 1–3 skin involvement; Grade 1 liver or gut involvement
Grade 3:	Stage 2 or 3 skin involvement
Grade 4:	Stage 1–4 skin involvement; stage 2–4 liver or gut involvement

The severity of aGvHD is evaluated according to a grading scale (Table 90.5). Major risk factors for aGvHD include HLA-mismatch between donor and recipient, older age of the recipient or donor, gender mismatch, and stem cell source.[5] However, aGvHD may also be observed following related HLA-identical HSCT, particularly when a conditioning regimen is used.

Finally, transfusion-associated aGvHD is a very severe complication after HSCT, which can be effectively prevented by using irradiated (1500 to 3000 rad) and filtered blood derivatives.

Chronic Graft-Versus-Host Disease

Chronic GvHD (cGvHD) has traditionally been defined as symptoms that persist or appear 100 days after transplantation. cGvHD is better defined by its distinct clinical manifestations rather than the time of onset alone.[6] These clinical manifestations include skin changes (scleroderma-like lesions, hyperpigmentation, hyperkeratosis, skin atrophy, ulcerations), tissue fibrosis and limitation of joint mobility, fibrosis of exocrine glands ("sicca syndrome"), fibrosis of lungs and liver, increased susceptibility to infections, immune dysregulation, and autoimmunity. Consequently, cGvHD may have a major effect on quality of life and can be fatal. Although the incidence of cGvHD is lower in children than in adults treated by allogeneic HSCT, the risk factors and the spectrum of clinical manifestations are similar.

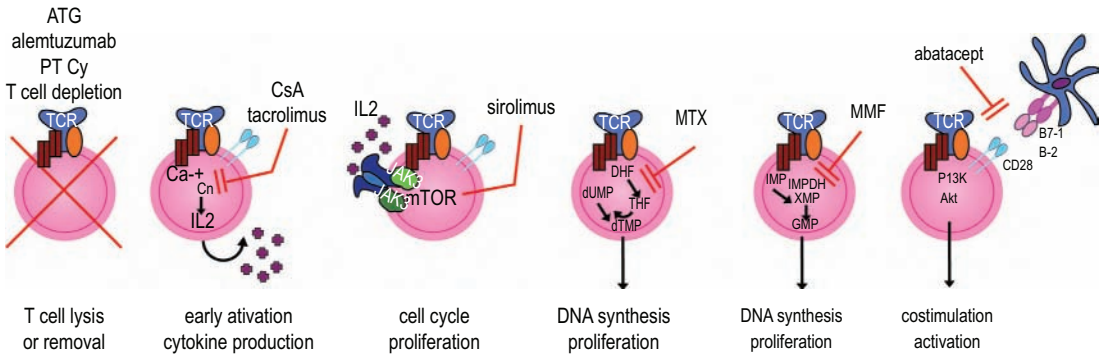


FIG. 90.3 Mechanism of action of medications or manipulations commonly used for graft-versus-host disease *(GvHD)* prophylaxis. Agents such as anti-thymocyte globulin *(ATG)*, alemtuzumab, post-transplant cyclophosphamide *(PTCy)*, or graft manipulation via T-cell depletion eliminate T cells by lysis or removal. Cyclosporine A *(CsA)* and tacrolimus inhibit calcineurin *(Cn)* a calcium (Ca^{2+})-dependent phosphatase triggered by T-cell receptor *(TCR)* stimulation by antigen, leading to elaboration of early cytokines including interleukin-2 *(IL2)*. Sirolimus inhibits mammalian target of rapamycin *(mTOR)*, a key signaling molecule downstream of cytokine receptors including the IL2 receptor which signals via Janus kinases 1 and 3 (JAK1, JAK3) to induce cell cycle entry and proliferation. Methotrexate *(MTX)* inhibits dihydrofolate reductase, which catalyzes the conversion of dihydrofolate *(DHF)* to tetrahydrofolate *(THF)*. Conversion of THF back to DHF by thymidylate synthase catalyzes the conversion of deoxyuridine monophosphate *(dUMP)* to deoxythymidine monophosphate *(dTMP)* needed for DNA synthesis and cell proliferation. Mycophenolate mofetil *(MMF)* inhibits inosine monophosphate dehydrogenase *(IMPDH)*, which catalyzes the conversion of inosine monophosphate *(IMP)* to xanthine monophosphate *(XMP)*, a key step in generating guanosine monophosphate *(GMP)* needed for DNA synthesis and cell proliferation. Abatacept is a fusion of CTLA4 with immunoglobulin that binds B7-1 and B7-2, ligands for CD28, whose engagement costimulates T cells triggered by antigen binding to the TCR and leads to T-cell activation.

Acute GvHD represents a major risk factor for cGvHD, yet cGvHD can be observed even without preceding aGvHD, and when present, it does not represent merely the continuation of aGvHD. The use of peripheral blood stem cells carries an increased risk of cGvHD compared with the use of bone marrow stem cells.

Prevention of GvHD

Prevention is the most effective approach to GvHD; the mechanism of action of agents commonly used for prevention and/or treatment of GvHD largely targets T cells (Fig. 90.3). Use of a fully matched donor remains the best, though not assured, method of prevention, but of course, it is not available to most patients. If a related HLA-mismatched donor is used for transplantation, vigorous T-cell depletion must be performed either *ex vivo* (using graft manipulation) or *in vivo* (using PTCy). Whenever a conditioning regimen is used in the transplantation protocol, pharmacological prophylaxis of GvHD must also be included, even in the case of HSCT from a related HLA-identical donor.

Long-standing approaches to prevention of GvHD include the use of a calcineurin inhibitor, either cyclosporine or tacrolimus for 6 months post-HSCT in combination with methotrexate (15 mg/m^2 on the first day after HSCT, and then 10 mg/m^2 at days +3, +6, and +11 after transplantation). Mycophenolate mofetil and sirolimus can be used instead of methotrexate. Serotherapy with ATG or alemtuzumab is often included in the preparatory regimen, to reduce the risk of GvHD. The timing of administration of alemtuzumab as part of the preparatory regimen has opposed effects on the prevention of aGvHD and the speed of T-cell reconstitution. In particular, use of alemtuzumab in the days that immediately precede transplantation has more potent effects on aGvHD prevention but is associated with delayed T-cell reconstitution, whereas the opposite is observed when the drug is administered a few weeks before HSCT. Finally, abatacept, a fusion protein of CTLA4 and Ig which disrupts engagement of CD28 by its ligands, has received breakthrough designation from the FDA for prevention of aGvHD (see Fig. 90.3). While in general, MUD grafts are infused unmanipulated with pharmacological prophylaxis, PTCy and TCRαβ/CD19 depletion have also been used with MUD grafts, particularly for mismatched grafts.

Treatment of Graft-Versus-Host Disease

Once GvHD has developed, treatment is based mainly upon the use of immunosuppressive drugs. Corticosteroids remain the first-line therapy and are usually effective, especially for mild and moderate aGvHD. A multitude of second-line agents for steroid-dependent or steroid-refractory disease include pentostatin, etanercept, sirolimus, daclizumab, and basiliximab; but few if any have shown dramatic efficacy. In 2019, ruxolitinib, an inhibitor of JAK1 and JAK2 kinases, became the first FDA-approved agent in the United States for treatment of steroid-refractory aGvHD for adults and pediatric patients ≥12 years of age; a milestone in the field.

Our understanding of the cellular mechanisms of cGvHD is growing, though therapies remain woefully limited in their efficacy.[7] Treatment of cGvHD is based on immunosuppression, either by inhibiting T cells, inhibiting B cells, or expanding regulatory T cells (Treg). Topical steroids and calcineurin inhibitors may alleviate mucosal and skin symptoms. Systemic steroids have been shown to improve survival, but at the risk of significant adverse effects. Extracorporeal photopheresis can be used to induce tolerance; typically, its benefits, if present, are delayed until 2 to 3 months after initiation of treatment. The promise of inhibiting B cells was realized in 2017 when ibrutinib was the first agent to be FDA-approved in the United States

for the treatment of refractory cGvHD. Many other agents including mycophenolate mofetil, anti-TNF-α mAb, etanercept, anti-CD20 antibodies, sirolimus, ruxolitinib, and chronic low-dose IL-2 have also been used with variable efficacy.

Infections

Infections are a major complication following HSCT. Patients with severe PID are intrinsically highly susceptible to infections. In infants with SCID and with other forms of combined immune deficiency, viral and opportunistic infections can develop before transplantation and are one of the factors that adversely affect the outcome of HSCT itself.[1,2,8] Regardless of the type of underlying PID, T-cell-depleted HSCT carries a high risk of infections because of the longer time required to achieve immune reconstitution. Furthermore, conditioning regimens with myeloablative and immunosuppressive drugs, and GvHD prophylaxis, increase susceptibility to infections after HSCT. Strict isolation of the patients during and after HSCT and prophylactic administration of antibiotics have been associated with better survival, particularly after related HLA-mismatched transplantation for SCID.[1]

Challenging viruses in infants with SCID include adenovirus, cytomegalovirus (CMV), parainfluenza type III virus, and EBV. In particular, CMV infection after HSCT can cause interstitial pneumonia, enteritis, hepatitis, and encephalitis. EBV can also cause lymphoproliferative disease. The risk of transfusion-associated infections can be reduced by the removal of leukocytes or selecting CMV-negative donors. Several antiviral drugs (acyclovir, ganciclovir, foscarnet, cidofovir, letermovir) are now available and are effective, especially against CMV[9]; and preemptive administration of anti-CD20 mAb helps to prevent EBV-driven lymphoproliferative disease. Furthermore, transfusion of virus-specific cytotoxic T lymphocytes (CTL), either donor-derived, autologous engineered, or third party, represents another important resource to fight severe viral infections.[10]

Pneumocystis jiroveci is a common cause of pneumonia in severely immunocompromised patients. The first-line treatment is intravenous sulfamethoxazole/trimethoprim.

Aspergillus infection is a severe complication in patients with chronic granulomatous disease (CGD) and patients with profound neutropenia. Voriconazole offers some advantage compared with liposomal amphotericin B for treatment of invasive aspergillosis, whereas prophylactic itraconazole reduces the incidence of fungal infections in patients with CGD before transplantation.

Bacterial infections are usually amenable to successful treatment if the pathogen is identified and aggressive use of appropriate antibiotics initiated. Prophylactic immunoglobulin (IgG) infusions following HSCT also reduce the frequency and severity of infections. Administration of Bacille Calmette-Guérin (BCG) vaccine at birth to prevent tuberculosis is still practiced in many countries worldwide. In severely immunocompromised infants who undergo HSCT, this live vaccine may cause disseminated disease, which often manifests at the time of engraftment with local and systemic immune reconstitution inflammatory syndrome (IRIS).

Toxicity Related to Conditioning Regimen

Chemotherapeutic agents in the conditioning regimen often cause significant short-term and long-term toxicity. Myeloablative regimens cause temporary anemia, thrombocytopenia, and leukopenia. Consequently, supportive treatment with RBC and platelet transfusions is necessary during the aplastic phase. Conditioning damages the lining of the intestine, leading to painful mucosal ulcers and need for nutritional support for several weeks. Finally, severe leukopenia, in conjunction with loss of intestinal barrier function, predisposes to life-threatening bacterial or fungal infections until neutrophil engraftment and healing occur. Chemotherapeutic drugs that damage the liver vascular endothelium, particularly busulfan and cyclophosphamide, can cause veno-occlusive disease (VOD), which is clinically marked by painful hepatomegaly, jaundice, ascites, fluid retention, weight gain, and can ultimately result in fatal multi-organ failure. Defibrotide is the most effective agent studied to date in the treatment of VOD. Busulfan can also cause seizures and lung damage. Cyclophosphamide can cause hemorrhagic cystitis, a syndrome of inappropriate antidiuretic hormone secretion, or more rarely, cardiac disturbances.

Long-term hormonal complications are more common when TBI is used. However, the busulfan and cyclophosphamide regimen can cause delayed puberty or sterility, and thyroid dysfunction is frequently observed, even in patients who have not received TBI. Delayed or incomplete tooth eruption is also a possible consequence of conditioning regimens given to infants or young children. Effects on final height and growth, as well as long-term neurocognitive effects, are emerging as more children surviving after treatment are followed up with. Patients with defects of DNA repair (*e.g.*, some forms of SCID) are particularly at risk for toxicities related to conditioning.

Hematopoietic Stem Cell Transplantation forv Severe Combined Immunodeficiency

General Considerations

SCID is a medical emergency and is uniformly fatal unless promptly diagnosed and successfully treated. While the alternative strategies, including gene therapy (Chapter 91), are gaining prominence, allogeneic HSCT remains the standard treatment for these disorders.

The virtual lack of T lymphocytes strongly impairs the ability of a SCID patient to reject the graft. Furthermore, donor-derived lymphoid progenitor cells have a striking advantage for *in vivo* T-cell differentiation. Use of pre-transplantation chemotherapy and immune suppression is thus not required to attain T-cell reconstitution for SCID. There is a general consensus that no conditioning is needed for HSCT from an HLA-identical related donor. In the last two decades, considerable advances have been made in the field of allogeneic transplantation, including improvements in HLA typing, in unrelated donor transplant techniques, T-cell depletion, prevention and treatment of infection, early diagnosis of SCID through universal newborn screening, and definition of the genes responsible for ~90% of cases of SCID. These advances, in combination with results published by multi-institutional consortia, the Inborn Errors Working Party of the European Blood and Marrow Transplant Society (IEWP-EBMT), and the Primary Immune Deficiency Treatment Consortium (PIDTC) in North America, have changed how HSCT is performed for SCID, leading to expanded donor options and tailoring of conditioning to genotype.

CLINICAL PEARLS

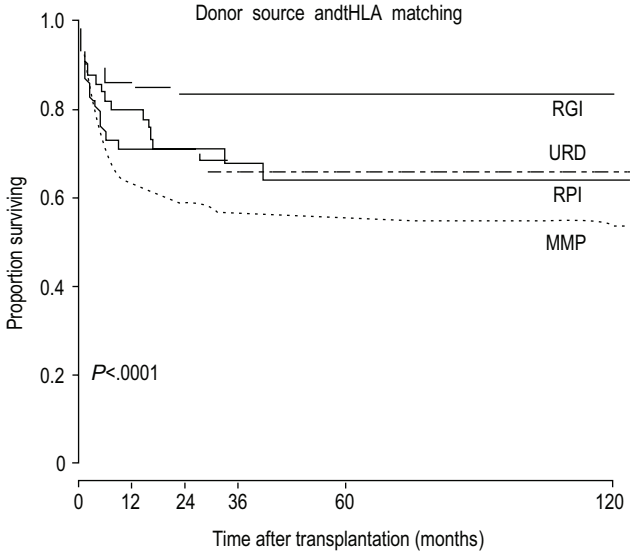
Unique Considerations of Stem Cell Transplantation for Severe Combined Immunodeficiency Versus Other Primary Immunodeficiencies

- Patients with SCID have little or no ability to reject allogeneic cells. Therefore no chemotherapy is required to achieve T-cell reconstitution following HSCT.
- All other patients must receive at least some conditioning to prevent rejection and remove enough recipient HSCs to allow donor HSC engraftment.
- After HSCT, early T-cell reconstitution in the first few months relies on the expansion of mature T cells infused with the graft. In SCID patients receiving MSD grafts without conditioning, graft-versus-host disease (GvHD) rarely occurs. In contrast, for patients undergoing HSCT with conditioning, medications to prevent GvHD or T-cell depletion must be used.
- Generation of naïve T cells that are tolerant to the recipient relies on differentiation of donor progenitors in the thymus, a process requiring 4 months or more after HSCT.
- Patients undergoing T-cell-depleted HSCT, via either graft manipulation or post-transplant cyclophosphamide, are dependent on thymus-dependent T-cell reconstitution.
- The longevity and quality of T-cell reconstitution in SCID patients after unconditioned HSCT is, in general, inferior to HSCT with conditioning. High-level engraftment of donor-derived HSC, generally achieved only after conditioning, leads to sustained T-cell generation.

Survival Following Hematopoietic Stem Cell Transplantation for Severe Combined Immunodeficiency

Two large retrospective reports of patients with SCID treated by HSCT have been published, the IEWP-EBMT cohort of 699 cases, and the PIDTC cohort of 663 cases.[1,2] The IEWP study reported on patients transplanted between 1968 and 2005, of who 203 patients received HSCT from MSD ($n=135$) or phenotypically matched other relatives ($n=68$), 415 from MMRD, and 81 from MUD. As expected, 10-year survival was best after MSD HSCT at 84% and superior to all non-MSD graft sources, which were similar in outcome, with 64%, 54%, and 66% survival, respectively (Fig. 90.4). Important factors this study defined in multivariate analysis as hurting survival included older age at transplantation (>12 months of age), poor clinical status of the patient (respiratory impairment, septicemia, viral infection), T-cell depletion of the graft, lack of supportive measures (protected environment, antimicrobial prophylaxis), and SCID phenotype (B^+ SCID being more favorable than B^- SCID). There was also a significant improvement in survival in the 2000 to 2005 era compared to the pre-1995 era (see Fig. 90.4).

The PIDTC report of Haddad *et al.*[8] spanned an overlapping timeframe from 1982 to 2012, confirming many findings in the IEWP study, while also extending and expanding on factors found to be important in a smaller report of 240 patients transplanted between 2000 and 2009. As expected, MSD recipient survival was superior at 94% compared to all other donor sources, and no single non-MSD graft type was superior to another. Analysis focused on the 571 patients receiving non-MSD grafts (phenotypically matched other related donors, MMRD, MUD) confirmed the importance of young age and freedom from infection, in that the best survival was achieved in patients <3.5 months of age at HSCT, and those of any age who had never been infected; those >3.5 months of age who had an active infection or even infection that had resolved, had poorer survival. Importantly a survival advantage was shown in this study according to the underlying genetic cause of SCID, with patients with defects in *IL2RG*, *JAK3*, *IL7R/CD3/CD45*, and *RAG1/RAG2* defects surviving better than those with ADA-deficient SCID or Artemis-deficient SCID due to defects in the *DCLRE1C* gene (Fig. 90.5). Patients with unknown genotypes also had relatively poor survival. Neither the use of conditioning nor whether the patient had typical SCID versus leaky SCID or Omenn syndrome had any effect on survival.[2]



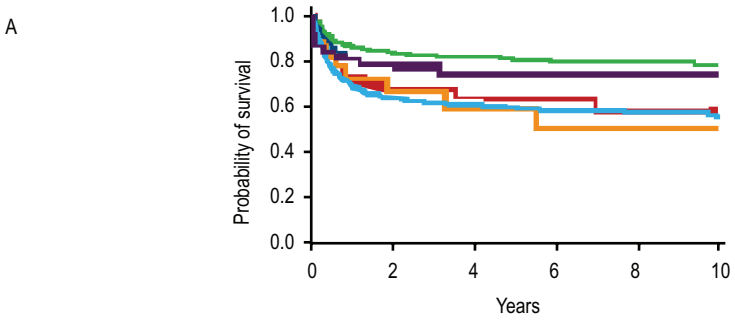
Months	0	6	12	24	36	60	120
Number at risk							
RGI	135	111	77	64	60	50	31
RPI	68	46	36	27	21	16	8
URD	81	53	36	33	23	18	6
MMR	415	259	207	175	154	111	46

FIG. 90.4 Cumulative probability of survival in patients with severe combined immunodeficiency after transplantation according to donor type and the period of transplantation. Survival after related genotypical identical *(RGI)*, related phenotypical identical *(RPI)*, unrelated *(URD)*, and mismatched related *(MMR)* donors is shown. (Redrawn after Fig. 1 of Gennery AR, Slatter MA, Grandin L, et al. Transplantation of hematopoietic stem cells and long-term survival for primary immunodeficiencies in Europe: entering a new century, do we do better? *J Allergy Clin Immun.* 2010;126[3]:602–610.e11.)

Quality and Kinetics of T-Cell Immune Reconstitution

The most critical goal of HSCT in SCID patients is a normalization of the number and function of T cells, which rescue the patient from death due to opportunistic infection. Factors that affect the quality and kinetics of T-cell immune reconstitution include the impact of donor type and HLA match, graft manipulation, use of conditioning, and underlying genetic cause of the SCID.

The high degree of success of MSD transplant without conditioning reflects the rapid reconstitution achieved when mature T cells are included in the graft and no immunosuppression, as part of conditioning or GvHD prophylaxis, is given. Mature HLA-matched T cells contained within the stem cell graft when infused into the lymphopenic SCID patient



No. at Risk						
ADA	40	23	18	13	12	10
DCLRE1C	18	12	8	6	5	4
γC/JAK3	198	159	127	102	85	68
IL7Ra/CD3/CD45	43	32	27	20	13	11
RAG	38	24	18	14	7	3
Other(PNP)/Unknown/ND	234	135	106	85	75	59

FIG. 90.5 Cumulative probability of survival in patients with severe combined immunodeficiency after transplantation according to genotype. Survival for patients with mutations in adenosine deaminase *(ADA)*, DNA cross-link repair protein 1C *(DCLRE1C)*, common gamma chain or interleukin-2 receptor gamma *(γC)*, Janus kinase 3 *(JAK3)*, interleukin-7 receptor *(IL7R)*, CD3 subunits, CD45, recombination-activating genes 1 and 2 *(RAG)*, other genes including purine nucleoside phosphorylase *(PNP)*, unknown genotype or not done *(ND)* are shown. (Redrawn after Fig. 2 of Haddad E, Logan BR, Griffith LM, Buckley RH, et al. SCID genotype and 6-month posttransplant CD4 count predict survival and immune recovery. *Blood*. 2018;132[17]:1737–1749.)

vigorously expand in response to high levels of cytokines such as IL-7. These donor T cells are detectable within 1 to 3 months and begin to mediate antigen-specific responses to their cognate antigens. Such T cells are primarily of a memory phenotype. In contrast, newly generated, naïve ($CD45RA^+$ $CD31^+$) T lymphocytes from donor-derived progenitors engrafting in the thymus do not appear in the peripheral circulation until 4 to 6 months after HSCT (see Fig. 90.2), irrespective of the type of transplant (HLA-identical or mismatched); their number tends to peak approximately a year after HSCT, by which time a fully polyclonal T-cell repertoire is usually observed. These naïve T lymphocytes are the product of ongoing active thymopoiesis and contain T-cell receptor excision circles (TRECs). TRECs are extrachromosomal DNA episomes generated during V(D)J recombination (Chapters 32 and 34) and are not duplicated during mitosis. TRECs are now used broadly as a method to identify babies with SCID at birth.[11] The rapid early reconstitution with mature lymphocytes followed by robust thymopoiesis and generation of T cells that are immediately functional, unhampered by GvHD prophylactic immunosuppression, underlie the success of this approach.

The speed and quality of T-cell reconstitution, after all other types of grafts, can be predicted based on the procedures used and the intrinsic properties of non-MSD grafts. MUD and UCB transplants are typically performed without T-cell depletion; however, the use of serotherapy to prevent GvHD is common, which in turn depletes mature T cells that could otherwise expand and mediate early immunity. Furthermore, T cells contained within UCB are antigen-inexperienced, which favors their safety for mediating GvHD but means that they will not provide antigen-specific immunological memory responses. MMRD grafts, the majority of which are haploidentical, require *ex vivo* or *in vivo* T-cell depletion, either by graft manipulation or post-transplant administration of PTCy. In such cases, T-cell reconstitution is almost exclusively dependent on thymopoiesis, which, as stated, occurs over many months. Indeed in the PIDTC cohort, MMRD recipients were more likely than MUD recipients to need another transplant, presumably due to poor engraftment of the initial one; whether this was primarily due to HLA mismatch or due to T-cell depletion of the graft is uncertain.[2] Finally, the health of the thymic epithelium must be considered. Transplantation performed early in life (at <3.5 months of age) leads to superior thymic output. This may reflect a lack of thymic damage (which is often observed in older infants after infections). Alternatively, it is possible that a younger thymus better supports thymopoiesis.

Conditioning using stem cell-targeted agents (RIC or MAC) was also associated with superior T-cell reconstitution. Engraftment of donor-derived long-term HSC achieved with RIC or MAC allows for continued seeding of the thymus with progenitors, which is thought to sustain thymopoiesis after the initial directly seeded short-term progenitors have died out. The study of genotype-specific outcomes lends support to the concept that T-cell reconstitution is influenced by the biology of the defect and the nature of the block in T-cell development. When compared to patients with *IL2RG/JAK3* defects, patients with RAG or Artemis SCID undergoing non-MSD transplant were less likely to have robust T-cell reconstitution 2 to 5 years post-HSCT, independent of conditioning.[2,12] One likely explanation is that arrest of thymocyte development in RAG and Artemis SCID, which both affect V-D-J recombination, interferes with the ability of donor-derived progenitors to engraft, particularly in the absence of conditioning. T-cell neogenesis is further compromised in the unconditioned setting due to a loss of the initially engrafted progenitors and lack of donor-derived HSC in the marrow to replace them. The presence of autologous NK cells in these NK^+ forms of SCID may also contribute to graft rejection. Finally, in a series of 106 patients with SCID caused by adenosine deaminase deficiency (ADA^- SCID), unconditioned transplantation from MMRDs was associated with a high risk of graft failure.[13]

CLINICAL PEARLS

Considerations related to underlying disease, donor match, and clinical condition

- Historically, only SCID and severe aplastic anemia were immediate indications for HSCT.
- Accumulated experience and improvements in tissue typing, supportive care, and GvHD prevention and management have made a cure with HSCT possible for more diseases.
- Lack of an MSD no longer precludes HSCT if a well- MUD can be found.
- Diseases for which HSCT is now indicated, include WAS, CGD, immune dysregulation, polyendocrinopathy, enteropathy, X-linked (IPEX), and other combined immunodeficiencies.
- Diseases that have both hematopoietic and non-hematopoietic manifestations should be approached with caution, as only the hematopoietic manifestations will be corrected by HSCT.
- The timing of HSCT depends on the natural history of the disease, the patient's age, existing morbidities, and time needed to secure a donor and optimize the patient's clinical condition.
- Organ damage already incurred is unlikely to be reversed by HSCT and increases the risks of this treatment.

Reconstitution of B- and NK-Cell Immunity

In contrast to T lymphocytes, reconstitution of B cells and B-cell function after HSCT for SCID is varied, and often incomplete. Factors now known to be associated with better B-cell reconstitution include the use of pre-transplant conditioning, donor type, and underlying genotype.

The PIDTC retrospective cohort study showed that certain genotypes are at risk of poor B-cell function post-HSCT, though not the exact same ones at risk of poor T-cell reconstitution. SCID patients with mutations in *RAG1/RAG2* and *DCLRE1C* are at higher risk of poor T-cell and B-cell function (the latter measured as freedom from immunoglobulin infusions) 2 to 5 years post-HSCT.[2] RIC or MAC are associated with improved T-cell reconstitution, as noted above, and also with a higher likelihood of B-cell function 2 to 5 years post-HSCT, particularly among recipients of non-MSD grafts.[2,8,14] These factors are independent predictors of B-cell function, but are interrelated in that only donor-derived HSC would be capable of differentiating into B cells in these two forms of B^- SCID and conditioning with RIC or MAC is required to engraft donor HSC in the bone marrow.

In contrast, patients with SCID due to defects *IL2RG* or *JAK3* have excellent T-cell reconstitution, but poor B-cell function compared to patients with defects in *IL7R/CD3/CD45* or ADA-deficient SCID. In this case, B-cell development is not the issue, but rather that IL2RG/JAK3-deficient B cells fail to respond *in vitro* to IL-21, an IL2RG-dependent cytokine, and consequently have poor class-switching and fail to secrete antibodies.[15] As a result, the rate of freedom from immunoglobulin infusions of patients with *IL2RG/JAK3* SCID, after non-MSD transplant, is much enhanced with the use of conditioning (16.7% vs. 74.2%).[2]

Finally, while all non-MSD recipients have similar T-cell reconstitution at 2 to 5 years post-HSCT, MMRD grafts have inferior B-cell function compared to those from phenotypically matched relatives and MUD, independent of genotype or conditioning. The reasons for this are not clear.

More limited data are available about the reconstitution of NK-cell function. In patients with NK^- SCID, NK cells are often the first cells to appear after haploidentical HSCT. However, lower NK-cell counts are observed at long-term follow-up after HSCT in patients with *IL2RG* or *JAK3* defects.

Complications Following Hematopoietic Stem Cell Transplantation for Severe Combined Immunodeficiency

Despite advances in prophylaxis and treatment, infections (especially those caused by viruses) remain a significant cause of death after HSCT for SCID. In a report of 166 transplantations performed at their center, Buckley *et al.*[16] indicated that viral infections, including CMV, adenovirus, EBV, enteroviruses, parainfluenza virus type 3, varicella, herpes simplex, and respiratory syncytial virus (RSV), accounted for 30 of the 40 deaths observed. In the PIDTC report, a total of 107 patients had infections associated with death, dominated by CMV, EBV, adenoviruses, parainfluenza, fungi, and bacteria.[2]

Viral and opportunistic infections are the most common early after HSCT, especially in recipients of T-cell-depleted haploidentical HSCT, because of the delay in achieving immune reconstitution. Incomplete recovery of immune function at 1 year after HSCT is associated with a higher risk of late infections. In a single-center study of 90 patients with SCID treated with HSCT, 11 (12%) developed significant infectious complications 2 to 17 years afterward.[17] Among late infections, chronic skin warts caused by papillomavirus have been observed in a significant fraction of infants with *IL2RG* or *JAK3* defects after HSCT.[17] This complication may also occur in patients who attain robust immune function and may result from signaling defects that involve extra-hematopoietic cells, such as keratinocytes.

GvHD is another major complication of HSCT for SCID. In the PIDTC series, cumulative incidence of grade 2 to 4 aGvHD and cGvHD were 23% and 16%, respectively.[2] The risk of aGvHD was related to conditioning, and possibly to maternal engraftment. While donor type and use of T-cell depletion did not impact aGvHD, in this series cGvHD was more often seen in patients who had received MMRD HSCT that was T-cell depleted by methods other than those that were standard at the time, soybean lectin agglutination and E-rosetting or $CD34^+$ selection.[2] Chronic GvHD has been reported in 10 (11%) of 90 patients who survived for at least 2 years after HSCT for SCID in Paris.[17] Six of them developed disseminated cGvHD, and 3 died of cGvHD and related infectious complications. Effective prevention of GvHD remains of paramount importance in SCID patients, who experience no benefit from it, in contrast to patients with leukemia for whom a graft-versus-leukemia (GvL) effect can be helpful.

Immune dysregulation, autoimmunity, and need for prolonged nutritional support represent additional complications of HSCT for SCID. Neven *et al.* have reported that among 90 long-term survivors after HSCT for SCID, 12 suffered from autoimmune and inflammatory complications more than 2 years after HSCT for SCID, and in 6 of them, the onset of such complications was within the first 2 years after transplantation.[17] These late manifestations of immune dysregulation are often associated with incomplete immune reconstitution and may lead to a poor outcome. The need for prolonged nutritional support was more frequent among patients treated by mismatched related or unrelated donor HSCT, especially if cGvHD, immune dysregulation, and poor immune reconstitution were also present. Infants with DCLRE1C (Artemis) deficiency (with defective DNA repair) are at a higher risk for late complications, including growth retardation, requirement for nutritional support, and dental abnormalities.[12,17]

Hematopoietic Stem Cell Transplantation for Combined Immunodeficiencies Other Than Severe Combined Immunodeficiency

Other predominant T-cell immunodeficiencies, such as purine nucleoside phosphorylase deficiency, cartilage hair hypoplasia, and other forms of T-cell activation deficiency can be treated with HSCT; overall, survival after HSCT for these disorders is worse than in typical SCID (approximately 50%).[1]

MHC class II deficiency remains a very difficult disease to treat with transplantation, although there have been substantial improvements over time. Compared to the survival of 30% to 40% in the 20th century, the survival of 19 patients in a single institution transplanted after 2008 was 94%.[1,18] Many patients with this deficiency fail to reconstitute a normal number of circulating CD4 T lymphocytes, probably because the lack of expression of HLA class II molecules on thymic epithelial cells interferes with the positive selection of CD4 lymphocytes.[18]

For patients with complete DiGeorge syndrome, HSCT or even transplantation of unmobilized peripheral blood mononuclear cells may be attempted if there is an HLA-identical donor; in such cases, immune reconstitution is provided by mature T lymphocytes contained in the graft. However, experimental implantation of allogeneic thymic tissue has become the preferred treatment, though not widely available at this time.[19,20]

Patients with CD40 ligand (*CD40L*) deficiency suffer from recurrent bacterial and opportunistic infections, particularly with *Pneumocystis jiroveci* and *Cryptosporidium parvum*, resulting in increased mortality in childhood and young adulthood. The early experience of HSCT for these patients was suboptimal, with only 26 of 38 patients transplanted in Europe surviving.[21] An updated experience of 106 patients transplanted after 2000 showed improved survival of 82.2%, attributable to patients being younger and having less organ dysfunction at the time of HSCT.[22] Successful outcome has also been reported after HSCT for CD40 deficiency.

X-linked immunodeficiency with ectodermal dystrophy, caused by mutations of the *IKBKG* (*NEMO*) gene that impair nuclear factor (NF)-κB signaling, may also present with features of combined immunodeficiency. Colitis is frequently seen and may be caused by infection and immune dysregulation as well as abnormalities of NF-κB signaling in the gut epithelial cells. Concern about the failure to correct features of the disease not intrinsic to the hematopoietic system has led clinicians to reserve HSCT for only severely affected cases. Of 29 patients in an international survey, 22 survived, with those who had pre-existing mycobacterial infection having poorer outcomes. Furthermore, colitis has been reported to persist after HSCT.[23,24] Conflicting results have also been reported after HSCT in patients with hypermorphic mutations of the *IKBA* gene, another cause of combined immune deficiency with ectodermal dystrophy.[25]

Dedicator of cytokinesis 8 (*DOCK8*) deficiency is a recently described form of combined immunodeficiency with elevated serum IgE, cutaneous viral infections, and a high risk of malignancy. HSCT is now established as an effective treatment for this disease, with a survival of 84% reported in an international retrospective study of 81 patients.[26] A trend to better survival in patients receiving RIC conditioning and suggestion of selective advantage for donor-derived T cells and switched memory B cells[27] argue in favor of RIC regimens for this disease, in which mixed chimerism may be sufficient for the amelioration of symptoms. HSCT has also been successfully used in patients with combined immune deficiency resulting from DOCK2 deficiency.[28]

Wiskott-Aldrich Syndrome

Bone marrow transplantation for correction of WAS was attempted as early as 1968, with partial success. Full correction following HSCT was first reported in 1978 when a more robust conditioning regimen was used. Since then, results of related HLA-identical HSCT in WAS have been consistently good, and excellent results have been also achieved with HSCT from MUD in children under 5 years of age. While two large multi-institutional surveys have shown survival in the 21st century of ~90% (Fig. 90.6), variability in disease correction due to mixed chimerism remains a significant challenge.[29,30] Persistent thrombocytopenia occurs when myeloid chimerism is below 50% (Fig. 90.7). Furthermore, RIC regimens for patients who cannot tolerate high-dose busulfan do not consistently achieve high-level donor chimerism.[30]

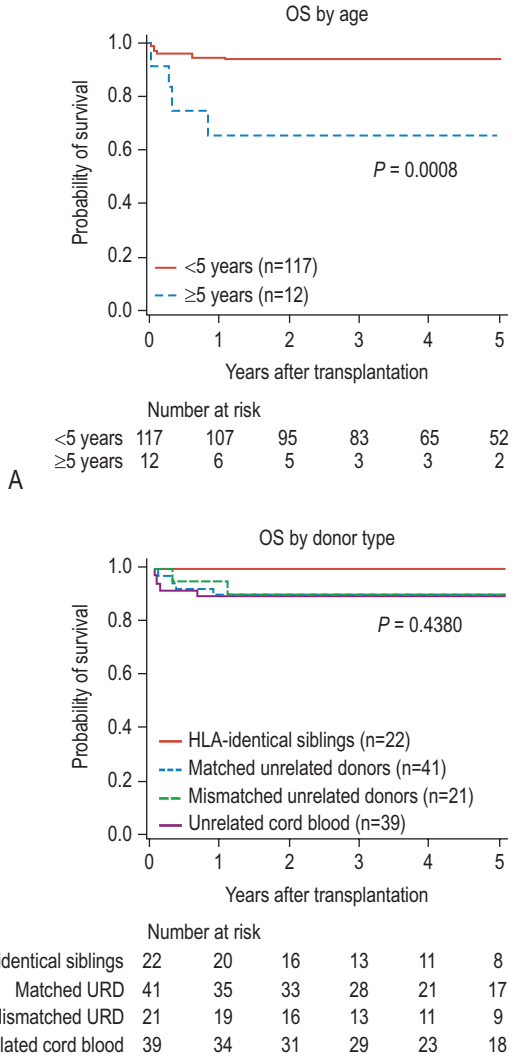


FIG. 90.6 Cumulative probability of overall survival *(OS)* of patients with Wiskott-Aldrich syndrome according to (A) age <5 years versus >5 years, and (B) donor type. *HLA*, Human leukocyte antigens; *UED*, unrelated. (Redrawn after Fig. 3 of Burroughs LM, Petrovic A, Brazauskas R, et al. Excellent outcomes following hematopoietic cell transplantation for Wiskott-Aldrich syndrome: a PIDTC report. *Blood*. 2020;135[23]:2094–2105.)

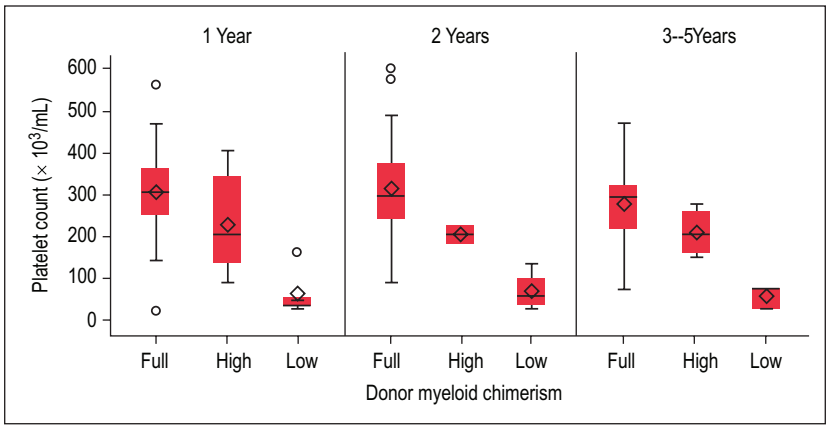


FIG. 90.7 Correlation of platelet count at the indicated times post-transplant of patients with Wiskott-Aldrich syndrome with donor myeloid chimerism. Box plots of platelet counts in patients with full donor (>95%), high donor (50% to 95%), and low donor (5% to 59%) myeloid chimerism are shown. (Redrawn after Fig. 5 of Burroughs LM, Petrovic A, Brazauskas R, et al. Excellent outcomes following hematopoietic cell transplantation for Wiskott-Aldrich syndrome: a PIDTC report. *Blood*. 2020;135[23]:2094–2105.)

Disorders Associated With Hemophagocytic Lymphohistiocytosis

Familial hemophagocytic lymphohistiocytosis (FHL) comprises a genetically heterogeneous group of disorders of T-cell and NK cell-mediated cytotoxicity. Although chemotherapy may induce remission, patients with FHL tend to relapse and ultimately die, mostly as a result of multiorgan failure that occurs in the accelerated phase of the disease. At present, HSCT is the only curative approach to FHL. However, patients with FHL often are critically ill, with extensive organ involvement and/or active infections, and may suffer from refractory disease. For these reasons, patients are unusually prone to developing transplantation-related toxicities and complications. While the use of RIC regimens has improved survival, mixed chimerism and graft failure remain significant problems.[31] Selection of optimal family donors is an important issue for HSCT in FHL. Functional and genetic studies must be used to screen potential familial donors to avoid those who may be genetically affected, but as yet asymptomatic.

Excellent results have been reported, also after HSCT for X-linked lymphoproliferative disease type 1 (XLP1). In an international series of 91 patients with XLP caused by an *SH2D1A* gene defect, survival was 81.4% in 43 patients treated with HSCT compared with 62.5% in 48 untransplanted patients.[32] Furthermore, the majority of untransplanted survivors required IgG replacement therapy, whereas good immune reconstitution was achieved in most transplanted patients. However, survival was poorer (50%) among patients with a previous history of hemophagocytic lymphohistiocytosis (HLH), suggesting that ideally the HSCT should be performed before the onset of EBV infection.[32] More recently, a single-center experience involving 16 patients with XLP1 showed that reduced-intensity conditioning with alemtuzumab, fludarabine, and melphalan is effective, even in patients with XLP1 with a history of HLH.[33] Survival was 80% at 1 year and 71% in the long term; mixed chimerism was common but, in most cases, was sufficient to control the disease. Infectious complications (especially viral infections) were recorded in the majority of the patients.[33]

XLP type 2 (XLP2), caused by mutations in the X-linked inhibitor of apoptosis (*XIAP*) gene, is associated with various phenotypes, including XLP, HLH, and severe colitis. An international survey of transplantation for this condition identified 19 patients who underwent HSCT following fully myeloablative ($n=7$), reduced-intensity ($n=11$), or intermediate-level ($n=1$) conditioning.[34] MAC was associated with poor survival (14%) and a high rate of serious, transplantation-related toxicities. Both this series and a series of 10 patients in Japan who underwent HSCT showed far better outcomes after reduced-intensity conditioning.[34,35]

The hematological and immune complications of Chediak-Higashi syndrome (CHS) can be cured with HSCT. In a series of 35 patients, 5-year survival after HSCT was 62%.[36] Mortality was higher in patients who were in the life-threatening accelerated phase of the disease at the time of transplantation and in those who received HSCT from an alternative related donor.[36] MUD HSCT represents a valid option in patients with CHS. However, the long-term outcome of patients with CHS treated with HSCT remains unclear, especially since neurological deterioration has been consistently observed several years after transplantation.

Griscelli syndrome type 2 (GS2) is a genetic disease characterized by hemophagocytic lymphohistiocytosis and a high risk of neurological complications. A recent series of 35 patients confirmed protection from HLH symptoms after HSCT, even with mixed chimerism, but the substantial risk of mortality, graft failure, and neurologic issues, particularly in those with HLH affecting the central nervous system.[37]

Phagocytic Cell Disorders

Although regular administration of prophylactic antibiotics and antifungal agents (with the possible addition of interferon [IFN]-γ) has improved survival in patients with CGD, this remains a severe disorder with a high risk of complications and death, especially in patients with oxidase-null phenotype (Chapter 39).[38] A real breakthrough in the treatment of CGD was provided by a prospective multicenter study in 56 patients with CGD, 42 of who had high-risk features, including treatment-refractory infections and severe inflammatory complications.[39] In this study, reduced-intensity conditioning (with low-dose or targeted busulfan administration, high-dose fludarabine, and serotherapy) and transplantation from matched related

($n=21$) or unrelated ($n=35$) donors resulted in a 2-year overall survival rate of 96% and an event-free survival rate of 91%. Stable (≥90%) myeloid chimerism was documented in 93% of surviving patients. A low rate of aGvHD of grade III/IV (4%) and cGvHD (7%) was observed. A retrospective review of outcomes of HSCT of 145 CGD patients performed by the PIDTC similarly showed high survival and engraftment, and importantly, showed in contrast to previous literature that patients with CGD-associated inflammatory bowel disease nonetheless had outcomes indistinguishable from those without these symptoms.[40] Studies of female carriers of X-linked CGD suggest that a threshold of 10% to 15% myeloid chimerism is likely sufficient for protection from infection, but that there is no such clear threshold for the avoidance of inflammatory symptoms.[41]

HSCT is a successful and lifesaving procedure in patients with the complete form of leukocyte adhesion defect type 1 (LAD-1). A multicenter study of 36 such patients who underwent HSCT between 1993 and 2007 reported an overall survival of 75%, with similar results in matched related and unrelated donors. Mortality was higher (four of eight cases) after haploidentical HSCT. Stable mixed multilineage chimerism is sufficient to cure the disease.[42] More recently, employing blockade of the P40 subunit of IL-23 and IL-12 with ustekinumab resolved inflammatory symptoms in a patient with LAD-1 after 1 year of therapy.[43]

Administration of recombinant G-CSF is the treatment of choice for patients with severe congenital neutropenia (SCN). However, a subgroup of these patients fail to respond to G-CSF, and some of them are at high risk for the development of myelogenous leukemia. A retrospective multicenter study of 136 patients with SCN who underwent allogeneic HSCT between 1990 and 2012 reported a 3-year overall survival rate of 82%, with 17% transplantation-related mortality. Transplantation at a younger age (<10 years) and the use of matched related or unrelated donors were associated with better outcomes.

Heterozygous mutations in the GATA2 gene result in a wide variety of clinical manifestations, including the Mono-Mac syndrome, pulmonary alveolar proteinosis (PAP), bone marrow failure, susceptibility to infection with mycobacteria and DNA viruses, myelodysplastic syndrome, and progression to acute myelogenous leukemia (AML). These hematopoietic-intrinsic defects are eminently correctable with HSCT using a variety of donors as reported in a 22 patient series.[44]

Regulatory T-Cell Defects

Immunodysregulation/polyendocrinopathy/enteropathy/X-linked (IPEX) syndrome is a severe disorder with immune dysregulation caused by mutations of the *FOXP3* gene, which plays a critical role in the development and function of regulatory T cells (Treg). Patients with IPEX syndrome often die early in infancy and manifest severe multi-system autoimmunity. HSCT is effective and is now established as a standard treatment. A report of 96 patients with IPEX, 58 of who underwent HSCT, showed 73.2% survival in contrast to 65.1% with immunomodulatory therapy alone. RIC and MAC regimens appeared to have an equal survival benefit, which may be related to the strong selective advantage for donor-derived Tregs observed in patients with mixed chimerism after HSCT.[45] Patients with more mild disease activity at the time of HSCT fared best, speaking to the importance of control of autoimmunity prior to the procedure.[46]

Small series have been published on HSCT for other disorders affecting Treg function, including activated PI3 kinase delta syndrome, STAT1 gain of function, and CTLA4 haploinsufficiency.[47–49] Outcomes in these disorders, for which donor-derived Tregs do not have a selective advantage as they do in IPEX, have been variable due to issues of poor engraftment and mixed chimerism.

Interferon-γ Pathway Disorders

IFN-γ receptor 1 deficiency leads to severe mycobacterial infections, with a high mortality rate of mortality early in life. Although HSCT should theoretically correct the disease, results have been disappointing, with few exceptions. An international survey of eight transplanted patients showed that only two were in full remission 5 years after transplantation.[50] The high level of IFN-γ in these patients inhibits engraftment of donor-derived HSCs, accounting for the high rejection rate. Correction of the disease has been reported with HSCT following control of mycobacterial infection, normalization of IFN-γ levels, and use of MAC.

ON THE HORIZON

The following strategies may improve the outcome of HSCT for primary immunodeficiency in the future:

- Antibody-based methods to increase and sustain engraftment of stem cells without systemic organ toxicity.
- Strategies to facilitate engraftment and reduce the occurrence of GvHD in patients with mismatched or matched unrelated donor transplants.
- Approaches to increase thymopoiesis to accelerate T-cell immune reconstitution.
- Improved antimicrobial protocols to reduce infectious complications of HSCT.

FUTURE DIRECTIONS FOR HEMATOPOIETIC STEM CELL TRANSPLANTATION IN THE TREATMENT OF PRIMARY IMMUNE DEFICIENCY

Although allogeneic HSCT has clear efficacy in patients with severe forms of PID, several challenges remain. The development of RIC regimens notwithstanding, efficient myeloablation without the toxicity of alkylating agents for PID patients, many of whom are young children or infants, is desperately needed. Cytolytic mAbs against CD117 (c-kit) are now being tested to remove autologous HSCs, making space for engraftment of donor-derived cells while avoiding the systemic toxicity of chemotherapy and radiotherapy.[51]

The use of cytotoxic drugs and GvHD are significant risk factors for post-transplantation T-cell deficiency as they interfere with normal thymic function.[52] Strategies aimed at improving thymic function after HSCT may include (1) protection of thymic stroma that supports thymopoiesis and (2) direct stimulation or infusion of early T-cell progenitors. It is critical that newly approved agents for the treatment and prevention of GvHD be tested in young children, so that young PID patients may benefit.

Now more than 50 years after the first successful use of HSCT, we are faced with a large cohort of adult survivors. It is important to fully capture and analyze their long-term outcomes. Natural history studies of these patients are critical to understanding the durability of HSCT, assessing the quality of

life and impact on patients' well-being, and elucidating early biomarkers that predict long-term success. By doing so, we may be able to identify patients at risk of poor long-term outcomes, deserving of further early intervention.

REFERENCES

1. Gennery AR, Slatter MA, Grandin L, et al. Transplantation of hematopoietic stem cells and long-term survival for primary immunodeficiencies in Europe: Entering a new century, do we do better? *J Allergy Clin Immunol*. 2010;126(3):602–610.
2. Haddad E, Logan BR, Griffith LM, et al. SCID genotype and 6-month posttransplant CD4 count predict survival and immune recovery. *Blood*. 2018;132(17):1737–1749.
3. Shah RM, Elfeky R, Nademi Z, et al. T-cell receptor alphabeta(+) and CD19(+) cell-depleted haploidentical and mismatched hematopoietic stem cell transplantation in primary immune deficiency. *J Allergy Clin Immunol*. 2018;141(4):1417–1426. e1.
4. Eapen M, Wang T, Veys PA, et al. Allele-level HLA matching for umbilical cord blood transplantation for non-malignant diseases in children: a retrospective analysis. *Lancet Haemat*. 2017;4(7):e325–e333.
5. Ferrara JLM, Levine JE, Reddy P, et al. Graft-versus-host disease. *The Lancet*. 2009;373(9674):1550–1561.
6. Jagasia MH, Greinix HT, Arora M, et al. National Institutes of Health Consensus Development Project on criteria for clinical trials in chronic graft-versus-host disease: I. The 2014 diagnosis and staging working group report. *Biol Blood Marrow Transplant*. 2015;21(3):389–401.e1.
7. Cutler CS, Koreth J, Ritz J. Mechanistic approaches for the prevention and treatment of chronic GVHD. *Blood*. 2017;129(1):22–29.
8. Pai S-Y, Logan BR, Griffith LM, et al. Transplantation outcomes for severe combined immunodeficiency, 2000–2009. *N Engl J Med*. 2014:371.
9. Tomblyn M, Chiller T, Einsele H, et al. Guidelines for preventing infectious complications among hematopoietic cell transplant recipients: a global perspective. Preface. *Bone Marrow Transplant*. 2009;44(8):453–455.
10. Naik S, Nicholas SK, Martinez CA, et al. Adoptive immunotherapy for primary immunodeficiency disorders with virus-specific T lymphocytes. *J Allergy Clin Immunol*. 2016;137(5):1498–1505.e1.
11. Kwan A, Abraham RS, Currier R, et al. Newborn screening for severe combined immunodeficiency in 11 screening programs in the United States. *JAMA*. 2014;312(7):729–738.
12. Schuetz C, Neven B, Dvorak CC, et al. SCID patients with ARTEMIS vs RAG deficiencies following HCT: increased risk of late toxicity in ARTEMIS-deficient SCID. *Blood*. 2014;123(2):281–289.
13. Hassan A, Booth C, Brightwell A, et al. Outcome of hematopoietic stem cell transplantation for adenosine deaminase-deficient severe combined immunodeficiency. *Blood*. 2012;120(17):3615–3624. quiz 26.
14. Heimall J, Logan BR, Cowan MJ, et al. immune reconstitution and survival of 100 SCID patients post hematopoietic cell transplant: A PIDTC natural history study. *Blood*. 2017
15. Miggelbrink AM, Logan BR, Buckley RH, et al. B-cell differentiation and IL-21 response in IL2RG/JAK3 SCID patients after hematopoietic stem cell transplantation. *Blood*. 2018;131(26):2967–2977.
16. Buckley RH. Transplantation of hematopoietic stem cells in human severe combined immunodeficiency: longterm outcomes. *Immunol Res*. 2011;49(1–3):25–43.
17. Neven B, Leroy S, Decaluwe H, et al. Long-term outcome after haematopoietic stem cell transplantation of a single-centre cohort of 90 patients with severe combined immunodeficiency: Long-term outcome of HSCT in SCID. *Blood*. 2009;113(17):4114–4124.
18. Lum SH, Anderson C, McNaughton P, et al. Improved transplant survival and long-term disease outcome in children with MHC class II deficiency. *Blood*. 2020;135(12):954–973.
19. Markert ML, Devlin BH, McCarthy EA. Thymus transplantation. *Clin Immunol*. 2010;135(2):236–246.
20. Davies EG, Cheung M, Gilmour K, et al. Thymus transplantation for complete DiGeorge syndrome: European experience. *J Allergy Clin Immunol*. 2017;140(6):1660–1670.e16.
21. Gennery AR, Khawaja K, Veys P, et al. Treatment of CD40 ligand deficiency by hematopoietic stem cell transplantation: a survey of the European experience, 1993–2002. *Blood*. 2004;103(3):1152–1157.
22. Ferrua F, Galimberti S, Courteille V, et al. Hematopoietic stem cell transplantation for CD40 ligand deficiency: Results from an EBMT/ESID-IEWP-SCETIDE-PIDTC study. *J Allergy Clin Immunol*. 2019;143(6):2238–2253.
23. Pai SY, Levy O, Jabara HH, et al. Allogeneic transplantation successfully corrects immune defects, but not susceptibility to colitis, in a patient with nuclear factor kB essential modulator deficiency. *J Allergy Clin Immunol*. 2008;122(6):1113–1118.e1.
24. Miot C, Imai K, Imai C, et al. Hematopoietic stem cell transplantation in 29 patients hemizygous for hypomorphic IKBKG/NEMO mutations. *Blood*. 2017;130(12):1456–1467.
25. Fish JD, Duerst RE, Gelfand EW, et al. Challenges in the use of allogeneic hematopoietic SCT for ectodermal dysplasia with immune deficiency. *Bone Marrow Transplant*. 2009;43(3):217–221.
26. Aydin SE, Freeman AF, Al-Herz W, et al. Hematopoietic stem cell transplantation as treatment for patients with DOCK8 deficiency. *J Allergy Clin Immunol Pract*. 2019;7(3):848–855.
27. Al-Herz W, Chu JI, van der Spek J, et al. Hematopoietic stem cell transplantation outcomes for 11 patients with dedicator of cytokinesis 8 deficiency. *J Allergy Clin Immunol*. 2016;138(3):852–859.e3.
28. Dobbs K, Dominguez Conde C, Zhang SY, et al. Inherited DOCK2 Deficiency in patients with early-onset invasive infections. *N Engl J Med*. 2015;372(25):2409–2422.
29. Moratto D, Giliani S, Bonfim C, et al. Long-term outcome and lineage-specific chimerism in 194 patients with Wiskott-Aldrich syndrome treated by hematopoietic cell transplantation in the period 1980–2009: an international collaborative study. *Blood*. 2011;118(6):1675–1684.
30. Burroughs LM, Petrovic A, Brazauskas R, et al. Excellent outcomes following hematopoietic cell transplantation for Wiskott-Aldrich syndrome: a PIDTC report. *Blood*. 2020;135(23):2094–3105.
31. Allen CE, Marsh R, Dawson P, et al. Reduced-intensity conditioning for hematopoietic cell transplant for HLH and primary immune deficiencies. *Blood*. 2018;132(13):1438–1451.
32. Booth C, Gilmour KC, Veys P, et al. X-linked lymphoproliferative disease due to SAP/SH2D1A deficiency: a multicenter study on the manifestations, management, and outcome of the disease. *Blood*. 2011;117(1):53–62.
33. Marsh RA, Bleesing JJ, Chandrakasan S, et al. Reduced-intensity conditioning hematopoietic cell transplantation is an effective treatment for patients with SLAM-associated protein deficiency/X-linked lymphoproliferative disease type 1. *Biol Blood Marrow Transplant*. 2014;20(10):1641–1645.
34. Marsh RA, Rao K, Satwani P, et al. Allogeneic hematopoietic cell transplantation for XIAP deficiency: an International survey reveals poor outcomes. *Blood*. 2013;121(6):877–883.
35. Ono S, Okano T, Hoshino A, et al. Hematopoietic stem-cell transplantation for XIAP deficiency in Japan. *J Clin Immunol*. 2017;37(1):85–91.
36. Eapen M, DeLaat CA, Baker KS, et al. Hematopoietic cell transplantation for Chediak-Higashi syndrome. *Bone Marrow Transplant*. 2007;39(7):411–415.
37. Al-Mofareh M, Ayas M, Al-Seraihy A, Siddiqui K, Al-Jefri A, Ghemlas I, et al. Hematopoietic stem-cell transplantation in children with Griscelli syndrome type 2: a single-center report on 35 patients. *Bone Marrow Transplant*. 2020;55(10):2026–2034.
38. Kuhns DB, Alvord WG, Heller T, Feld JJ, Pike KM, Marciano BE, et al. Residual NADPH oxidase and survival in chronic granulomatous disease. *N Engl J Med*. 2010;363(27):2600–2610.
39. Gungor T, Teira P, Slatter M, et al. Reduced-intensity conditioning and HLA-matched hematopoietic stem-cell transplantation in patients with chronic granulomatous disease: a prospective multicentre study. *Lancet*. 2014;383(9915):436–448.
40. Marsh RA, Leiding JW, Logan BR, et al. Chronic granulomatous disease-associated IBD resolves and does not adversely impact survival following allogeneic HCT. *J Clin Immunol*. 2019;39(7):653–667.
41. Marciano BE, Zerbe CS, Falcone EL, et al. X-linked carriers of chronic granulomatous disease: Illness, lyonization, and stability. *J Allergy Clin Immunol*. 2018;141(1):365–671.

42. Qasim W, Cavazzana-Calvo M, Davies EG, et al. Allogeneic hematopoietic stem-cell transplantation for leukocyte adhesion deficiency. *Pediatrics*. 2009;123(3):836–840.
43. Moutsopoulos NM, Zerbe CS, Wild T, et al. Interleukin-12 and Interleukin-23 blockade in leukocyte adhesion deficiency type 1. *N Engl J Med*. 2017;376(12):1141–1146.
44. Parta M, Shah NN, Baird K, et al. Allogeneic hematopoietic stem cell transplantation for GATA2 deficiency using a Busulfan-based regimen. *Biol Blood Marrow Transplant*. 2018;24(6):1250–1259.
45. Horino S, Sasahara Y, Sato M, et al. Selective expansion of donor-derived regulatory T cells after allogeneic bone marrow transplantation in a patient with IPEX syndrome. *Pediatr Transplant*. 2014;18(1):E25–30.
46. Barzaghi F, Amaya Hernandez LC, et al. Long-term follow-up of IPEX syndrome patients after different therapeutic strategies: An international multicenter retrospective study. *J Allergy Clin Immunol*. 2018;141(3):1036–1049.e5.
47. Nademi Z, Slatter MA, Dvorak CC, et al. Hematopoietic stem cell transplant in patients with activated PI3K delta syndrome. *J Allergy Clin Immunol*. 2017;139(3):1046–1049.
48. Leiding JW, Okada S, Hagin D, et al. Hematopoietic stem cell transplantation in patients with gain-of-function signal transducer and activator of transcription 1 mutations. *J Allergy Clin Immunol*. 2018;141(2):704–717.e5.
49. Schwab C, Gabrysch A, Olbrich P, et al. Phenotype, penetrance, and treatment of 133 cytotoxic T-lymphocyte antigen 4-insufficient subjects. *J Allergy Clin Immunol*. 2018;142(6):1932–1946.
50. Roesler J, Horwitz ME, Picard C, et al. Hematopoietic stem cell transplantation for complete IFN-gamma receptor 1 deficiency: a multi-institutional survey. *J Pediatr*. 2004;145(6):806–812.
51. Czechowicz A, Kraft D, Weissman IL, Bhattacharya D. Efficient transplantation via antibody-based clearance of hematopoietic stem cell niches. *Science*. 2007;318(5854):1296–1299.
52. Krenger W, Blazar BR, Holländer GA. Thymic T-cell development in allogeneic stem cell transplantation. *Blood*. 2011;117(25):6768–6776.

Gene Therapy for Primary Immune Deficiency Diseases

Caroline Y. Kuo and Donald B. Kohn

Gene therapy, as it is being applied for primary immune deficiency (PID) diseases, represents an autologous hematopoietic stem cell transplant (HSCT), in which a patient's own stem cells are genetically corrected ex vivo and then transplanted back (Fig. 91.1). Thus, gene therapy for PID builds upon decades of experience using allogeneic HSCT from a healthy donor, in which replacement of some or all of a PID patient's bone marrow hematopoietic stem cells (HSCs) with HSCs from a healthy donor can cure the disease. Although initial efforts in gene therapy did not provide clinical benefit to the patients involved, there has been steady progress with the methods to the point that there is now clear-cut therapeutic efficacy using gene therapy/autologous HSCT for the three most commonly transplanted PIDs: severe combined immune deficiency (SCID), Wiskott-Aldrich syndrome (WAS), and chronic granulomatous disease (CGD) (see Chapters 34 and 35).

Most gene therapy efforts to date have used **gene addition** methods in which a normal copy of the relevant gene is added to the patient's cells, usually using a viral vector. More recently, methods are under development to perform **gene editing**, most commonly using site-specific endonucleases (such as CRISPR/Cas9) to disrupt a target gene or modify or add new sequences by homologous recombination (HR).

The key potential advantage of autologous transplant with gene therapy, compared to allogeneic HSCT, is that it may have reduced risks and a better safety profile, because it eliminates the need for pre-transplant immune-suppressive conditioning and post-transplant immune suppression; essentially, there should be no risk of graft-versus host disease (GvHD) from the autologous graft (Table 91.1). Additionally, gene therapy could have increased efficacy compared to allogeneic HSCT in some conditions, due to the potential to overexpress the relevant gene product (e.g. adenosine deaminase [ADA] enzyme) and lead to a supraphysiological effect from the engineered graft.

However, gene therapy may have unique risks: the potential for causing malignant transformation and leukoproliferative complications from the manipulation of the genome of the stem cells, by either gene addition or gene editing. Moreover, gene therapy could have decreased efficacy if the percentage of transplanted stem cells that are successfully gene-corrected is low, if the level of expression of the inserted transgene is suboptimal, or if the process of ex vivo gene manipulation impairs the stem cells' capacity for long-term hematopoiesis. Additionally, developing gene therapies for PID requires a separate translational research process for each genetic etiology (e.g., ADA SCID, XSCID, Artemis SCID, Rag1 SCID, Rag2 SCID, etc.), whereas allogeneic HSCT is a more "one size fits all" approach, requiring less individualization for similar classes of disorders.

KEY CONCEPTS

- Some severe primary immune deficiencies (PIDs) can be treated by transplantation of hematopoietic stem cells (HSCs) from a healthy donor, but this is optimal with a well-matched immune-compatible donor and may yet entail immune complications.
- Gene therapy using autologous HSCs that have been gene-corrected (by gene addition or endogenous correction) may avoid the immune complications of allogeneic transplant and confer similar benefits.
- Stable gene addition to HSC can be done using integrating viral vectors, derived from retroviruses or lentiviruses.
- Gene therapy using γ-retroviral vectors led to immune reconstitution for several forms of severe combined immune deficiency, Wiskott-Aldrich syndrome, and chronic granulomatous disease, but some patients developed leukoproliferative complications.
- More recent trials using safer vectors are continuing to yield clinical efficacy without vector-related adverse effects and with good safety profiles.
- New techniques are being developed for precise gene editing in HSCs, which may allow application to a wider spectrum of PIDs.

GENE TRANSFER TO HEMATOPOIETIC STEM CELLS

For gene therapy to have an enduring effect in PID, the gene addition or editing must occur in the long-term pluripotent HSCs by some method that will lead to the corrective gene being passed on to the billions of progeny blood cells arising from each HSC. Gene modification of the far more numerous, but short-lived, lineage-committed progenitor cells would lead only to transient presence of gene-corrected mature blood cells. The technical challenge is for the gene delivery to HSCs to be efficient (high percentage of the cells modified), to yield persistent expression, and to have minimal immediate cytotoxicity or long-term genotoxicity.

Many of the gene transfer methods used for research purposes, such as transfection, electroporation, and nano-particle delivery, are themselves transient, and the inserted gene would be lost by dilution as the stem cells proliferate. Thus, most of the studies to date have used viral vectors derived from the Retroviridae family that integrate their genomes into the chromosomes of the target cell to achieve persistence of the normal gene (Fig. 91.2). Genes delivered by murine γ-retroviral, human lentiviral (HIV-1-based), spumaviral (foamy), or alpha-retroviral vectors remain permanently covalently linked to the cellular chromosomal DNA for stable transmission to progeny cells. The steps involved in using these types of vectors to produce a gene-modified HSC graft are described in the Therapeutic Principles box.

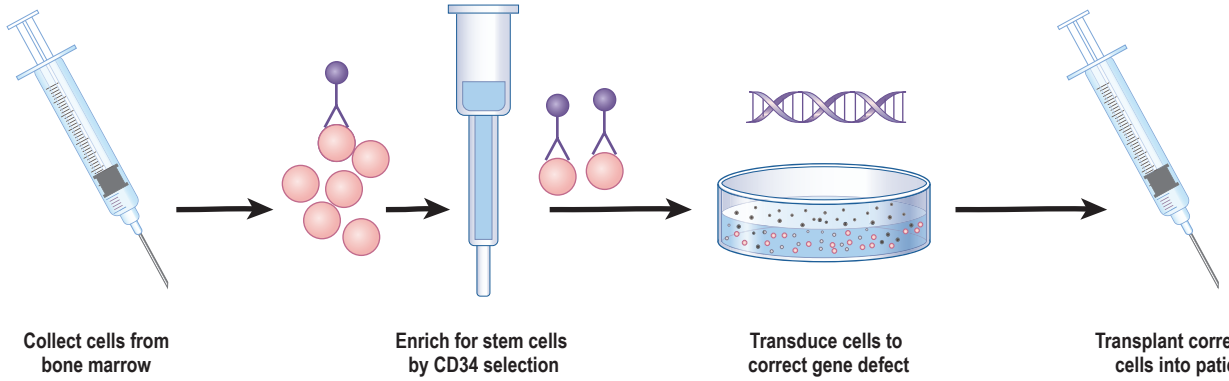


FIG. 91.1 Gene Therapy to Treat Primary Immune Deficiencies (PIDs) by Autologous Transplantation of Gene-Corrected Hematopoietic Stem Cells (HSCs). HSCs may be obtained from the bone marrow, by mobilization with granulocyte–colony-stimulating factor (G-SCF) and leukapheresis (peripheral blood stem cells [PBSCs]), or from umbilical cord blood of a patient with PID. The HSCs may be enriched by isolating the $CD34^+$ fraction of cells using immunomagnetic separation methods. The $CD34^+$ cell population is then cultured for gene manipulation (gene addition or gene editing). The gene-corrected autologous HSCs are then transplanted back to the patient.

TABLE 91.1 Advantages and Disadvantages of Allogeneic Versus Autologous (Gene Therapy) Hematopoietic Stem Cell Transplantation for Primary Immune Deficiencies

Transplant Type	Advantage	Disadvantage
Allogeneic HSCT	Normal function of relevant gene assured; benefit expected if sufficient donor chimerism	Need suitable matching donor
	Well-established and long-term experience with benefits/risks	Immune: GvHD and rejection risks
	Excellent outcomes with matched sibling donor (although most patients lack a matched sibling)	Requires immune modulation: immune ablative conditioning; graft manipulation; and GvHD prophylaxis and treatment. These may contribute to morbidity
Autologous gene therapy HSCT	Patient is donor	Potential genotoxicity from gene addition or editing causing cell loss, dysfunction, or transformation
	No risks of GvHD	Need to gene-modify high percentage of primary HSC with minimal cytotoxicity
	May not require immune suppression before (e.g., flu/ATG) and after (e.g., corticosteroids, calcineurin inhibitor) HSCT	Low fractional correction may blunt efficacy
	Risks of rejection may be less than with allogeneic cells	Prior chemotherapy or marrow dysfunction may preclude use of autologous stem cells
	Transgene overexpression may have benefits (e.g., adenosine deaminase)	Immunogenicity of transgene products is not well defined
Gene addition (e.g., lentiviral vector)	Currently showing efficacy for SCID, WAS, and CGD (and for non-PIDs; e.g., X-ALD, MLD, β-thalassemia, sickle cell disease, MPS-1)	Risks for insertional mutagenesis (gene disruption, gene activation) from semi-random insertions into target cell genomes
	Current generation of self-inactivating (SIN) lentiviral vectors has reduced genotoxicity compared to earlier-used γ-retroviral vectors	Transgene not under normal transcriptional control. Function of transgene (level, lineage, longevity) may vary and be variegated
Gene editing (e.g., nuclease/HDR)	Corrected endogenous gene should have normal function.	Risks of local or off-target gene disruption (insertion/deletion)
		Risk of translocations

ATG, Anti-thymocyte globulin; *CGD*, chronic granulomatous disease; flu, fludarabine; *GvHD*, graft-versus-host disease; *HDR*, homology directed repair; *HSC*, hematopoietic stem cells; *HSCT*, hematopoietic stem cell transplantation; *MLD*, metachromatic leukodystrophy; *MPS-1*, mucopolysaccharidosis type I {Hurler syndrome}; *PID*, primary immune deficiencies; *SCID*, severe combined immune deficiency; *WAS*, Wiskott-Aldrich syndrome; *X-ALD*, X-linked adrenoleukodystrophy.

CLINICAL TRIALS OF GENE THERAPY FOR PRIMARY IMMUNE DEFICIENCIES

To date, clinical trials of autologous transplant/gene therapy have been performed for several PID disorders, including multiple trials for ADA-deficient SCID, XSCID, CGD, and WAS, and more recent studies for Artemis-deficient SCID and leukocyte adhesion deficiency (LAD-1) (Table 91.2). These have occurred over roughly demarcated eras as the technology and clinical approaches advanced in the past three decades (Table 91.3). Because each genetic form of PID requires a separate development research path from preclinical studies to define efficacy and safety through clinical trials, gene therapy for each individual PID will be discussed separately. However, they share a common progression through the different eras.

Over these eras, the preferred vector for gene delivery to HSCs has progressed from murine γ-retroviral vectors (γ-RV) to HIV-1-based lentiviral vectors in which the enhancers in the

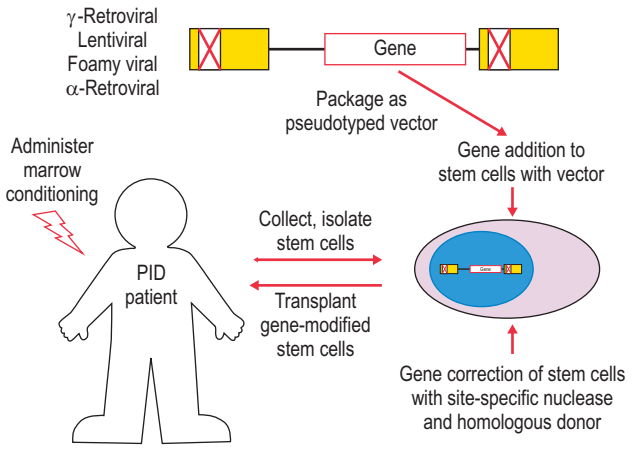


FIG. 91.2 Autologous Transplantation of Gene-Corrected Hematopoietic Stem Cells (HSCs). Gene addition may be performed using Retroviridae-derived vectors (from γ-retroviruses, lentiviruses, foamy viruses, or α-retroviruses) to transfer a normal copy of a relevant gene into HSCs collected and isolated from a patient with primary immune deficiency *(PID)*. The gene-containing vector is packaged in a suitable envelope (pseudotyped) for gene addition to human HSC. Alternatively, the HSC may be gene-edited using site-specific endonucleases to augment homologous recombination–directed gene editing. The patient may receive marrow cytoreductive conditioning prior to transplant to enhance engraftment of the gene-modified stem cells.

THERAPEUTIC PRINCIPLES

Steps for Gene Transfer to Hematopoietic Stem Cells for Clinical Transplantation

Package Retro-Lentiviral Vector

Transfect packaging cell with vector-encoding plasmid and virion protein and envelope-expressing plasmids.

Collect released vector from cell culture medium (digest residual plasmid with DNAse).

Purify and concentrate vector (e.g., ion-exchange chromatography, tangential flow filtration).

Aliquot. Certify (including absence of replication-competent virus). Release for use.

Transduce Stem Cells with Vector

Isolate $CD34^+$ stem/progenitor cells from clinical source (bone marrow, mobilized peripheral blood, cord blood).

Grow $CD34^+$ cells in serum-free medium plus recombinant cytokines (e.g., ckit ligand, flt-3 ligand, thrombopoietin) for 16–24 h.

Add vector to cells and culture for transduction for 16–24 h, during which the vector inserts gene sequences covalently into chromosomal DNA of $CD34^+$ HSC.

Formulate and characterize cell product for release.

Administer Marrow Cytoreductive Conditioning

Deliver single or combination chemotherapeutic agents or monoclonal antibodies to "make space" in HSC niches in the bone marrow.

Transplant Gene-Modified Stem Cells

Infuse gene-modified cell product intravenously, after which stem cells engraft and transmit the transgene to all progeny blood cells.

The transgene produces the necessary gene product to correct the genetic deficiency.

viral long-terminal repeats are self-inactivated during the reverse transcription steps (SIN vectors). Lentiviral vectors have several potential advantages over γ-RV, including the ability to more effectively transduce human HSCs in a shorter period of ex vivo culture; the capacity to carry longer and more complex genetic sequences (such as cellular gene enhancers and promoters); and less tendency to be inserted near the 5′ ends of genes, which may decrease risks for *trans*-activating expression of the adjacent cellular genes. The combination of recombinant cytokines used to activate the HSCs to facilitate transduction was advanced as factors were identified that acted on the earliest HSCs, including the flt-3 ligand and thrombopoietin, combined with the c-*kit* ligand.

The Role of Cytoreductive Conditioning to Facilitate Engraftment

In the initial trials of gene therapy for PID, there was no administration of pre-transplant cytoreductive conditioning, due to potential risks of chemotherapy or radiation with unknown prospects of benefit from the gene transfer procedure. It is well-known from multiple transplant studies in mice and large animal models that there is minimal, if any, engraftment of autologous HSCs when given without prior conditioning, unless extraordinarily high numbers of cells (e.g., 30 to 50× higher than standard cell dose/kg) are given; even mega-dose transplants without conditioning lead to only low levels of engraftment (e.g., 1%), although this may be persistent.[1] While there was initial reluctance to use conditioning in gene therapy trials when there had not previously been efficacy, now with the clear-cut benefits that may be obtained from gene therapy, the necessity to use conditioning for autologous gene therapy HSCT (with a dose generally lower than that needed for allogeneic HSCT) is now recognized as the standard.

The amounts and types of chemotherapy drugs used for conditioning have varied, depending on the disease setting. For SCID, where a relatively low number of gene-corrected HSCs may support immune reconstitution, reduced-intensity conditioning (RIC), such as low-dose busulfan alone (e.g., busulfan at 4 to 6 mg/kg), may be sufficient. For other disorders where less of a selective advantage may exist for the gene-corrected cells, a higher level of engraftment of modified HSC may be needed; and thus stronger conditioning regimens have been used, reaching levels for myeloablative conditioning (MAC) (e.g., busulfan, 12 to 16 mg/kg). For WAS, where it may be necessary to ablate the pre-transplant immunity to reduce risks for post-transplant autoimmunity, immune-suppressive drugs (e.g., fludarabine, rituximab) have been added to conditioning regimens. These combined iterative approaches to improving gene therapy have led to the current state, where clinical benefits are being routinely achieved, as detailed below for each disorder.

New methods are being explored to "make space" in the marrow using alternatives to chemotherapy and radiation, such as monoclonal antibodies to markers present on HSCs. Antibodies to a cytokine receptor on HSC (c-kit), a common leukocyte antigen (CD45), and a macrophage inhibitory receptor (CD47) have all been shown in animal models to improve engraftment of HSC with minimal clinical toxicity. An anti-cKit antibody is being assessed in a clinical trial to improve engraftment in patients with SCID (NCT02963064). These novel approaches to cytoreduction to facilitate engraftment with lower short-term and long-term risks may replace the current approaches using chemotherapeutic agents.

TABLE 91.2 Clinical Trials of Gene Therapy for Primary Immune Deficiencies

PID	Investigator(s)	Year	Vector/Target	Conditioning	NCT #
ADA SCID	Blaese, Anderson, Culver	1990	LASN (MLV LTR) PBMC	None	—
	Bordignon, Mavilio	1992	G1ADA (MLV LTR) PBMC and BM	None	—
	Hoogerbrugge, Valerio	1993	MLV-ADA BM CD34+	None	—
	Kohn, Parkman	1993	G1NA-ADA (MLV LTR) UCB CD34+	None	—
	Onodera/Sakiyama	1995	LASN (MLV LTR) PBMC	None	—
	Aiuti/Roncarolo	1998	G1ADA (MLV-LTR) BM CD34+	Busulfan 4 mg/kg	00599781
	Gaspar/Thrasher	1999	SFFV-ADA-wpre BM CD34+	Melphalan; busulfan	NCT01279720
	Otsu/Ariga	2003	GCsap-M-ADA BM CD34+	None	—
	Kohn/Candotti	2001	MND-ADA and GCsap-M-ADA BM CD34+	None (4). Busulfan (6)	—
	GlaxoSmithKlein	2008	G1ADA BM CD34+	Busulfan	00598481
	Kohn/Candotti	2009	MND-ADA BM CD34+	Busulfan 90 mg/m^2	00794508
	Gaspar	2012	EFS-ADA CD34+ BM/PBSC	Busulfan	01380990
	Kohn	2013	EFS-ADA BM CD34+	Busulfan	01852071
XSCID	Cavazzana-Calvo/ Fischer	1998	MFG-IL2RG BM CD34+	None	—
	Thrasher/Gaspar	1999	MLV-IL2RG BM CD34+	None	—
	Thrasher/Fischer/ Cavazzana/Pai/ Williams	2011	SIN γRV EFS-IL2RG BM CD34+	None	Fr: 01410019 UK: 01175239 US: 01129544
	Sorrentino	2012	EFS-IL2RG-Ins BM CD34+	None/busulfan	01512888
	Malech/DeRavin	2011	EFS-IL2RG-Ins CD34+ PBSC	Busulfan	01306019
Artemis SCID	Cowan/Puck	2018	ART-ART CD34+ BM	Busulfan	03538899
LAD	Hickstein, Bauer	1999	MLV-CD18 BM	None	—
	Kohn, Bueren, Thrasher (Rocket Pharma)	2019	pChim-ITGB2 CD34+ PBSC	Busulfan	03812263
CGD	Malech	1995	MLV-p47 CD34+ PBSC	None	—
	Malech	1998	MLV-gp91phox	None	—
	Ott/Grez	2009	SFFV-gp91phox CD34+ PBSC	8 mg/kg	00927134
	Kang/Malech	2010	MT-gp91phox CD34+ PBSC	10 mg/kg	—
	Thrasher	2013	pChim-gp91phox (G1XCGD) CD34+ PBSC	pK adjusted, ~12 mg/kg	01855685
	Kohn/Williams/Kang	2015	pChim-gp91phox (G1XCGD) CD34+ PBSC	pK adjusted, ~12 mg/kg	02234934
WAS	Klein	2008	SFFV-WAS γRV	Busulfan	—
	Aiuti	2011	1.6hWASP-WPRE LV	Bu/flu/rituximab	01515462
	Hacein-Bey Abina…Cavazzana	2011		Bu/flu (rituximab or alemtuzimab for autoimmunify)	02333760
	Thrasher	2011		Bu/flu	01347242
	Pai	2011		Bu/flu	01410825

NCT#, Clinical Trial registration number (https://clinicaltrials.gov/).
ART, Artemis; *Bu/flu*, busulfan/fludarabine; *CGD*, chronic granulomatous disease; *LAD*, leukocyte adhesion defect; *PID*, primary immune deficiencies; *SCID*, severe combined immune deficiency; *WAS*, Wiskott-Aldrich syndrome; *XSCID*, X-linked SCID.

TABLE 91.3 Eras in the Advancement of Clinical Gene Therapy for Primary Immune Deficiencies

Era	Predominant Vector	Growth Factors	Conditioning Regimen	PIDs Treated	Typical Outcomes
Early (~1990–1999)	γ-RV	IL3/IL6/SCF	None	ADA XSCID CGD LAD	Insufficient engraftment No efficacy
Middle (~1998–2006)	γ-RV	SCF/FLT3L/TPO	RIC, MAC	ADA XSCID CGD WAS	Efficacy Genotoxicity in some
Current (~2007–present)	SIN LV	SCF/FLT3L/TPO	RIC, MAC	ADA XSCID ART SCID XCGD WAS LAD	Efficacy Good safety profile

γ-RV, γ-retroviral vectors; *ADA*, adenosine deaminase; *ART*, artemis; *FLT3L*, FMS-tyrosine kinase 3 ligand; *CGD*, chronic granulomatous disease; *LV*, lentiviral vector; *MAC*, myeloablative conditioning; *PIDs*, primary immune deficiencies; *RIC*, reduced-intensity conditioning; *SCF*, stem cell factor; *SCID*, severe combined immune deficiency; *SIN*, self-inactivating; *TPO*, thrombopoietin; *WAS*, Wiskott-Aldrich syndrome; *XCGD*, X-linked Chronic Granulomatous Disease; *XSCID*, X-linked SCID.

Adenosine Deaminase-Deficient Severe Combined Immune Deficiency (Chapter 34)

The first clinical trial of gene therapy for an inherited disorder (other than a premature attempt at gene therapy for beta-thalassemia in the 1970s) was directed against ADA-SCID. ADA-SCID was the first form of human SCID for which the responsible gene was identified and cloned, allowing gene therapy to be approached.[2] As a ubiquitously expressed housekeeping enzyme, the non-regulated expression of an introduced ADA gene was expected to be tolerated and potentially beneficial. Moreover, the ability to treat ADA-SCID with a bone marrow transplant from a matched sibling donor without the use of cytoreductive or immune-suppressive conditioning was taken to imply that there is a selective advantage for ADA-replete T lymphocytes over ADA-deficient cells and that only a modest number of engrafted gene-corrected HSC might provide clinical benefit.

A series of clinical trials were performed using γ-retroviral vectors targeting either peripheral blood T cells or bone marrow HSC. While initial studies in the early 1990s did not achieve clinical efficacy, subsequent trials that applied RIC prior to transplant of gene-corrected HSCs to increase engraftment have led to immune restoration in the majority of more than 80 treated patients, without vector-related complications.[3–7] The relatively low-dose chemotherapy and absence of GvHD makes these autologous transplants well-tolerated with essentially none of the clinical side effects seen in allogeneic HSCT with conditioning. One of the retroviral vectors used in studies at Telethon Gene Therapy Program for Genetic Diseases, of the Hospital San Raffaele, in Milan, Italy was licensed by GlaxoSmithKline (GSK) and has received licensure approval in the European Union, a major advance for gene therapeutics that is available for patients. Other ongoing clinical trials in the United States and the UK are using a lentiviral vector for ADA-SCID with excellent initial clinical results in terms of efficacy and safety.[8] Thus, gene therapy for ADA-SCID has become a major treatment option for patients, as the efficacy and safety profile have been favorable and may exceed those of matched unrelated or haplo-identical transplants, although there have been no randomized controlled trials.

X-linked Severe Combined Immune Deficiency (Chapter 34)

The second genetic form of human SCID for which the relevant gene was identified and cloned is the X-linked form (XSCID). The responsible gene, *IL2RG*, encodes the common cytokine receptor γ chain (or γ_c), a component of several multimeric receptors for a family of lymphopoietic cytokines: IL-2, IL-4, IL-7, IL-9, IL-15, and IL-21. Defects of γ_c severely impair the development of multiple components of the immune system, and affected patients typically have severely reduced numbers of T and natural killer (NK) cells, and may have immature but non-functional B cells. Retroviral-mediated gene transfer of a normal human *IL2RG* cDNA was shown to restore cytokine-induced signaling activity and lymphocyte function in patients' cells and in murine models.[9,10]

XSCID patients were first treated by gene therapy using γ-retroviral vectors.[11,12] No conditioning was given, relying on the potent selective advantage for survival and proliferation that gene-corrected lymphoid cells were expected to have in patients with lymphopenic SCID. Indeed, robust reconstitution of T-cell immunity was achieved, although there were variable responses of B cells. However, after 2 to 3 years from treatment, several of the patients developed a severe complication from the retroviral vector of leukemia-like clonal leukoproliferation.

New vectors (self-inactivating or "SIN" vectors) that lacked the strong enhancer elements of the retroviral vectors that caused the insertional oncogenesis were developed. One published study using a SIN γ-retroviral vector demonstrated immune restoration without any evidence of leukoproliferation. New clinical trials are in progress using lentiviral vectors.[13] The first of these studies treated a group of patients XSCID who had undergone allogeneic transplant some years earlier, but who had not achieved complete immune reconstitution and had significant morbidity from poor growth and development.[14] Following gene therapy using a lentiviral vector and non-myeloablative conditioning, all have improved general well-being, with development of sufficient antibody production to be able to stop immunoglobulin replacement therapy. Studies treating more typical newly diagnosed infants with XSCID are now underway at several centers in the United States and Europe, with results from one published showing good immune reconstitution in the treated infants.[15]

Chronic Granulomatous Disease (Chapter 39)

CGD is a rare primary immunodeficiency (1 in 200,000 live births in the United States), first termed fatal granulomatous disease of childhood in 1959 to describe individuals affected by recurrent, invasive bacterial and fungal infections complicated by granuloma formation. Its clinical presentation is explained by defects in any of the components that comprise the nicotinamide adenine dinucleotide phosphate (NADPH) oxidase complex leading to phagocyte dysfunction. In its fully assembled form, the NADPH oxidase complex is composed of five proteins, two of which are membrane-bound (gp91phox and p22phox) and three of which are cytosolic (p47phox,

p67phox, and p40phox). All five components are necessary for proper NADPH oxidase function without which affected individuals are particularly predisposed to infections with *Aspergillus* spp., *Staphylococcus aureus*, *Burkholderia cepacia*, *Serratia marcescens*, *Nocardia* spp., and less commonly, *Salmonella* and BCG infections outside of North America. Defects in gp91phox encoded by the *CYBB* gene account for the most common X-linked form of CGD while the remaining four causal genetic defects are inherited in an autosomal recessive (AR) mode. Since its initial description almost six decades ago, CGD has evolved from a disease with early mortality to one with relatively good outcomes and multiple treatment options.

While prophylactic antimicrobial and immunomodulatory agents have dramatically decreased infection rates and inflammatory bowel disease in patients with CGD, curative therapy can only be achieved with HSCT. Transplant outcomes have continued to improve over the years, with good results reported even in patients with high-risk features such as intractable infections and auto-inflammation using RIC protocols.[16] Nevertheless, allogeneic HSCT can still be complicated by GvHD and pre-existing infections and is not preferred for those without an HLA-matched stem cell donor. For these patients, autologous HSCT with gene-modified cells is becoming a more viable and realistic option.

The first gene therapy trials for CGD began in the mid-1990s using γ-retroviral vectors. As with ADA-SCID, initial trials in which pre-transplant conditioning was not used did not lead to efficacy as there was minimal engraftment of gene-corrected stem cells. In subsequent studies where non-myeloablative conditioning with busulfan was used, increasingly higher levels of engraftment of corrected stem cells were achieved.[17] The best and worst of these studies was one using a γ-retroviral vector with a potent transcriptional control element to drive high-level expressions of the *CYBB* gene.[18] The three treated subjects had initial development of neutrophils with restored oxidase function and their severe resistant infections cleared. However, insertional oncogenesis occurred in this study also, and myelodysplasia or frank myeloid leukemia developed.

More recently, trials are using a lentiviral vector with a myeloid-specific transcriptional control element intended to drive *gp91phox* expression in mature myeloid cells, where it is needed, but to not have activity in HSCs that are the likely targets for transformation.[19] Initial results are showing safety and evidence of efficacy. A report of results from nine patients with XCGD treated in the United States and United Kingdom showed sustained restoration of gp91phox expression and functional oxidase activity using the dihydrorhodamine assay in six of seven patients who survived beyond the immediate post-transplant period.[20]

WISKOTT-ALDRICH SYNDROME (CHAPTER 34)

WAS, initially described as an X-linked syndrome in kindreds in Germany and the United States, presents with multiple clinical manifestations, including the classical triad of immune deficiency, eczema, and thrombocytopenia. The complex immune deficiency involves defects of T, B, NK, and antigen-presenting cells. The identified *WAS* gene encodes a 501 amino acid proline-rich protein that has multiple identified functional domains, placing it at the nexus of intracellular signaling and cytoskeleton control. Once the WAS cDNA was obtained, γ-retroviral vectors were constructed and shown to correct several manifestations of the disorder in patient-derived cells and in murine models. However, it is not fully known what levels of *WAS* transgene expression are needed to safely and effectively correct the major manifestations of lymphocyte and platelet dysfunction. A concern has been raised that suboptimal levels or low frequency of expression of WAS protein (WASP) could allow auto-immunity to develop. If correction is only partial (e.g., some proportion of B cells are not gene corrected and have defective auto-regulatory function), autoimmunity may occur.[21]

A first trial of gene therapy for WAS used a γ-retroviral vector with a very potent retroviral long terminal repeat promoter enhancer. Granulocyte colony-stimulating factor (G-CSF)-mobilized peripheral blood stem cells were used as the HSC source and conditioning with myeloablative dosages of busulfan was given prior to transplant. There were excellent initial results in terms of immune reconstitution and platelet counts, demonstrating that gene therapy can be therapeutic for this disorder. However, there was subsequently a very high incidence of acute leukemia among these patients, developing within a few years from treatment, with acute lymphocytic leukemia (ALL), acute myeloid leukemia (AML), or both occurring in a shocking seven of nine subjects.[22] The mechanism of insertional oncogenesis from the γ-retroviral vector was clearly the cause, with the leukemias showing clonal vector integrations adjacent to known proto-oncogenes, such as *LMO2* and *MDS1*.

While that trial was being initiated, work was ongoing through a multicenter collaboration in the European Union (EU) to develop a SIN lentiviral vector for WAS that uses the promoter from the *WAS* gene per se to drive expression of the WAS cDNA. Extensive preclinical studies showed the efficacy of this vector to improve immunological and hematological parameters in murine and human disease models.[23,24] Moreover, this lentiviral vector displayed significantly lower risks for genotoxicity than the type of γ-retroviral vector used in the first trial.

The lentiviral vector with the WAS promoter has now been used in parallel but independent phase I/II clinical trials performed in several countries.[25,26] Relatively high-intensity conditioning was administered in an attempt to obtain high-level engraftment of gene-corrected HSCs, with some differences among the centers in the precise conditioning regimen used. All used moderate doses of busulfan (pharmacokinetically [pK]-adjusted) and fludarabine, and variably, rituximab or other serotherapy agents. These trials have uniformly demonstrated efficacy and safety. Findings have included clinical improvement in general health, bleeding incidents, and eczema, and as expected for an autologous transplant, no problems from GvHD. Most have had good recovery of T, B, and NK cell numbers and functions. However, there have been only modest and variable improvements for platelet levels (e.g., 20 to 60,000/μL), with evidence that a higher transplanted dose of transduced cells led to higher platelet counts.[25] No new onset of autoimmunity has been reported, although in some cases, pre-existing autoimmune problems have persisted. In fact, decreases of indices of autoimmunity and improved B-cell tolerance have been documented after gene therapy for WAS.[27] Polyclonal vector integration patterns were seen, with no reported development of clonal expansions or frank leukoproliferative complications. Integration site patterns resemble those seen in other trials using lentiviral vectors into human HSCs, with highly diverse vector integration sites and no predilection for cancer-related genes, significantly different from the pattern seen in the γ-retroviral vector trials.

These results are promising and of significant clinical benefit and safety but would be better if more normal levels of platelets

could be achieved. It is possible that the level of expression of WASP from the *WAS* gene promoter in the lentiviral vector is inadequate to support normal platelet production or survival. If so, introducing more than one copy of the vector per cell could lead to higher levels of WASP and higher platelet counts. Alternatively, the absolute number of gene-corrected HSCs engrafted may mediate the platelet levels and a higher percentage of transduced cells or higher numbers of gene-corrected cells may support higher platelet levels. Relative risks of multiple integrants versus a stronger promoter are not known. As for the other disorders discussed here, direct gene correction should yield normal, physiological expression of the *WAS* gene and may lead to improvements in all relevant lineages.

Leukocyte Adhesion Defect-1 (Chapter 39)

LAD-1 is another severe PID that may be a good candidate for HSC gene therapy, as it is caused by a single gene defect (*ITGB2* encoding CD18). CD18 functions as a heterodimer with CD11a, -b or -c, which may limit activity only to the appropriate cell types where the CD11 co-chains are expressed. Bauer and Hickstein performed gene therapy for two patients with LAD-1 using a γ-retroviral vector expressing human CD18 to transduce bone marrow cells without conditioning.[28] There were no beneficial effects of gene therapy and both later succumbed to complications of their immune deficiency. More recently a clinical trial for LAD-1 using a lentiviral vector has begun enrollment (NCT03812263).

GENE THERAPY CONSIDERATIONS FOR OTHER PRIMARY IMMUNE DEFICIENCIES

Currently, a new development project is needed to bring gene therapy to clinical application for each individual PID-causing locus, such as the more than 20 human genes that may cause SCID, 5 or more CGD loci, several for hemophagocytic lymphohistiocytosis (HLP), X-linked lymphoproliferative disease (XLP), etc. Each gene and disease setting pose different challenges in terms of necessary gene transfer efficiency, level of expression, need for regulation of the transferred gene, safety considerations, and measurable endpoints.

X-linked agammaglobulinemia (XLA; Chapter 33) is another logical disease to consider treating by autologous transplant with gene therapy, since normal B-cell development from HSC with a normal Bruton tyrosine kinase *(BTK)* gene should correct the immune deficiency. Because of the good clinical effects from immunoglobulin replacement therapy for XLA and the toxicities from HSCT, especially from chemotherapy and GvHD, HSCT is rarely done for patients with XLA. A few patients with XLA have had allogeneic transplants from healthy donors and have developed B-cell reconstitution.

Gene therapy studies in *BTK* gene knock-out mice have shown that lentiviral vectors with B lymphoid-specific promoters driving BTK expression can lead to immune reconstitution.[29,30] While no adverse effects were seen from constitutive expression of the *BTK* gene in these murine studies, these do not constitute formal toxicology studies, which would be needed before clinical applications. It is likely that regulated expression of *BTK*, rather than constitutive, ubiquitous expression, is needed for highest efficacy and safety. In theory, lentiviral viral vectors using components of the *BTK* gene transcriptional regulatory sequences could yield vectors with the desired expression specificity. Alternatively, gene correction of the *BTK* gene, using the methods discussed below, could restore precise *BTK* expression regulation and may have the greatest safety profile.

Common variable immune deficiency (CVID) comprises the most common severe human PID (Chapter 33). While immunoglobulin replacement therapy can ameliorate the immune deficiency caused by hypogammaglobulinemia, the high rate of other clinical complications that CVID patients may experience necessitate new treatments. However, gene therapy requires knowing the responsible pathogenic gene, and has been limited to monogenic disorders. To date, no single gene defect has been identified in the majority of patients with CVID. There are some known CVID-causing genes, including *TNFRSF13B* (encoding TACI 8% to 10%), *TNFRSF13C* (encoding BAFF-R), NFKB1, ICOS, CD19, and MSH5, which in total may be responsible for 30% to 50% of all patients with CVID.[31] A separate gene therapy project would be needed to develop treatment for each causal gene, through the full spectrum from preclinical activities to clinical trial performance. Moreover, because most of these known CVID-causing genes are involved in cell stimulation and signaling, they may require regulated, rather than ubiquitous, constitutive expression for safety. Because of these constraints, it is not currently possible to apply gene therapy for the majority of patients with CVID.

There has been growing recognition of numerous immune dysregulation and auto-inflammatory syndromes due to deficiency of regulatory T cells (e.g., immunodeficiency, polyendocrinopathy, enteropathy, X-linked {IPEX}), cytotoxic T lymphocyte antigen 4 {CTLA4} and lipopolysaccharide-responsive beige-like anchor protein {LRBA} deficiency, and autosomal dominant gain-of-function mutations (e.g., signal transducer and activator of transcription 3 {STAT3}, MEFV, IL-1, nuclear factor kappa B {NFkB}, interferon pathways). The Primary Immune Deficiency Treatment Consortium named these disorders collectively as primary immune regulatory disorders (PIRD).[32] PIRD are characterized more by autoimmune complications than infections, in contrast to the classical PIDs. The responses of PIRD to allogeneic HSCT have been variable, and, to the extent that GvHD has been a significant component of poor outcomes, autologous gene therapy/HSCT may improve results.[33] Gene modification of a patient's HSC by addition of a shRNA cassette or gene disruption using site-specific endonucleases may suppress expression of the dominant gene. Direct correction of the pathogenic mutation could also be beneficial, as discussed below. It is likely that the gene modification would need to be highly efficient to yield a high fractional correction of the engrafting stem cells. In this setting, some pre-transplant immune suppression may be needed to ablate pre-existing autoimmunity.

Perhaps most challenging for gene therapy are the PIDs that also have major somatic or neurodevelopmental or neurodegenerative problems, such as chromosomal abnormalities, ataxia-telangiectasia, and others. Here, gene therapy with HSC may benefit the immunological components of the disorder but would not address the others; systemic delivery of genes or delivery to the CNS is under study, but it is not yet sufficiently efficient for most clinical needs.

GENE EDITING FOR GENE THERAPY OF PRIMARY IMMUNE DEFICIENCIES

A major paradigm shift in gene therapy is underway as methods of performing precise editing of the genomes of somatic cells are being developed. As an alternative to the semi-random insertion

of normal copies of the relevant gene delivered by a viral vector, as discussed for all of the studies above, techniques are being established to either correct specific bases in the DNA, or to insert (or remove) gene sequences at specific sites by harnessing cellular DNA repair pathways. These DNA repair mechanisms normally correct the many double-stranded DNA breaks that occur during DNA replication or from environmental genotoxic agents (ionizing radiation, chemicals). There are two major DNA repair pathways for rejoining the sequences that flank a double-stranded DNA break: non-homologous end joining (NHEJ) and HR. NHEJ reconnects the broken ends of the chromosomes in a way that often leads to insertion or deletion of DNA bases (indels) at the junctional site. This is, in essence, a mutagenic process and may be used to disrupt genes to knock-out their activity, examples being the HIV-1 co-receptor CCR5, dominant-active transcriptional factors such as CTLA-4, or STAT3 gain-of-function alleles, *BCL11a*, a transcriptional repressor of fetal hemoglobin, etc. There have been clinical trials for patients with HIV infection in which the gene for CCR5 HIV co-receptor was disrupted using a zinc finger nuclease (ZFN) to make a double-stranded DNA break while NHEJ was allowed to repair the break, leading to indels that inactivated the gene and co-receptor expression.[34] More recently, clinical trials have been initiated using either ZFN or clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9 nucleases to disrupt the *BCL11a* gene and induce expression of fetal hemoglobin as a treatment for patients with sickle cell disease and beta-thalassemia.

HR is a more precise repair mechanism that normally uses a copied sister chromosome as a template to repair double-strand DNA breaks that may occur during DNA replication or spontaneously; sequences of the template are copied into the repair site, and, if there are differences in their sequences, the template acts as a donor for the new sequences. For gene correction, an artificial donor template is provided to the cells to instruct the introduction of the intended sequence changes (Fig. 91.3). Beyond its use for modification of single base pairs (bp), as illustrated here, HR can be used to introduce whole gene sequences into the target site by flanking a gene cassette with "homology arms" that consist of the DNA sequences homologous to the target site (Fig. 91.4).

Codon#:	1	2	3	4	5	6	7	8	9
(1) Wild-type Gene	5'-ATG	CCT	TGA	AAT	TCG	GGG	CGA	TTG	ACC-3'
	3'-TAC	GGA	ACT	TTA	AGC	CCC	GCT	AAC	TGG-5'
(2) Mutant Gene	5'-ATG	CCT	TGA	AAT	ACG	GGG	CGA	TTG	ACC-3'
	3'-TAC	GGA	ACT	TTA	TGC	CCC	GCT	AAC	TGG-5'
(3) Donor template				5'-GA	AAT	TCG	GGG	C-3'	
(4) Corrected Gene	5'-ATG	CCT	TGA	AAT	TCG	GGG	CGA	TTG	ACC-3'
	3'-TAC	GGA	ACT	TTA	AGC	CCC	GCT	AAC	TGG-5'

FIG. 91.3 Site-Specific Gene Editing by Homologous Recombination (HR). In this example, instead of the wild-type gene sequence (1), a patient's gene (2) has a mutation from a base-pair substitution of an A for a T at the start of the 5th codon *(red)*. An artificial donor template (3), here as a single-stranded deoxyoligonucleotide of 12 bases in length *(green)*, is provided with the correct base present at the site of the patient's mutation *(blue)*. If the donor template is used to repair a double-stranded DNA break induced near the mutation by a site-specific endonuclease *(red arrow)*, sequences from the donor *(green)* will be incorporated into the patient's gene (4), introducing the normal corrective base pair *(blue)*.

Although HR has been used to introduce genes into cells, such as murine embryonic stem cells to make gene knockout and knock-in mice, it is generally a low-frequency event (occurring in $1/10^6$ to $1/10^4$ cells) and requires the use of selectable markers to isolate the rare desired recombinant. Whereas cloning and expansion of murine (and human) embryonic stem cells can be done to produce populations of the rare recombinant cells, this is not currently possible with primary HSCs, which cannot be expanded to any great degree from single cells while retaining their totipotent differentiation capability. Methods achieving high efficiency of gene modification with low cytotoxicity in large numbers of primary stem cells will be needed for clinical applications to autologous HSCT. A major breakthrough in this area comes from the observation that HR is vastly more frequent when a double-stranded DNA break is introduced close to the target site; then, the repair donor molecule can be used at efficiencies in the range of approximately 1% to 50% of treated cells.[35,36]

Several classes of designer site-specific endonucleases have been derived, including ZFN, introduced above; transcription activator-like effector nucleases (TALENs); and more recently CRISPR/Cas9, all of which permit the introduction of a double-stranded DNA break at unique sites in the mammalian genome with high specificity. Current methods introduce the nuclease into HSCs by a method where it will only be present transiently, for example, electroporation of in vitro transcribed messenger RNA encoding the nuclease proteins or as recombinant protein.

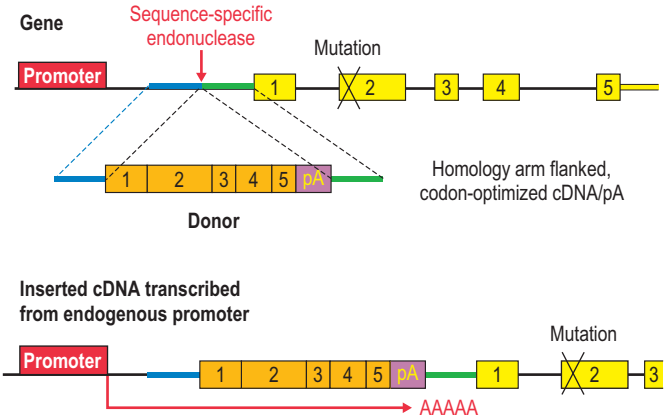


FIG. 91.4 Site-Specific Insertion of a cDNA to Override Downstream Gene Mutations. A prototypical gene is shown with 5 exons (upper diagram, *yellow boxes* 1–5) and an upstream promoter *(red box)*. An inherited mutation in exon 2 (black X) inactivates the gene. A sequence-specific endonuclease *(red arrow)* is designed to introduce a double-stranded DNA break, in this case in the 5′ untranslated region of the gene (*blue* and *green* lines). A cDNA molecule is constructed that contains the contiguous exons of the gene (*orange* 1–5), codon-optimized to increase expression and to eliminate homology with the endogenous exons to avoid illegitimate recombination events, and a polyadenylation signal (pA) to terminate transcription. The cDNA is flanked by the sequences matching the endonuclease cleavage region (*blue* and *green* lines). This donor cDNA is inserted into the nuclease target site by homologous recombination (lower diagram), positioned to be under transcriptional control of the endogenous gene promoter. The resulting cDNA transcript *(red arrow)* would override any downstream mutations in the endogenous gene to produce the correct gene product under the physiological control of the normal gene expression regulatory elements.

The homologous donor template is introduced either as a short single-stranded deoxy-oligonucleotide (e.g., 50 to 100 bases) or longer (200 to 10.000 bp) double-stranded DNA cut from plasmids or produced by polymerase chain reaction that are co-electroporated with the nuclease; or as a long sequence carried by a viral vector that does not integrate into the target cell chromosome, including adeno-associated virus (AAV) or integrase-defective lentiviral vector (IDLV), but efficiently delivers the DNA template to the cell's nucleus.

An important caveat is that potential off-target activity of the nucleases could cause genotoxicity from either disruption of unintended genes or even introduction of chromosomal translocations between two cut sites produced simultaneously (e.g., one on-target and one off-target). Current studies are assessing the consequences of this potential genotoxicity in human HSCs, while ongoing work seeks to improve the specificity of the nucleases to eliminate or greatly minimize off-target activity. An array of assays has been developed to identify off-target cleavage site by endonucleases and, in many cases, nucleases with minimal detectable off-target activity have been found.[37]

Gene editing will have several advantages over gene addition. It would avoid the potential problems from randomly inserting vectors that may disrupt or overstimulate adjacent cellular genes, as discussed above for the retroviral vectors. Crucially, correcting the endogenous gene keeps its expression under normal physiological control. This intrinsic regulation of gene expression will be essential for the gene therapy of many PIDs, where constitutive expression from a viral vector could produce malignancy or other unwanted effects. For example, retroviral vector delivery of a normal CD40 ligand gene corrected murine models of X-linked hyper-IgM syndrome (XHIM), subsequently led to the development of lymphomas in the mice from constitutive rather than regulated CD40 ligand expression.[38] BTK, defective in XLA, similarly may need to be expressed specifically at certain stages of B-lymphocyte development for safety and efficacy.

A long list of other loci involved in PIDs may similarly be best approached by gene correction that allows physiological gene expression control, including RAG1, SLAM, XIAP, JAK3, FOXP3, IL-10, IL7R, TACI, CTLA4, etc. While current gene correction techniques are below the frequency of efficiency needed to yield clinical benefits for most disorders, this field is moving at a lightning pace, with new advances in the activity and specificity of a whole host of nucleases and other genome editing tools.

Newer techniques for site-specific gene editing have been described that do not entail making double-stranded DNA breaks in the genome. Base editors use the Cas9/sgRNA system to bring to a specific genome site a catalytically inactive Cas9 protein fused to an enzyme capable of converting one nucleotide to another.[39] The first of these base editors was a fusion to cytidine deaminase that converts CG base pairs to TA base pairs. Other editors capable of making other nucleotide changes have been described; one day it may be possible to have a library of such editors that can be selected in order to make any desired specific base change.

USE OF PLURIPOTENT STEM CELLS AS A SOURCE OF HEMATOPOIETIC STEM CELLS FOR GENE THERAPY OF PRIMARY IMMUNE DEFICIENCIES

The establishment of human pluripotent stem cells (hPSCs), initially as human embryonic stem cells (hESCs) and subsequently as induced pluripotent stem cells (iPSCs), has brought the promise of novel models to study human diseases ("disease in a dish") and to provide renewable sources of patient-compatible cells for cellular therapies. The essentially unlimited ability to expand hPSCs and their capacity to produce any of the cell types in the body has led to investigations to harness them for regenerative medicines. In the treatment of PIDs, hPSCs could provide an ideal target to produce autologous HSCs with precise gene correction. While techniques have been developed to perform the gene modification strategies that may be sufficiently robust for clinical applications, the major roadblock is the current inability to produce clinically relevant numbers of transplantable HSCs from hPSCs. The HSC is a relatively evanescent state and it has not been possible to "freeze" differentiation from hPSC at that state, although it has been possible to proceed right through the HSC stage to produce relatively pure populations of mature blood lineages. As with gene correction, the pace of scientific progress with these phenomenal cells is proceeding rapidly, and it is likely that new sources of gene-corrected autologous HSCs will be available for clinical transplants in the near future.

GENE THERAPY FOR PRIMARY IMMUNE DEFICIENCIES INVOLVING SERUM PROTEIN DEFICIENCIES

Whereas many of the severe PIDs are due to blood cell defects and thus responsive to HSCT, others result from deficiencies of serum proteins, such as complement components. Here, gene therapy may be approached, as is being done for hemophilia, by direct in vivo administration of a gene-containing vector that can permanently insert the gene into target tissues—such as the liver, skeletal muscle, or endothelium—that can serve as protein factories. In vivo gene delivery is being studied using retroviral and lentiviral vectors or AAV vectors, and methods for in vivo gene correction using site-specific nucleases and homologous donor sequences are also under development. To this end, Crystal and collaborators have recently reported studies in a murine model of hereditary angioedema using an AAV vector to express the *SERPING1* gene encoding the C1 inhibitor.[40]

ADVANCING GENE THERAPY FOR PRIMARY IMMUNE DEFICIENCIES FROM EXPERIMENTAL TO STANDARD OF CARE

Initial trials of gene therapy for PIDs were all performed at tertiary-level academic medical centers, often with federal or disease-foundation research grant funding, and they tested initial hypotheses concerning safety and efficacy. These centers may continue to perform early-phase clinical trials for specific disorders, especially where there is local expertise on the disease being studied. However, there is an ongoing transition to industry-sponsored trials focused on drug product development, as the effective vectors and the gene-modified stem cell products they compose are advanced to licensure and marketing as pharmaceutical agents. A common model used by these new gene therapy companies is one of central processing, at one or a few high-grade commercial good manufacturing practice (GMP) facilities. The autologous patient cells are procured at the patient's home institution (by leukapheresis or bone marrow harvest), shipped to the central processing site for genetic manipulation and cryopreservation, and then returned to be administered locally. Currently, gene therapy transplants have been performed at a limited number of clinical trial sites due to the cell processing expertise needed; presumably under the

commercial model, such transplants may be done at any suitable HSCT center, with the company selling the processed cell product, as now occurs with medical devices or an unrelated stem cell product. Alternatively, self-contained cell processing and manipulation devices are being developed that could allow gene modification of stem cells to be done at individual institutions, without requiring highly trained staff or high-grade GMP facilities.

Major issues remain to be determined about cost and reimbursement for gene therapy. Effective gene therapy for the severe diseases being approached would be expected to lead to large lifetime savings in medical costs. The one-time price can be compared to the costs faced by the patient encumbered by the progressive nature of the underlying disease, to the costs for long-term protein-based therapies and possibly even the costs for allogeneic transplantation. However, those one-time costs are high for clinical gene therapy transplants using pharmaceutical-grade vectors and commercial-level cell processing with the attendant, essential quality control. Thus, a single large expenditure for gene therapy may eventually be cost-effective compared to ongoing medical costs; however, the method of financing the large up-front charge, at least in the United States with its multiple insurance companies, remains to be determined.

Recently approved gene therapies for an inherited retinopathy and B-lineage malignancies are being reimbursed through various models intended to distribute costs over time and in some cases to make reimbursement dependent on the measured benefits of the therapy.

CONCLUSION

In the past decades, gene therapy for PIDs has advanced from a dream for the future to a clinical reality. Gene therapy for ADA SCID, XSCID, WAS, CGD has been safe and effective in most treated patients using ex vivo gene delivery with lentiviral vectors. Approaches to direct gene editing are being developed that may broaden indications for PIDs that may be treated by gene therapy (e.g., XHIM, XLA, etc.). Thus, the continued efforts of scientists and physicians to develop gene therapy are leading to a new therapeutic modality, ideally to permanently and safely cure these diseases.

ON THE HORIZON

- Continued expansion of the genetic types of primary immune deficiencies (PIDs) being treated by gene addition using integrating vectors (e.g., other forms of severe combined immune deficiency [SCID], other forms of chronic granulomatous disease [CGD], leukocyte adhesion deficiency [LAD], hemophagocytic lymphohistiocytosis [HLH], etc.).
- Development of safe and effective marrow conditioning regimens that are not chemotherapy based (e.g., monoclonal antibodies).
- Application of gene editing to a broader spectrum of PIDs: (e.g., X-linked agammaglobulinemia, X-linked hyper-IgM syndrome [CD40 ligand deficiency], recombinase activating gene 1 (RAG1) SCID, IPEX, gain-of-function Signal Transducer and Activator of Transcription 1 [STAT1] disease).
- Understanding the molecular pathogenesis of more PIDs (e.g., CVID, auto-immune single gene disorders) and developing effective gene therapy approaches. Some may require correction of more tissue types than just HSCs (e.g., ataxia telangiectasia, in which neurodegeneration is a predominant feature).
- Production of autologous hematopoietic stem cells (HSCs) by cellular reprogramming of somatic cells to pluripotent state coupled to effective expansion of HSCs.

REFERENCES

1. Stewart FM, Crittenden RB, Lowry PA, et al. Long-term engraftment of normal and post-5-fluorouracil murine marrow into normal nonmyeloablated mice. *Blood*. 1993;81:2566–2571.
2. Blaese RM, Culver KW, Miller AD, et al. T lymphocyte-directed gene therapy for ADA- SCID: initial trial results after 4 years. *Science*. 1995;270:475–480.
3. Aiuti A, Slavin S, Aker MF, et al. Correction of ADA-SCID by stem cell gene therapy combined with nonmyeloablative conditioning. *Science*. 2002;296:2410–2413.
4. Cicalese MP, Ferrua F, Castagnaro L, et al. Update on the safety and efficacy of retroviral gene therapy for immunodeficiency due to adenosine deaminase deficiency. *Blood*. 2016;128:45–54.
5. Gaspar HB, Cooray S, Gilmour KC, et al. Hematopoietic stem cell gene therapy for adenosine deaminase-deficient severe combined immunodeficiency leads to long-term immunological recovery and metabolic correction. *Sci Transl Med*. 2011;3(97):97ra80.
6. Candotti F, Shaw KL, Muul L, et al. Gene therapy for adenosine deaminase—deficient severe combined immune deficiency: clinical comparison of retroviral vectors and treatment plans. *Blood*. 2012;120:3635–3646.
7. Shaw KL, Sokolic R, Davila A, et al. Phase II clinical trial of gene therapy for adenosine deaminase-deficient severe combined immune deficiency (ADA-SCID). *Mol Ther*. 2014;22(suppl. 1):S107.
8. Kohn DB, Shaw KL, Garabedian E, et al. Lentiviral gene therapy with autologous hematopoietic stem and progenitor cells (HSPCs) for the treatment of severe combined immune deficiency due to adenosine deaminase (ADA-SCID): results in and expanded cohort. *Blood*. 2019;134(suppl):3345.
9. Candotti F, Johnston JA, Puck JM, et al. Retroviral-mediated gene correction for X-linked severe combined immunodeficiency. *Blood*. 1996;87:3097–3102. Available from: http://www.ncbi.nlm.nih.gov/pubmed/8605322.
10. Cavazzana-Calvo M, Hacein-Bey S, de Saint Basile G, et al. Role of interleukin-2 (IL-2), IL-7, and IL-15 in natural killer cell differentiation from cord blood hematopoietic progenitor cells and from gamma c transduced severe combined immunodeficiency X1 bone marrow cells. *Blood*. 1996;88:3901–3909.
11. Hacein-Bey-Abina S, Hauer J, Lim A, et al. Efficacy of gene therapy for X-linked severe combined immunodeficiency. *N Engl J Med*. 2010;363:355–364. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2957288&tool=pmcentrez&rendertype=abstract.
12. Gaspar HB, Cooray S, Gilmour KC, et al. Long-term persistence of a polyclonal T cell repertoire after gene therapy for X-linked severe combined immunodeficiency. *Sci Transl Med*. 2011;3:97ra79.
13. Hacein-Bey-Abina S, Pai S-Y, Gaspar HB, et al. A modified γ-retrovirus vector for X-linked severe combined immunodeficiency. *N Engl J Med*. 2014;371:1407–1417. Available from: http://www.nejm.org/doi/abs/10.1056/NEJMoa1404588.
14. De Ravin SS, Wu X, Moir S, et al. Lentiviral hematopoietic stem cell gene therapy for X-linked severe combined immunodeficiency. *Sci Transl Med*. 2016;8:335ra57 Available from: http://stm.sciencemag.org/content/8/335/335ra57.figures-only.
15. Mamcarz E, Zhou S, Lockey T, et al. Lentiviral gene therapy combined with low-dose busulfan in infants with SCID-X1. *N Engl J Med*. 2019;380:1525–1534.
16. Güngör T, Teira P, Slatter M, et al. Reduced-intensity conditioning and HLA-matched haemopoietic stem-cell transplantation in patients with chronic granulomatous disease: a prospective multicentre study. *Lancet*. 2014;383:436–448. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24161820.
17. Kang EM, Choi U, Theobald N, et al. Retrovirus gene therapy for X-linked chronic granulomatous disease can achieve stable long-term correction of oxidase activity in peripheral blood neutrophils. *Blood*. 2010;115:783–791.
18. Ott MG, Schmidt M, Schwarzwaelder K, et al. Correction of X-linked chronic granulomatous disease by gene therapy, augmented by insertional

activation of MDS1-EVI1, PRDM16 or SETBP1. *Nat Med*. 2006 Apr;12(4):401–409.
19. Santilli G, Almarza E, Brendel C, et al. Biochemical correction of X-CGD by a novel chimeric promoter regulating high levels of transgene expression in myeloid cells. *Mol Ther*. 2011;19:122–132.
20. Kohn DB, Booth C, Kang EM, et al. Net4CGD consortium. Lentiviral gene therapy for X-linked chronic granulomatous disease. *Nat Med*. 2020 Feb;26(2):200–206.
21. Becker-Herman S, Meyer-Bahlburg A, Schwartz MA, et al. WASp-deficient B cells play a critical, cell-intrinsic role in triggering autoimmunity. *J Exp Med*. 2011;208:2033–2042.
22. Boztug K, Schmidt M, Schwarzer A, et al. Stem-cell gene therapy for the Wiskott-Aldrich syndrome. *N Engl J Med*. 2010;363:1918–1927.
23. Charrier S, Dupré L, Scaramuzza S, et al. Lentiviral vectors targeting WASp expression to hematopoietic cells, efficiently transduce and correct cells from WAS patients. *Gene Ther*. 2007;14:415–428. Available from: http://www.nature.com/doifinder/10.1038/sj.gt.3302863%5Cnpapers3://publication/doi/10.1038/sj.gt.3302863.
24. Scaramuzza S, Biasco L, Ripamonti A, et al. Preclinical safety and efficacy of human CD34$^+$ cells transduced with lentiviral vector for the treatment of Wiskott-Aldrich syndrome. *Mol Ther*. 2013;21:175–184. Available from: https://doi.org/10.1038/mt.2012.23/nature06264.
25. Hacein-Bey Abina S, Gaspar HB, Blondeau J, et al. Outcomes following gene therapy in patients with severe Wiskott-Aldrich syndrome. *JAMA*. 2015;313:1550–1563. Available from: http://jama.jamanetwork.com/article.aspx?doi=10.1001/jama.2015.3253.
26. Aiuti A, Biasco L, Scaramuzza S, et al. Lentiviral hematopoietic stem cell gene therapy in patients with Wiskott-Aldrich syndrome. *Science*. 2013;341:1233151 Available from: http://www.ncbi.nlm.nih.gov/pubmed/23845947.
27. Pala F, Morbach H, Castiello MC, et al. Lentiviral-mediated gene therapy restores B cell tolerance in Wiskott-Aldrich syndrome patients. *J Clin Invest*. 2015;125:3941–3951. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4607131&tool=pmcentrez&rendertype=abstract.
28. Bauer Jr TR, Hickstein DD. Gene therapy for leukocyte adhesion deficiency. *Curr Opin Mol Ther*. 2000;2:383–388.
29. Ng YY, Baert MRM, Pike-Overzet K, et al. Correction of B-cell development in Btk-deficient mice using lentiviral vectors with codon-optimized human BTK. *Leukemia*. 2010;24:1617–1630.
30. Kerns HM, Ryu BY, Stirling BV, et al. B cell-specific lentiviral gene therapy leads to sustained B-cell functional recovery in a murine model of X-linked agammaglobulinemia. *Blood*. 2010;115:2146–2155.
31. Abolhassani H, Hammarström L, Cunningham-Rundles C. Current genetic landscape in common variable immune deficiency. *Blood*. 2020;135:656–667.
32. Chan AY, Leiding JW, Liu X, et al. Hematopoietic cell transplantation in patients with primary immune regulatory disorders (PIRD): a Primary Immune Deficiency Treatment Consortium (PIDTC) survey. *Frontiers in Immunology*. 2020 https://doi.org/10.3389/fimmu.2020.00239.
33. Hagin D, Burroughs L, Torgerson TR. Hematopoietic stem cell transplant for immune deficiency and immune dysregulation disorders. *Immunol Allergy Clin North Am*. 2015;35:695–711.
34. Tebas P, Stein D, Tang WW, et al. Gene editing of CCR5 in autologous CD4 T cells of persons infected with HIV. *N Engl J Med*. 2014;370:901–910. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4084652&tool=pmcentrez&rendertype=abstract.
35. Rouet P, Smih F, Jasin M. Expression of a site-specific endonuclease stimulates homologous recombination in mammalian cells. *Proc Natl Acad Sci U S A*. 1994 Jun 21;91(13):6064–6068.
36. Porteus MH, Baltimore D. Gene targeting in human cells. *Science*. 2003;300(May):75390.
37. Tsai SQ, Joung JK. Defining and improving the genome-wide specificities of CRISPR-Cas9 nucleases. *Nat Rev Genet*. 2016;17:300–312.
38. Brown MP, Topham DJ, Sangster MY, et al. Thymic lymphoproliferative disease after successful correction of CD40 ligand deficiency by gene transfer in mice. *Nat Med*. 1998;4:1253–1260.
39. Komor AC, Kim YB, Packer MS, et al. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature*. 2016;533:420–424.
40. Qiu T, Chiuchiolo MJ, Whaley AS, et al. Gene therapy for C1 esterase inhibitor deficiency in a Murine Model of Hereditary angioedema. *Allergy*. 2019 Jun;74(6):1081–1089.

92

Hematopoietic Stem Cell Transplantation for Malignant Diseases

Pashna N. Munshi, Scott D. Rowley, and Robert Korngold

Hematopoietic stem cell transplantation (HSCT) is effective treatment for most hematological malignancies, including leukemia, lymphoma, multiple myeloma (MM), and clonal myelodysplastic and myeloproliferative diseases (MPDs), as well as a variety of nonmalignant diseases, such as autoimmune disorders and hemoglobinopathies. Autologous HSCT (auto-HSCT) is commonly used as therapy for patients with malignancies sensitive in a dose-responsive manner to chemotherapy or radiotherapy. These patients receive intensive cytoreductive regimens designed to eliminate all tumor cells; in so doing, however, they also destroy the patient's hematopoietic function needed for blood formation. Infusion of previously collected hematopoietic stem cells (HSCs) will rescue the patient from the marrow-ablative effects of this treatment. Allogeneic HSCT (allo-HSCT), in addition to reconstitution of bone marrow function, achieves an immunotherapeutic benefit from the donor natural killer (NK) and T cells within the graft attacking residual tumor cells that persist after the conditioning regimen, thereby greatly reducing the risk of later relapse of the disease. Thus allo-HSCT, in contrast to auto-HSCT, does not require administration of dose-intensive regimens to achieve complete tumor cell kill, and lower-risk nonmyeloablative regimens may be used to "condition" the host for transplantation.

Auto-HSCT (including syngeneic twins) is justified by the dose sensitivity of most hematological malignancies. Although there is some evidence that a more robust immunological recovery after auto-HSCT predicts lower risk of relapse, possibly opening an area of research in graft (or host) modification to enhance such recovery, treatment of the disease is primarily a result of the dose-intensive, myeloablative chemotherapy or radiotherapy administered. Infusion of cells is required only to recover hematopoiesis, and the stem cell infusion is, therefore, intended to treat the deleterious effect of chemotherapy on bone marrow function and not the disease itself. Marrow-toxic agents that cannot easily be incorporated into standard treatment regimens can be used in auto-HSCT. The primary complications of auto-HSCT result from the administration of a dose-intensive regimen and include a period of marrow hypoplasia, possibly requiring blood transfusions and antibiotics. Nonhematological toxicities, including mucositis resulting in inanition and diarrhea, and damage to other organs such as the lung, liver, and kidney, limit the amount of chemotherapy that can be administered. Currently, the treatment-related mortality (TRM) risk for most treatment regimens is ≤5%. Relapse of disease, particularly for patients who come to transplantation with chemotherapy-refractory disease, is the primary cause of failure of auto-HSCT. Improvements in the outcome of auto-HSCT will require innovative strategies to reduce this risk.

THERAPEUTIC PRINCIPLES

Autologous and Allogenic Transplantation

Autologous Transplantation

- Based on chemotherapy or radiotherapy dose sensitivity of disease being treated
- Requires collection and storage of adequate hematopoietic stem cells (HSCs), preferably before extensive alkylating agent or purine analogue therapy
- Lower risk of graft failure (no immunological rejection)
- No routine posttransplantation immunosuppression
- Minimal risk of graft-versus-host disease (GvHD)
- No graft-versus-tumor (GvT) effect
- More rapid posttransplantation immune reconstitution
- Risk of tumor cell contamination in HSC product
- Not useful for diseases in which normal HSCs cannot be collected (e.g., chronic myelogenous leukemia, myelodysplasia)

Allogeneic Transplantation

- Will rescue bone marrow function if dose-intensive therapy is administered
- Effective with reduced-intensity conditioning regimens
- Achieves a GvT effect in many malignancies
- Risk of GvHD distinct from the beneficial GvT effect
- Higher risk of transplantation-related complications that may offset the benefit of the GvT effect
- Risk of immunological graft rejection
- Slower posttransplantation immune reconstitution
- No risk of tumor-cell contamination with healthy donor

Allo-HSCT has a much lower relapse risk compared with auto-HSCT as a result of a beneficial immunological graft-versus-tumor (GvT) effect achieved by engraftment of the donor immune system. Allograft recipients, however, face a much higher risk of TRM from the detrimental immunological GvH response against healthy tissues of the patient. The principal nonrelapse complication of allo-HSCT is graft-versus-host disease (GvHD), which can occur early (acute GvHD, within the first several weeks) (Table 92.1) or late (chronic GvHD, months to several years) after transplantation. The rate of overall incidence of moderate to severe acute GvHD (aGvHD) is 35% to 80% for all patients receiving cells from a human leukocyte antigen (HLA)–matched related or unrelated stem cell donor, and aGvHD is a primary cause of death in 10% to 20% of these patients. Chronic GvHD (cGvHD), a clinicopathologically distinctive form of this alloreaction, occurs in up to 80% of recipients; it may involve aspects of regulatory T-cell (Treg) dysfunction and autoimmune-like responses, and may require years of therapy before tolerance is achieved, allowing withdrawal of immunosuppressant medications. Other

TABLE 92.1 Acute Graft-versus-Host Disease

Clinical Staging

Stage	Skin	Liver	Gut
1	Maculopapular rash <25% BSA	Bilirubin 2–3 mg/dL	Diarrhea 500–1000 mL/day, or persistent nausea
2	Maculo-papular rash 25%–50% BSA	Bilirubin 3–6 mg/dL	Diarrhea 1000–1500 mL/day
3	Generalized erythroderma	Bilirubin 6–15 mg/dL	Diarrhea >1500 mL/day
4	Desquamation and bullae formation	Bilirubin >15 mg/dL	Pain ± ileus

Clinical Grading

Overall Grade	Skin	Liver	Gut	Functional Impairment
0 (none)	0	0	0	0
1 (mild)	1–2	0	0	0
2 (moderate)	1–3	1	1	1
3 (severe)	2–3	2–3	2–3	2
4 (life-threatening)	1–4	1–4	1–4	3

BSA, Body surface area.
Adapted from: Harris AC, Young R, Devine S, et al. International, multicenter standardization of acute graft-versus-host disease clinical data collection: a report from the Mount Sinai Acute GVHD International Consortium. *Biol Blood Marrow Transplant*. 2016;22:4–10.

significant complications of allo-HSCT relate to problems of inadequate reconstitution of the immune system of the patient and the concomitant risk of opportunistic infections, including viral and fungal infections that are rarely observed in healthy hosts. Despite the apparent risks associated with allo-HSCT, this treatment holds great promise as a curative therapy for several tumor types, particularly with strategies either to enhance the efficacy of the GvT effect or decrease the toxicity of the graft-versus-host (GvH) response. Efforts have been made in recent years to make HSCT less toxic by development of low-dose, nonmyeloablative conditioning regimens that allow the treatment of older patients or those with comorbid health issues that otherwise preclude treatment with high doses of chemotherapy.[1]

IMMUNE MECHANISMS RELATED TO ALLO-HEMATOPOIETIC STEM CELL TRANSPLANTATION

Histocompatibility

The HLA major histocompatibility complex (MHC; Chapter 5) is the primary consideration in the selection of a donor for allo-HSCT, since its loci contribute significantly to host-versus-graft (HvG), leading to immunological rejection of donor HSC, and to GvH (leading to GvHD and GvT) reactions. HLA antigens are classified as class I (HLA-A, -B, -C) and class II (HLA-DR, -DQ, -DP) molecules, and typing of donors and patients can be performed using low- or high-resolution techniques. The low-resolution serological techniques can determine a phenotypic mismatch (e.g., A02 vs A03), whereas high-resolution molecular techniques can identify allelic genotypic differences (*e.g.*, HLA-*A*02:01* vs. *HLA-A*02:02*). The risks of aGvHD, cGvHD, and transplantation-related mortality increase with the number of HLA mismatches, and ideally, genotypically matched unrelated donors are sought for patients lacking an HLA-identical sibling. Disparity at HLA-A, -B, -C, and DRB1 alleles are definite risk factors for survival after unrelated donor transplantation, whereas single HLA-DQ or -DP mismatches appear to be better tolerated and/or permissive, meaning they do not have a deleterious impact on clinical outcome.[2]

Not all mismatches result in a deleterious clinical outcome, and identification of permissive mismatches will increase the number of donors available to the patient. Algorithms accounting for possible permissive mismatching in the selection of unrelated donors are being developed and tested.[3]

GvHD and GvT occurring after HLA-compatible sibling transplantation demonstrate the importance of minor histocompatibility antigens (miHAs) on the outcome of allo-HSCT. miHAs are derived from polymorphic sites in normal proteins between individuals and are constantly processed by proteasome activity and presented on the cell surface by MHC molecules; they can thus be recognized by T cells from an HLA-matched donor. As a result, hundreds of miHAs may be variably expressed on host tissues and can trigger an alloresponse from donor T cells, thereby causing GvHD (and GvT). Unfortunately, only a few miHAs have been identified, such as those on the Y chromosome, leading to a higher risk of GvHD in male recipients of cells from female donors.[4] miHA matching of a male donor for a male recipient is not a component of current algorithms for donor selection.

Research in the field of additional criteria for donor selection is ongoing. Human NK cells possess clonally distributed, inhibitory receptors termed "killer cell immunoglobulin-like" (KIR) receptors that recognize epitopes shared by groups of HLA class I alleles (KIR ligands). KIR–ligand mismatching in the GvH direction appears to result in lower risk of relapse- and nonrelapse-related mortality after allo-HSCT. Activating KIRs transduce signals to activate NK cells, and the presence of these cells is associated with a lower risk of leukemia relapse after unrelated and haploidentical transplantation as well as protection against certain viral infections, such as human immunodeficiency virus/acquired immunodeficiency syndrome (HIV/AIDS) or hepatitis C virus (HCV) infection.

Graft-versus-Host Disease

GvHD is caused by mature donor T cells within the HSC inoculum, which recognize HLA or miHA differences expressed by host antigen-presenting cells (APCs) and tissues (Fig. 92.1).[5] Cytokines released from host cells after a patient has received dose-intensive tumor cytoreductive chemotherapy or radiotherapy conditioning create an inflammatory environment that enables the generation of a response of infused donor T cells against host antigens. This initiates a cascade of T-cell activation events, which results in proliferation, release of additional inflammatory cytokines, and the generation of effector T cells that can infiltrate target tissue, particularly the lymphoid system, intestinal tract, skin, and liver and mediate the destruction of host cells in those organs. Both CD4 and CD8 T cells can be involved in GvHD, depending on the specific class I or class II HLA or miHA disparities involved.

The simplest way to avoid the development of GvHD is to deplete the donor HSC graft of T cells before infusion to a cell

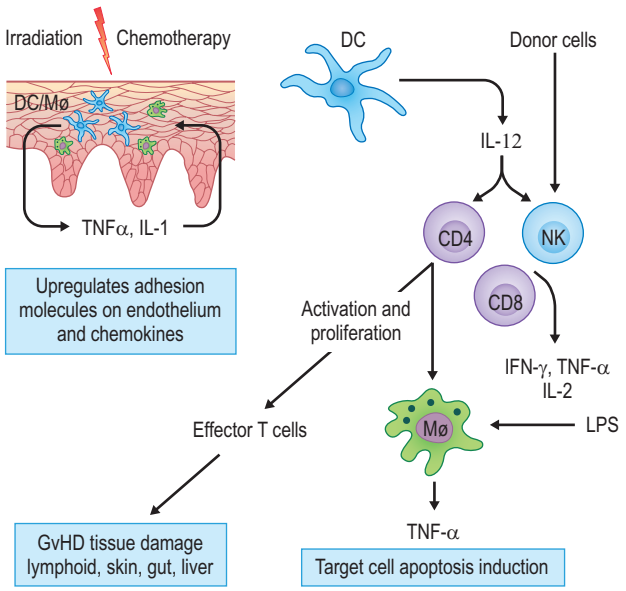


FIG. 92.1 Development of Graft-versus-Host Disease. Cytoreductive preconditioning treatment of patients with hematological malignancies with total body irradiation (TBI) or chemotherapeutic drugs causes damage to epithelium in the skin and gastrointestinal tract and activates the release of inflammatory cytokines by dendritic cells *(DCs)* and macrophages *(Mϕ)* in those tissues. These cytokines include tumor necrosis factor-*α (TNF-α)* and interleukin *(IL)*-1, which upregulate adhesion molecules and chemokine release in the vascular endothelium of the tissues. Activated DCs also migrate to the lymphoid system, where they can present recipient histocompatibility antigens to infused donor T cells that are in the hematopoietic stem cell graft. DCs release IL-12, which helps activate CD4 and CD8 T cells, as well as natural killer *(NK)* cells. These responding cells proliferate and produce additional inflammatory cytokines, including interferon-*γ (IFN-γ)*, TNF-*α*, and IL-2. M*ϕ* are activated by both IFN-*γ* and lipopolysaccharide *(LPS)* produced by bacteria found in the intestinal tract, and these cells then produce high levels of more TNF-*α*. TNF-*α* has many properties, including direct induction of apoptosis on cells in the tissues of target organs of graft-versus-host disease *(GvHD)*, but it also helps effector T cells to home to and enter tissue sites through the vascular endothelium. Effector T cells specific for host histocompatibility antigens then get reactivated and perform their effector functions, including release of inflammatory and cytolytic cytokines and direct killing of recipient-type cells in the lymphoid compartment and in the skin, gut, and liver.

KEY CONCEPTS

Graft-versus-Host Disease

- Caused by donor–recipient differences in:
  - Major histocompatibility complex (MHC) molecules
  - Minor histocompatibility antigens (miHAs)
- Mediated by mature donor CD4 and/or CD8 T cells
- Requires inflammatory cytokines
- Primary target organs include lymphoid system, skin, gastrointestinal tract, and liver
- Distinct forms of presentation and target organ involvement classified as acute or chronic

dose below 10^5 cells/kg body weight. This approach has succeeded in significantly diminishing the incidence of GvHD, but other complications related to the ensuing delay in immune reconstitution of recipients, an increased risk of relapse from loss of the GvT effect, and a higher rate of engraftment failure (also from loss of the GvH effect) have resulted in inconsistent improvement in long-term survival, compared with T cell–replete products. The conundrum is that the same alloreactive donor T cells that mediate GvHD can also cause a GvT response, although there may also be additional tumor-specific or hematopoietic tissue-specific T cells.[6] Thus the overriding goal is to be able to manipulate the donor HSC inoculum in such a way as to avoid GvHD, but to still be able to mediate a GvT effect.

Clinical Aspects of Acute Graft-versus-Host Disease

Usually developing within the first 3 months after transplantation, aGvHD is a clinical diagnosis with characteristic, but nondiagnostic, pathological findings, with the most common presenting manifestations including skin rash, nausea, anorexia, diarrhea and jaundice, depending on the target organ(s) most affected.[7] In addition to the increased risks of developing aGvHD related to the extent of HLA and miHA disparity, additional factors include the advanced age of either the donor or the recipient, gender disparity (female donor–male recipient), donor parity (female donors), and infusion of T-cell–replete HSC products. Conditioning with reduced-intensity regimens, with lower regimen-related toxicities to nonhematological tissues, also results in a lower risk of aGvHD and may delay the onset of its initial manifestation. Interactions between microbial-associated molecules and innate immune receptors (*e.g.*, Toll-like receptors [TLRs]) appear to be involved in GvHD pathogenesis, as demonstrated in both murine models and human transplantation. Research has shown an interaction between the host's gut microbiota and immune system.[8] On the basis of this knowledge, experiments to decrease transplantation-related GvHD by altering the gut microbiome are being actively pursued.

Pharmacological agents are the mainstay of aGvHD prophylaxis. Most patients receive a combination of a calcineurin inhibitor (tacrolimus or cyclosporine) along with an antimetabolite, such as methotrexate or mycophenolate mofetil (MMF). Methotrexate is associated with delayed engraftment, mucositis, idiopathic pneumonia syndrome, and other transplantation-related complications, which has prompted the development of other combination regimens, such as a calcineurin inhibitor in combination with sirolimus or MMF, or reduced doses of methotrexate. The addition of anti-thymocyte globulin (ATG) to the conditioning regimen lowers the incidence of both GvH and HvG reactions because of its persistence for several days after HSC infusion, effectively depleting T cells from the graft as well as the host. However, patients treated with ATG may face higher risks of infectious complications, including Epstein-Barr virus (EBV)–associated posttransplantation lymphoproliferative disorder as a result of the greater immunosuppression achieved.

The administration of high-dose cyclophosphamide, an alkylating agent of the nitrogen mustard family, after transplantation reduces the risks of acute and/or chronic GvHD and is a prime example of drug-induced immunological tolerance, a concept first demonstrated by Schwartz and Dameshek in their experiments using 6-mercaptopurine.[9] Subsequently, in 1963, Berenbaum and Brown demonstrated immune tolerance to skin allografts in adult mice by using cyclophosphamide.[10] Posttransplantation cyclophosphamide given on days 3 and 4 after transplantation is now

a standard regimen used in haploidentical allogeneic transplantations in combination with other immunosuppressive drugs and as a single agent in matched unrelated and related transplantations, reducing the complications of graft failure and life-threatening aGvHD associated with transplantation of HLA disparate grafts.[11] Given the successes of this approach, posttransplantation cyclophosphamide is now used for GvHD prophylaxis in unrelated donor transplants with comparable outcomes of survival, GvHD,[12] and immune reconstitution.

Glucocorticoids with a calcineurin inhibitor remain the standard approach to initial systemic management of clinically significant aGvHD. About 30% to 50% of patients will respond to initial therapy, and patients who fail to respond have a poor prognosis, as additional agents added for control greatly increase the risk of opportunistic infections and other treatment-related complications. The use of higher doses of corticosteroids, or the addition of ATG, for example, in the initial treatment of aGvHD do not improve patient outcomes and should be reserved for patients who fail initial therapy. A number of drugs, including ATG, pentostatin, switching to tacrolimus from cyclosporine, and newer monoclonal antibodies (mAbs) have shown limited activity in the salvage treatment of patients with steroid-refractory aGvHD. Extracorporeal exposure of peripheral blood mononuclear cells (PBMCs) to the photosensitizing agent 8-methoxypsoralen and ultraviolet A (UVA) radiation (photopheresis) is effective in the treatment of selected diseases mediated by T cells, including both aGvHD and cGvHD, although the mechanism of this effect remains to be elucidated. The recent approval of ruxolitinib[13] for steroid-refractory GvHD is a result of better understanding of lymphocyte activation pathways and provides encouraging treatment options for this otherwise challenging clinical condition.

Autologous Graft-versus-Host Disease

A form of aGvHD can also occur after auto-HSCT, probably as a manifestation of immune system dysregulation during reconstitution of the immune system after dose-intensive therapy. Initially studied in murine models of auto-HSCT, it was demonstrated that the abrupt withdrawal of cyclosporine could induce clinical features indistinguishable from those observed after allo-HSCT. A mechanism that involves the depletion of central memory cells has been proposed, although depletion of Treg with inhibition of peripheral tolerance may also be involved. Autologous GvHD is also reported in patients not receiving posttransplantation immunomanipulations (spontaneous GvHD). Based on the clinical benefits of GvT observed in allo-HSCT, several clinical trials with administration of cyclosporine with or without interferon (IFN) were performed in patients undergoing auto-HSCT without obvious GvT effect.[14] Both induced and spontaneous autologous aGvHD are usually self-limiting complications, and are usually easily managed with a course of corticosteroids in patients with more extensive involvement, in contrast to the more extensive and difficult-to-manage aGvHD that occurs after allo-HSCT. Furthermore, cGvHD does not occur after auto-HSCT.

A spontaneously occurring severe, and sometimes fatal, reaction, "engraftment syndrome," involves the gastrointestinal (GI) system with severe diarrhea during the periengraftment period in some patients undergoing auto-HSCT for MM and other malignancies.[15] The presentation is clearly distinct from the spontaneous or induced GvHD previously observed in patients undergoing auto-HSCT. Engraftment syndrome may be a consequence of prior treatments with immunomodulatory drugs, such as lenalidomide or thalidomide, the proteasome inhibitor bortezomib, or high doses of potent glucocorticoids before collection and storage of autologous HSCs, resulting in a period of immune dysregulation occurring early after autologous HSCT. Patients with engraftment syndrome may require intensive treatment with corticosteroids and a calcineurin inhibitor, with courses more similar to those experienced by patients undergoing allo-HSCT.

Clinical Aspects of Chronic Graft-versus-Host Disease

cGvHD is the leading cause of late TRM among those undergoing allo-HSCT and resembles autoimmune disorders, such as scleroderma, Sjögren syndrome, and primary biliary cirrhosis. Diagnosis is, as with aGvHD, based on clinical observations and secondary confirmation with laboratory or pathology tests (Table 92.2). A falling performance status, progressive weight loss, or recurrent infections are typical signs of severe cGvHD. About 50% of long-term survivors will develop cGvHD at a median of 9 months after transplantation, and patients must be monitored closely for this

TABLE 92.2 Staging of Chronic Graft-versus-Host Disease (GvHD)

Target Organ	Score 0	Score 1	Score 2	Score 3
Performance score	KPS 100%	KPS 80%–90%	KPS 60%–70%	KPS <60%
Skin	No symptoms	<18% BSA	19%–50% or sclerotic, still able to pinch	>50% or "hidebound"
Mouth	No symptoms	Mild symptoms, no limitations	Moderate symptoms, decreased oral intake	Severe symptoms with major decrease in intake
Eyes	No symptoms	Mild dry eyes	Moderate dry eyes, drops >3×/day	Severe dry eyes affecting daily activities
Gastrointestinal tract	No symptoms	Symptoms without weight loss	Symptoms with moderate weight loss (5%–15%)	Symptoms with weight loss >15%
Liver	Normal LFTs	LFTs elevated <2× upper limits of normal	LFTs elevated 2–5× upper limits of normal	LFTs elevated >5× upper limits of normal
Lungs	No symptoms	Mild symptoms FEV 60%–79%	Moderate symptoms FEV 40%–59%	Severe symptoms FEV <40%
Joints and fascia	No symptoms	Mild tightness not affecting daily activities	Tightness affecting daily activities	Contractures with significant loss of range of motion
Female genital tract	No symptoms	Symptomatic with middle signs on examination	Symptomatic with dyspareunia	Symptomatic with strictures

At least one diagnostic and one distinctive sign is necessary to make a diagnosis of cGvHD.
BSA, Body surface area; *FEV*, forced expiratory volume; *KPS*, Karnofsky performance status; *LFT*, liver function test.
Adapted from: Filipovitch AH, Weisdorf D, Pavletic S, et al. National Institutes of Health Consensus Development Project on Criteria for Clinical Trials in Chronic Graft-versus-Host Disease: 1. Diagnosis and Staging Working Group Report. *Biol Blood Marrow Transplant.* 2005;11:945–956.

complication for at least 3 years so that appropriate treatment can be initiated before extensive end-organ damage ensues. Permanent ocular or oral sicca syndrome and pulmonary dysfunction can result if cGvHD is not appropriately treated, leading to a marked deterioration in quality of life.

Factors predictive of the development of cGvHD include the degree of HLA and miHA disparity, as well as prior aGvHD, older patient age, the source of HSCs (greater risk after peripheral blood stem cell [PBSC] transplantation than after bone marrow transplantation), gender (female donor–male recipient), and donor lymphocyte infusion (DLI) after transplantation. Inflammatory events, such as sunburn or surgical procedures, can precipitate cGvHD. Patients who develop cGvHD have a higher risk of TRM but a lower risk of relapse as a result of the immunological GvT effect.[16] T-cell depletion of the graft or treatment with ATG may decrease the risk of cGvHD, although this has not been demonstrated in all studies. Most patients require at least two drugs for effective treatment of cGvHD, with the standard initial treatment being glucocorticoids and a calcineurin inhibitor. About half the patients do not achieve a complete remission (CR) with first-line therapy, although the manifold signs and symptoms of cGvHD complicate the definition of response to treatment. There are no clear recommendations regarding second-line treatments and various pharmacological and immunological techniques have been used. Photopheresis has an overall response rate of 50% to 60%, with many patients achieving CR. At least transient responses can be achieved with treatment with rituximab, an anti-CD20 chimeric antibody, which illustrates the humoral immunity contribution to cGvHD. The demonstration of activation of the Janus kinase (JAK) pathways in activated T lymphocytes has led to studies demonstrating a potential benefit of Jak2 inhibitors in the management of steroid-refractory cGvHD.

CLINICAL PEARLS

Factors Predicting Chronic Graft-versus-Host Disease

- Degree of human leukocyte antigen (HLA) incompatibility:
  - Major histocompatibility complex (MHC)
  - Minor histocompatibility antigen (miHA) disparity
- Presence of prior acute GvHD
- Source of stem cells (higher risk in peripheral blood vs. bone marrow)
- Donor gender (female donor → male recipient)
- Use of donor lymphocyte infusion (DLI) following hematopoietic stem cell transplantation (HSCT)
- Inflammatory events; surgeries, phototoxicity, alcohol consumption

Graft-versus-Tumor Responses

Although recognized in mouse models of transplantation in the 1950s, the first clinical report of a relationship between GvHD and GvT was published in 1979.[17] This relationship between the incidence (but not the severity) of aGvHD or cGvHD and the relapse rate of chronic myeloid leukemia (CML) and, to a lesser extent, acute myeloid leukemia (AML), acute lymphocytic leukemia (ALL), and MM has been observed in patients who received allo-HSCT. Relapses after HSCT occur because of survival of malignant cells after administration of the pretransplantation conditioning regimen and their outgrowth several months later. Relapse remains the major cause of treatment failure after allo-HSCT despite the intended GvT effect of this treatment modality. The ability to mediate an effective GvT response likely depends on several factors, including the presentation of appropriate antigens by MHC class I and/or class II molecules on the tumor cells that can be recognized by effector CD4 or CD8 T cells; lack of strong Treg activity that may be induced by cytokines from the tumor cells; tumor cell susceptibility to lysis by effector T cells (*e.g.*, the level of B-cell lymphoma 2 [BCL-2] expression and the ability to resist apoptosis induction); ability of T cells to home to sites of tumor growth; and the direct effect of immunosuppressive cytokines, such as transforming growth factor-β (TGF-β), produced by the tumor cells.[18] Many types of tumor cells downregulate expression of MHC on their surface, and perhaps CML and AML are most susceptible to GvT responses because the myeloid lineage is adapted for antigen presentation and high MHC expression. A number of novel immunotherapeutic approaches are being developed to overcome these obstacles and enhance GvT responses, keeping in mind that GvHD has to also be avoided or minimized to improve outcomes (Fig. 92.2). Immunotherapy remains the major strategy to combat relapse occurring after allo-HSCT. This can be achieved by reducing immunosuppression, offering a second allo-HSCT, or infusing additional lymphocytes from the HSC donor (DLI) or immune checkpoint blockade (*e.g.*, ipilimumab, nivolumab, pembrolizumab) to "boost" the donor T cells to attack the cancer.[19] As one can expect, all these modalities pose the challenge of associated morbidities, particularly GvHD.[20] More defined immunotherapeutic approaches using chimeric antigen receptors (CARs) expressed on T cells (CAR T cells; Chapter 81) also show some promise to target residual tumor cells that remain after allo-HSCT, with a possible low risk of GvHD.

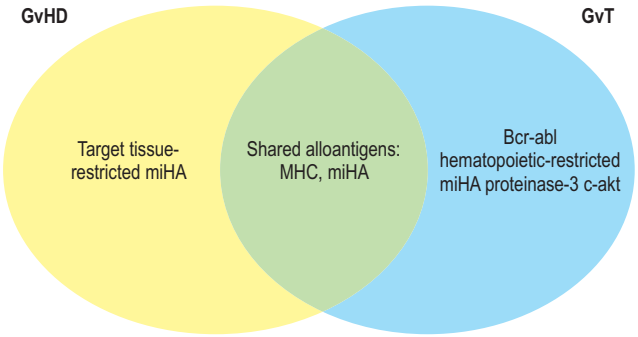


FIG. 92.2 Graft-versus-Host Disease and Graft-versus-Tumor Responses. Donor T-cell responses to recipient antigens can cause graft-versus-host disease *(GvHD)* but can also target residual leukemia cells. T cells causing GvHD may recognize ubiquitous or tissue-restricted antigens (either major histocompatibility complex *[MHC]* or minor histocompatibility antigens *[miHAs]*). Many of these recipient antigens may also be expressed by the leukemia cells and allows for a graft-versus-tumor *(GvT)* response. Additional leukemia-specific (*e.g.*, bcr-abl, proteinase 3, or c-akt) or tissue-restricted antigens (some miHAs, as those expressed only by certain lineages of hematological cells) may be dominantly expressed by the tumor cells and can be targeted by donor T cells without causing GvHD.

KEY CONCEPTS

Major Issues Related to Success of Allogeneic Hematopoietic Stem Cell Transplantation

- Graft-versus-host disease (GvHD)
- Graft-versus-tumor (GvT) responses
- Kinetics and completeness of immune reconstitution
- Opportunistic infections from delayed immune reconstitution
- Patient comorbid conditions increasing risk of toxicities
- Chemotherapy sensitivity of disease being treated

Immunomodulation is frequently the first treatment option for patients in relapse after allo-HSCT, with rapid withdrawal of immunosuppressive medications and infusion of donor T cells.[21] This treatment approach requires that donor T-cell chimerism be sustained to avoid rejection of effector cells through a HvG mechanism. Patients with low-grade lymphoid malignancies, such as CLL, indolent non-Hodgkin lymphoma (NHL), mantle cell NHL, and CML have the highest likelihood of response. However, it remains the main treatment strategy to treat relapse following allo-HSCT. DLI can be used in combination with chemotherapy to maintain disease control during the time required for the development of the GvT effect.[22]

The number of lymphocytes infused is important in achieving the DLI effect, although it may be possible to induce GvT using doses of lymphocytes that are less likely to result in GvHD. A large retrospective analysis demonstrated that using an initially lower cell dose reduced GvHD and improved survival.[23] One method to reduce GvHD risk is to genetically insert suicide genes into the T cells being given for DLI, allowing specific ablation of these cells if this complication occurs.[24] Another method to lower GvHD rates complicating DLI is to deplete the DLI product of GvHD-inducing cells. Studies in mice have shown that naïve subset of CD8 T cells lead to more GvHD, whereas the effector memory subsets (T_{EM}) of CD8 and CD4 T cells moderate the graft-versus-leukemia (GvL) effect without causing GvHD.[25,26] The delay in response between DLI and the development of a GvT effect suggests that only a minority of the cells infused recognize the tumor cell antigens and must undergo in vivo expansion before the therapeutic effect is achieved. It may be possible to develop leukemia-specific cytotoxic T cells in the laboratory, decreasing the delay in effect and possibly increasing the GvT potential. Donor immunity can be transferred, at least transiently, to the host, as demonstrated by delayed-type transfusion reactions to host red blood cells (RBCs), mediated by donor lymphocytes transfused along with the HSC product.[27] Adoptive transfer of donor immunity against specific targets can be achieved, but its persistence requires immunization of both the donor and the recipient. A clinical utility for such manipulation of the immune system has not yet been demonstrated, but it is clearly of interest as a technique to prevent posttransplantation infections and/or disease relapse.

A rapid and major leap in the field of adoptive transfer of immunity is the development of CAR T-cell therapies for patients with aggressive lymphomas, ALL, MM, and CLL, resulting in long-lasting remissions following heavy pretreatment.[28] Developed in the late 1980s and requiring decades of fine-tuning and extensive clinical trials, CAR T cells have the ability to recognize and target tumor cells via their reprogrammed T-cell receptor (TCR) toward the malignant cell and costimulatory molecules engineered into the patient's own T cells. This form of treatment, engaging a combination of gene therapy, cell therapy, and immunotherapy, has led to unprecedented results in patients without robust treatment options for their aggressive malignancies or following relapse after allo-HSCT. The primary limitation of CAR T-cell therapy is the need to identify tumor-specific antigens to avoid the serious toxicity of "on target" but "off tumor" response.

Another potential approach to augment GvT responses following HSCT to prevent or treat relapse is the use of checkpoint inhibitors, currently accomplished by mAb blockade of negative regulatory signals from either cytotoxic T-lymphocyte antigen-4 (CTLA4) or programmed death (PD)-1 proteins on the surface of activated T cells upon interaction with their respective ligands, B7-2 (CD86) and PD-L1 (B7-H1), on APCs and tumor cells (Chapter 80). Checkpoint inhibitors show strong clinical efficacy in the treatment of relapsed hematological malignancies, such as CLL and Hodgkin lymphoma.[29–31]

Adjuvant Therapy With Hematopoietic Stem Cell Transplantation

The relatively higher incidence of relapse after dose-intense therapy (compared to allo-HSCT) leads to the hypothesis that chemoablation therapy with auto-HSCT, curative for some patients, could be viewed as a platform for other approaches effective in eliminating the minimal residual disease (MRD) in patients destined to experience relapse. Additional or yet higher-dose chemotherapy or radiotherapy, unless directly targeted to the tumor, increases the risk of nonhematopoietic toxicity and TRM from causes other than relapse. The correlation of more rapid lymphocyte recovery with a decreased risk of relapse, although likely a reflection of host factors and not direct evidence of a GvT effect after auto-HSCT, supports attempts to use HSCT as a tumor-debulking platform for posttransplantation immunotherapies. Immunotherapies are of interest in this regard and include administration of posttransplantation cytokines; the addition of tumor-specific antibodies, used before and/or after HSCT as an "in vivo purge;" and the development of tumor-specific vaccines, such as with tumor antigen–pulsed dendritic cells (DCs). These therapies are under review and overall clinical benefit remains elusive.

CLINICAL HEMATOPOIETIC STEM CELL TRANSPLANTATION

Sources of Hematopoietic Stem Cells

Bone marrow, PBSCs, or umbilical cord blood (UCB) are all appropriate sources of HSCs for transplantation, with the primary differences being the quantities of HSCs in each product and the quantities and functions of other blood cells comprising most of the cellular content of the HSC product that may influence the immunological outcomes of transplantation. PBSCs have virtually replaced bone marrow as the HSC source for auto-HSCT and are widely used for allo-HSCT. Transplantation of PBSCs achieves more rapid hematological recovery in both the auto- and allo-HSCT settings, resulting in lower complication rates and lower costs of treatment. PBSCs appear to improve overall survival and disease-free survival of patients with advanced hematological malignancies who undergo allo-HSCT, although at the cost of an increased risk of extensive cGvHD.[32] Use of PBSCs results in a decreased CD4 T-helper cell-1 (Th1)/Th2 ratio, which may adversely affect the ability to counteract infections and may favor cGvHD development. It is also interesting to note that PBSC products tend to contain about 10 times more T cells than bone marrow products, yet those T cells are less functional, possibly because of inhibition by granulocytes activated by the granulocyte–colony-stimulating factor (G-CSF) used for HSC mobilization.[33]

UCB is a rich source of HSCs, with the major limitation for clinical use being the small quantity of cells collected, resulting in slower recovery of hematological function and increased risk of failure of sustained engraftment. A very significant advantage is the lower risk of GvHD because of the relative immaturity of the donor immune system, allowing for the use of

HLA-mismatched units without a prohibitory increase in the risk of GvHD. In general, as with other sources of HSCs, outcomes of UCB transplantation reflect patient characteristics, with lower survival probabilities for patients with advanced diseases or poorer performance status at time of transplantation.

KEY CONCEPTS

Selection of Hematopoietic Stem Cell Products for Transplantation

Bone Marrow
- Adequate quantities of cells can be obtained from most patients and donors
- Cytokine administration before harvesting may increase the quantities of hematopoietic stem cells (HSCs) collected and the number and function of accessory cells affecting transplant outcomes
- Lower risk of chronic graft-versus-host disease (GvHD) than peripheral blood stem cells (PBSCs)

Peripheral Blood Stem Cells
- Faster hematological recovery than with bone marrow or umbilical cord blood (UCB)
- Better survival after related-donor hematopoietic stem cell transplantation (HSCT) for patients with advanced malignancy
- Lower tumor cell contamination than with bone marrow collected from autologous patients
- Cytokine administration before harvesting may increase the quantities of HSCs collected and the number and function of accessory cells affecting transplantation outcomes

Umbilical Cord Blood
- Relative immaturity of donor immune system permits multiple antigen-mismatched transplantation
- Transplantation outcomes similar to mismatched unrelated bone marrow transplantation
- Slower hematological recovery than with either PBSC or bone marrow
- Availability of stored units facilitates transplantation for patients with immediate need of treatment

Cell dose is an important predictor of outcome for both auto-HSCT and allo-HSCT, and HSCs comprise a very small portion (generally <1%) of the marrow, PBSC, or UCB product. It is now recognized that successful establishment of donor cell chimerism after allo-HSCT is a complex interplay of pre- and posttransplantation suppression of the host immune system, dose of HSCs and accessory cells (including donor lymphocytes) contained in the graft, and donor HLA compatibility (Fig. 92.3). HLA mismatching, T-cell depletion, and less-intensive pretransplantation conditioning regimens all raise the risk of graft failure. Importantly, the duration of aplasia predicts the incidence of TRM after auto-HSCT or allo-HSCT.[34] Auto-HSCT has a negligible risk of engraftment failure if the viability of HSCs is maintained during processing and storage. The speed of hematological recovery is related to the quantity of HSCs reinfused in an exponential relationship.[35] Increasingly higher $CD34^+$ cell (a surrogate marker of immature HSCs) doses result in greater likelihood of rapid recovery of hematopoietic function. At lower doses, there is considerable heterogeneity in engraftment speed, especially for platelet recovery, with some patients experiencing quick engraftment despite low doses of PBSCs. Products containing ≥2 to 3×10^6 $CD34^+$ cells/kg recipient weight have more consistent rapid granulocyte and platelet engraftment.

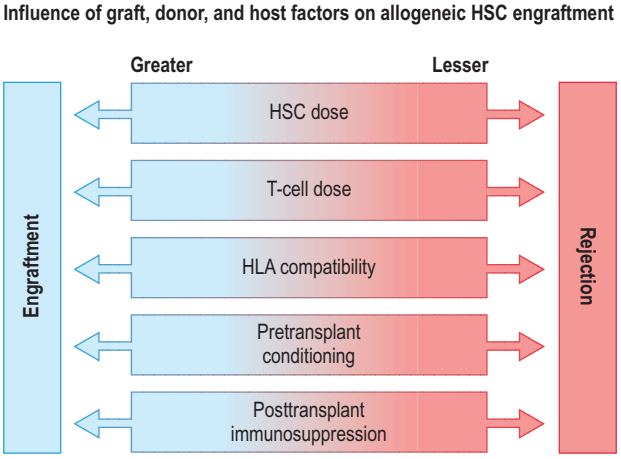


FIG. 92.3 Effect of Donor and Graft Characteristics on Allogeneic Hematopoietic Stem Cell Transplantation. Hematological engraftment requires that the host-versus-graft (HvG) reaction be overwhelmed, either by administration of a more intensive conditioning regimen before transplantation, use of more immunosuppressive medications after transplantation, closer matching of donor with host, and infusion of donor lymphocytes that can affect development of graft-versus-host disease *(GvHD)*. *HLA*, Human leukocyte antigen; *HSC*, hematopoietic stem cell.

Similarly, cell dose is a predictor of outcome for patients undergoing allo-HSCT. The actuarial 5-year TRM, overall survival, and disease-free survival significantly favored patients receiving marrow or PBSC grafts with higher cell doses,[36] but exceeding a $CD34^+$ cell dose of about 10×10^6/kg results in lower survival, likely as a result of higher risk of chronic GvHD.[37]

HSC products are not uniform in their characteristics, and the relative contributions of $CD34^+$ cell dose and the doses of accessory cells, including CD4, CD8, and NK-cell populations, on the outcome of transplantation remain unconfirmed. Whether or not this is clinically significant, the ratio and quantity of cell populations collected will be affected by changes in the cytokines and chemokines, used singly or in combination, to mobilize HSCs into peripheral blood for collection or to alter bone marrow products collected from patients and donors.

Hematological Recovery

HSC engraftment encompasses two concepts: *(i)* recovery of hematopoietic and immunological function and *(ii)* the rate at which this recovery occurs. Delay in or failure of sustained engraftment after myeloablative-conditioning regimen administration greatly increases treatment morbidity and cost. Engraftment failure can occur as a result of inadequate HSC quantity from poor collection or loss in postcollection processing, inadequate host support (stromal cell function) of the infused cells, posttransplantation events or medications, or HvG rejection (see Fig. 92.3). Engraftment failure is a very rare complication of auto-HSCT and is most likely a consequence of poor preservation of HSCs after collection. Secondary graft failure, occurring despite initial engraftment, is a delayed complication rising from infections or induced by other iatrogenic causes. In allo-HSCT, the risk of engraftment failure is proportional to the donor HLA–miHA disparity, occurring more commonly in unrelated donor transplants than in sibling donor transplants, and with HLA-mismatched transplants. Engraftment failure is also

increased by T-cell depletion of marrow inoculum because of the loss of the GvH effect against residual host immune cells.

Chimerism assessment is important in evaluating graft function after allo-HSCT. A fall in PBSC counts could indicate HvG rejection of the graft or early relapse after transplantation or could result from GvHD or viral infection. Documentation of stable persistence of donor T cells (*e.g.*, $CD3^+$ cells) in the recipient's blood will help discriminate between these possibilities. It is also important that sustained lymphoid chimerism be demonstrated if DLI is to be used in the treatment of disease relapse after transplantation. The level of donor–host chimerism after allo-HSCT is best demonstrated through evaluation of single nucleotide tandem repeats (short tandem repeats [STRs]) by using molecular analysis. Obviously, such studies are of no value in assessing engraftment after auto-(syngeneic) HSCT.

Much of the emphasis in transplantation has been on myeloid engraftment because initial patient survival depends on recovery of phagocytes and, to a lesser extent, platelets. Immune reconstitution, and in particular donor T-cell reconstitution, in patients receiving HSCT is often hampered by older recipient age, diminished functional status of the thymus, cytokine milieu at the time of transplantation, and posttransplantation immunosuppressive treatments. The thymus involutes rapidly after childhood, and in the older adult, it is only able to contribute a very small portion to the mature T-cell compartment. The thymic tissue may be damaged as a result of a myeloablative conditioning regimen, or it can also be a target of alloreactive donor T cells mediating GvHD. As a result, restoration of the T-cell compartment in patients is often slow, particularly for CD4 T cells, and may be at suboptimal levels for many months to over a year. This situation, of course, endangers the ability of the patient to stave off opportunistic infections, such as from the herpes family of viruses and fungal pathogens. If donor T cells are provided in the HSC inoculum, some reconstitution of the T-cell repertoire, mostly CD8 T cells, is provided by the mechanism of nonthymic homeostatic expansion, although the level of diversity may be limited. Experimentally, administration of cytokines, such as interleukin-7 (IL-7), after HSCT, can enhance thymic function and help donor T-cell reconstitution. B-cell reconstitution, in contrast, is not that problematic in terms of the regeneration of the immune repertoire, although the ability to actually respond effectively to an infection with antibody production may still depend on the availability of antigen-specific CD4 T cells. Administration of Ig to patients with low IgG levels can prevent some of the infectious complications. Patients receiving HSCT who are conditioned with myeloablative regimens usually attain high levels of donor chimerism in their lymphoid compartment within a few months of transplantation. This often correlates with the ability of alloreactive donor T cells, capable of mediating GvHD, to target residual recipient HSC elements so that the primary source of de novo lymphoid reconstitution will be from the donor. By the same token, high donor chimerism is also associated with a lower incidence of relapse of malignancy.

CONDITIONING REGIMENS

Dose-Intensive and Reduced-Intensity Chemotherapy

The pretransplantation regimen is intended to accomplish two goals: ablate the tumor and achieve adequate immunosuppression to allow donor engraftment. For auto-HSCT, only the dose sensitivity of the tumor being treated need be considered. Lower-dose, nonmyeloablative regimens are not used in auto-HSCT because with such regimens, infusion of HSCs to reconstitute marrow function would not be needed. Total body irradiation (TBI) was initially used for conditioning of transplant recipients. This modality achieves tumor cytotoxicity; treatment of sanctuary sites of disease such as the central nervous system (CNS) and testes; and profound immunosuppression. TBI is usually combined in sequence with chemotherapy agents, such as cyclophosphamide or etoposide. Busulfan-based regimens were developed as alternatives to TBI for patients who had received prior dose-limiting radiotherapy and to avoid the effects of TBI on growth and development in children. A review of several studies that compared the use of busulfan and TBI found no statistically significant difference in survival or disease-free survival for patients with CML or AML.[38] Specific regimens are commonly used in the treatment of certain malignancies, such as dose-intensive melphalan in the treatment of MM and carmustine (BCNU)–containing regimens used in the treatment of lymphoma.

CLINICAL PEARLS

Conditioning Regimens for Hematopoietic Stem Cell Transplantation

Autologous Hematopoietic Stem Cell Transplantation
- Allows dose-intensive therapy
- Allows use of marrow-toxic agents
- Regimens designed for optimal tumor cytotoxicity

Allogeneic Hematopoietic Stem Cell Transplantation
- Must achieve adequate patient immunosuppression (reduce host-versus-graft [HvG] reaction) to achieve engraftment
- Allows but does not require dose-intensive therapy
- Allows but does not require marrow-toxic agents
- Need not be tumor specific

The myeloablative conditioning regimens currently used have been tested in dose-escalation studies to achieve the maximal tolerated doses in otherwise healthy patients. Nonmarrow toxicities, such as pneumonitis, mucositis, and hepatic venoocclusive disease, limit further dose escalation of standard TBI- or chemotherapy-based regimens. New approaches include the addition of targeted therapies to the conditioning regimen, such as tumor-directed mAbs or radioimmunoconjugates that will not increase the toxicity to other organs. Tandem transplantation with the combination of a dose-intensive regimen with auto-HSCT followed, after recovery from the immediate regimen-related toxicities, by allo-HSCT using a reduced-intensity regimen is a novel approach to combine the benefits of each transplantation modality.

Hematological malignancies are predominantly diseases of older adults. The potent GvT effect that is observed after allo-HSCT allows allograft recipients to be treated with lower-dose nonmyeloablative regimens with the immunosuppressive properties of the regimen to reduce HvG reactions and facilitate engraftment becoming more important than direct cancer cytotoxicity. The primary requirement in developing a reduced-intensity regimen is the need to achieve adequate immunosuppression to permit the development of hematopoietic chimerism, which became feasible with the development of the purine analogue family of drugs. A variety of regimens are available, including combinations of fludarabine with melphalan and fludarabine with busulfan. Among the least toxic are regimens that involve a single fraction of TBI, based on the work by Storb and colleagues,[39] who proposed that the HvG

reaction leading to HSC rejection and the GvH reaction could both be modified by an appropriate immunosuppressive regimen administered after transplantation, allowing a reduction in the intensity of the pretransplantation conditioning regimen.[39,40]

HEMATOPOIETIC STEM CELL TRANSPLANTATION FOR INDIVIDUAL DISEASES

The treatment of diseases by auto- or allo-HSCT continues to evolve as the understanding of the biology of these diseases becomes more clearly understood. Improvements in nontransplantation treatments available to patients, such as the development of targeted tyrosine kinase (TK) inhibitors, which are very effective in the initial treatment of CML, will reduce the numbers of patients requiring HSCT in the management of their disease. Organizations such as American Society of Blood and Marrow Transplantation have published guidelines for treatment and reviews of the efficacy of treatment, identifying areas requiring additional research.[41,42]

Acute Myelogenous Leukemia

The primary clinical questions in HSCT in the treatment of AML concern patient selection and timing of treatment. Most patients with AML will achieve remission with initial chemotherapy, but even with appropriate postremission consolidation, the majority of patients (≈65%) will relapse within 1 to 2 years. Older age, the presence of defined cytogenetic abnormalities, inability to achieve a minimal-residual disease negative CR with the initial course of therapy, and history of a preceding marrow disorder or receipt of prior chemotherapy ("secondary AML") are predictors for failure of nontransplantation therapy. Patients with these adverse risk factors may be offered HSCT in first remission in place of nontransplantation consolidation chemotherapy. Numerous studies of patients entering their first remission compared standard consolidation therapy with dose intensification with auto- or allo-HSCT. In general, auto-HSCT was not shown to be more effective than nontransplantation consolidation chemotherapy, whereas allo-HSCT had the lowest risk of relapse. Allo-HSCT in first remission is particularly beneficial for patients with adverse risk cytogenetics or leukemia that arises from prior chemotherapy or other marrow diseases, achieving a 50% to 70% disease-free survival (DFS). Features predictive of posttransplant disease relapse include adverse-risk cytogenetic mutations and presence of MRD at time of transplantation. Prophylactic use of DLIs or posttransplant maintenance with targeted agents may reduce the risk of relapse of patients with adverse-risk features or who are found to be developing disease recurrence early after transplantation.

Myelodysplastic Syndromes

Myelodysplastic syndrome (MDS) comprises a heterogeneous group of clonal hematological disorders characterized by expansion of abnormal HSCs engendering variable degrees of cytopenia and frequent evolution to AML. Currently, allo-HSCT is the only modality of treatment that can achieve long-term control of disease; auto-HSCT is not feasible because of the inability to collect normal HSCs from these patients. The best results are seen in patients with earlier-stage disease, although for patients with earlier-stage MDS, a "watchful waiting" approach, with transplantation performed at the time of disease progression, may be appropriate.

Chronic Myelogenous Leukemia

Allo-HSCT is an appropriate treatment for CML, with long-term survival rates >80% for younger patients undergoing related donor transplantation within the first year after diagnosis. However, inhibitors of the TK encoded by the Philadelphia chromosome have relegated transplantation to the treatment of patients with advanced disease or the rare patient with CML not responsive to that targeted therapy. The effect of prior therapy with a TK inhibitor on the outcome of transplantation is not known, although patients who fail multiple TKIs before coming to transplantation will experience a lower survival probability. CML is highly responsive to the immunological GvT effect, and many patients in relapse after transplantation can be salvaged by the administration of DLI. CML is a stem cell disease, and the collection of normal HSCs is unlikely for most patients, precluding auto-HSCT.

Myeloproliferative Diseases

MPDs are clonal stem cell diseases and include CML (discussed above), polycythemia vera, primary myelofibrosis, essential thrombocytosis, chronic myelomonocytic leukemia, and MPD not otherwise characterized. These disorders display overlapping clinical features and may also show features more characteristic of MDS, but they exhibit different natural histories and different therapeutic requirements. Allo-HSCT is effective in ablating the abnormal clone in these disorders. Staging systems that predict the course of the disease are used to determine the timing of transplantation. The timing of treatment is important because of the frequently long natural history of these diseases.

Acute Lymphoblastic Leukemia

In contrast to treatment of the pediatric patient with ALL (Chapter 77), only very few adults with ALL are cured with nontransplantation induction and consolidation regimens. As with AML, adverse-risk features include certain cytogenetic changes and the inability to achieve minimal residual disease–negative remission with initial induction chemotherapy. Allo-HSCT in first remission is clearly effective in the management of patients with defined adverse-risk cytogenetics, such as translocation involving chromosomes 4 and 11, or 9 and 22 (Philadelphia chromosome). Recent randomized studies also demonstrate a survival advantage for patients who have intermediate risk of disease and are undergoing allo-HSCT, especially since current salvage regimens for patients in relapse have a very low likelihood of achieving a second remission. There is little evidence to support the effectiveness of auto-HSCT despite its theoretical potential.

Chronic Lymphocytic Leukemia

Many patients and physicians are reluctant to support aggressive treatment of CLL with allo-HSCT because of the frequently indolent nature and long natural history of this disease, as well as the advanced age of most patients at time of diagnosis. Yet disease progression is inevitable, and the aggressiveness of this disease can be predicted by various features, including chromosome 17p and 11q deletions and detection of ZAP-70 mutation and CD38 expression. Therefore HSCT should be considered a treatment choice, especially for younger patients or those with adverse-risk disease features.

The exquisite sensitivity of the CLL cells to the GvT effect allows for the use of reduced-intensity regimens with lower risks of regimen-related mortality in these, generally older,

patients. Durable control of disease can be achieved for about 50% of patients. The extensive infiltration of bone marrow by malignant lymphocytes and the lack of the GvT effect preclude auto-HSCT. In the few clinical studies that have been reported, high CR responses have been obtained with auto-HSCT, but relapses were the frequent cause of treatment failure.

Multiple Myeloma

MM is a malignancy of plasma cells (Chapter 79), with a median patient age of 65 years at time of diagnosis. In MM, dose-intensive melphalan with auto-HSCT can achieve about a 40% complete response rate. Tandem auto-HSCT with two cycles of dose-intensive therapy at 2- to 6-month intervals can improve progressive free survival and overall survival, particularly for patients with adverse-risk cytogenetic features such as 17p deletion. Ultimately, most patients experience relapse. Although auto-HSCT is not considered curative for this disease, the median duration of time without treatment is 2 to 3 years, and ≈15% of patients may not need treatment for progressive disease for 5 years or longer. Allo-HSCT with use of a reduced-intensity conditioning regimen decreases the risk of TRM to less than 10%, but patients with advanced disease are unlikely to respond to this treatment. The lower risk of complications with nonmyeloablative conditioning regimens has led to studies of tandem auto-HSCT, using an intensive chemotherapy regimen to achieve tumor debulking, followed 2 to 6 months later by allo-HSCT using a reduced-intensity regimen to achieve the GvT effect. This approach to the treatment of MM was studied in large multicenter trials, with a strong GvT effect detected. However, the benefit of the GvT effect was offset by the increased regimen-related toxicity compared to tandem autologous transplantation.

Non-Hodgkin Lymphoma

The lymphomas (Chapter 78) represent a diverse group of malignant diseases of B and T lymphocytes, comprising some of the slowest- to the fastest-growing human malignancies, with a range of curability achieved by nontransplantation therapies. As a group, the lymphomas exhibit a strong dose–response relationship to chemotherapy or radiotherapy, and the benefit of dose-intensive treatment with auto-HSCT has been well established. Auto-HSCT avoids the morbidity of allo-HSCT and is the preferred approach for the majority of patients, with some prominent exceptions discussed below. Popular chemotherapy regimens include cyclophosphamide, BCNU, and etoposide (CBV) or BEAM (BCNU, etoposide, cytarabine, and melphalan). However, no single chemotherapeutic or radiation-based regimen has emerged as a superior treatment.

Low-Grade Non-Hodgkin Lymphoma

Low-grade NHL, in general, exhibits a variable and prolonged natural course, with many patients not requiring treatment until symptoms or organ toxicity appear. Therefore, most of the experience with HSCT has been in patients after relapse rather than at the time of initial diagnosis. Few randomized studies comparing auto-HSCT with nontransplantation therapies for patients have been reported. A number of phase II and registry data have been published for auto-HSCT, and although response rates are high, a continuing pattern of relapse has been observed. In contrast, patients who undergo allo-HSCT experience a higher probability of TRM but a lower risk of relapse after transplantation. Indolent B-cell NHL, similar to other low-grade diseases, such as CLL, appears to be very sensitive to the GvT effect of allo-HSCT, and using reduced-intensity regimens will result in lower transplantation-related complications. The difference in relapse rates between auto-HSCT and allo-HSCT could result from the possible reintroduction of lymphoma cells in the HSC product, but is more likely from the lack of a GvT effect of auto-HSCT.

Aggressive Non-Hodgkin Lymphoma

Auto-HSCT is the standard of care for patients with diffuse large B-cell NHL in first "chemotherapy-sensitive" relapse. The success of this therapy reflects the extent and the responsiveness of the disease to chemotherapy at the time of transplantation, with relapse the major cause of treatment failure. The Parma randomized trial for high-dose therapy followed by auto-HSCT reported a significantly higher event-free survival for patients treated with auto-HSCT than for the group receiving standard-dose treatment (46% vs. 12%). Furthermore, it is notable that no patients assigned to the conventional-dose salvage therapy could be rescued at the time of the second relapse with delayed transplantation. The efficacy of auto-HSCT is diminished for patients previously treated with rituximab, an anti–B-cell chimeric antibody, and who relapse within 12 months of initial therapy.[43] Other circumstances, in which HSCT may be indicated, include HSCT for patients who respond slowly to initial therapy, have high-risk disease, are resistant to initial therapy, or are in relapse with chemotherapy-insensitive disease.[42] Results for patients with truly chemotherapy-refractory disease are, however, poor, and such patients should be considered for allo-HSCT or treatment with autologous CAR T cells (Chapter 81).

Mantle cell lymphoma is known for its unremitting clinical course when treated conventionally and has proven relatively resistant to dose-intensive treatment, especially when used for management of relapsed disease. Mantle cell NHL appears to be very sensitive to the GvT effect of allo-HSCT, allowing treatment with reduced-intensity regimens.

Burkitt lymphoma, Burkitt-like lymphomas, and lymphoblastic lymphomas are high-grade NHLs associated with relatively poor long-term survival rates. The role of both auto-HSCT and allo-HSCT in these disorders remains unclear. There are no convincing risk models that identify clear indications for transplantation as part of the initial therapy. The lack of an obvious GvT effect in these disorders suggests that auto-HSCT can provide a reasonable treatment. Transplantation registry data suggest that disease status at the time of transplantation is the most important predictor of outcome in patients with high-grade disease.[44]

T-cell lymphomas are less common than BCLs. There is no well-defined management strategy for these disorders, and treatment is based on the disease stage and immunopathological grade. Disease control with T cell–type treatment regimens, compared with the more common B cell–type treatment regimens, is lower, although the comparison may not be appropriate, as T-cell diseases present at more advanced stages compared with B-cell diseases. The GvT effect may be more evident in T-cell NHL, particularly for virus-associated NHL, and allo-HSCT is the preferred treatment for some patients. Prospective comparative studies to guide treatment decisions have not been reported.

Hodgkin Lymphoma

Many patients with Hodgkin lymphoma will achieve durable remissions with nontransplantation chemotherapy and/or radiation therapy, and algorithms for staging and treatment of this disease are well defined. Dose-intensive therapy with auto-HSCT is available to those patients who do not achieve a remission or who relapse

after initial therapy. The outcome for patients whose remission lasted less than 1 year is dismal, with standard-dose second-line treatments, and approximately 40% to 50% of patients with Hodgkin lymphoma who suffer a relapse within 1 year will achieve durable remissions after auto-HSCT. This approach can also overcome drug resistance and lead to an overall survival rate of 34% to 50% for patients with chemotherapy-refractory disease. Outcomes are much improved for patients who achieve "positron emission tomography (PET) negativity" before undergoing auto-HSCT.

Allo-HSCT is not the first choice for the treatment of relapsed Hodgkin lymphoma because of the higher transplantation-related complications, despite the evidence of an effective GvT effect. Allo-HSCT is offered to patients who suffer relapse after auto-HSCT or who are not candidates for auto-HSCT because of aggressive, chemotherapy-refractory disease. With the front-line use of checkpoint inhibitors such as nivolumab for treatment of newly diagnosed and relapsed Hodgkin lymphoma, the risk of GvHD following allogeneic HSCT is considerable. This risk is mitigated by an appropriate "wash-out" period for the drug or potentially other strategies such as use of posttransplantation cyclophosphamide for GvHD prevention.[45]

Solid Tumors

Skin and colonic mucosa are primary targets of both acute and cGvHD, and this would suggest that allo-HSCT would be effective therapy in the treatment of cancers of these organs. Yet allo-HSCT is not effective in the control of these cancers, illustrating the discrimination between target antigens of normal tissues of GvH, such as the colonic crypt cells, compared with antigens expressed by tumors derived from these tissues targeted by GvT. With the exception of allo-HSCT in the treatment of renal cell cancer, in which a GvT effect was seen to be clinically evident in some but not all studies, transplantation in the treatment of solid tumors, such as neuroblastoma, Wilms tumor, and germ cell tumors, is limited to auto-HSCT with one or more cycles of dose-intensive chemotherapy.

FUTURE DIRECTIONS

Advances in HLA typing, chemotherapy conditioning regimens, supportive care, and our understanding of the biology of the HvG and GvH reactions have greatly reduced the toxicity of both auto-HSCT and allo-HSCT. Although the diseases treatable with HSCT are those characterized by dose sensitivity, currently available dose-intensive conditioning regimens have been pushed to maximal tolerable doses. Newer approaches to the treatment of these malignancies will likely include the addition of therapies that target the malignancy, such as radioimmunoconjugates that will add minimal toxicity to other organs. The availability of reduced-intensity regimens permits tandem transplantation, using dose-intensive therapy with auto-HSCT to achieve maximal tumor-cell debulking, followed, after recovery from transplantation-related toxicities, by allo-HSCT, using a reduced-intensity regimen to achieve an immunological GvT effect. Advancements in the field of haploidentical HSCT have facilitated immediate availability of donors for patients with high-risk malignancies when selection of unrelated donors is limited by ethnicity and time constraints. New testing techniques, such as detection of tumor-derived cell-free DNA, may be predictive of relapse, facilitating treatment of this complication of transplantation. Finally, the ability to use HSCT as a platform for protein- or cellular-based vaccination strategies is the subject of considerable interest.

ON THE HORIZON

Allogeneic Hematopoietic Stem Cell Transplantation

Continuing improvement in donor selection algorithms
Strategies to maximize graft-versus-tumor (GvT) effect
Refining haploidentical hematopoietic stem cell transplantation (HSCT) techniques for different diseases
Use of killer cell immunoglobulin-like testing for donor selection
Combination conditioning regimens
Developing regimens with lower toxicities
Developing testing predictive of early relapse
Posttransplantation maintenance/treatment to reduce risk of relapse
Immune effector cell therapies
Platform for other treatments; chimeric antigen receptor (CAR) T-cell therapy, checkpoint inhibitor blockade

Autologous Hematopoietic Stem Cell Transplantation

Tumor-specific conditioning regimens
Improve ratio of toxicity to tumor-cell cytotoxicity
Developing testing predictive of early relapse
Posttransplantation vaccination strategies
Tandem autologous transplantation with autologous CAR T-cell therapy
Checkpoint inhibitor blockade in clinical trials for high-risk malignancies

REFERENCES

1. Cooper JP, Soror BE, Granot N, et al. Allogeneic hematopoietic cell transplantation with non-myeloablative conditioning for patients with hematologic malignancies: Improved outcomes over two decades. *Haematologica*. 2021; 106;1599–1607.
2. Tiercy J-M. How to select the best available related or unrelated donor of hematopoietic stem cells? *Haematologica*. 2016;101:680–687.
3. Confer DL, Abress LK, Navarro W, et al. Selection of adult unrelated hematopoietic stem cell donors: beyond HLA. *Biol Blood Marrow Transplant*. 2010;16:S8–S11.
4. Nakamura R, La Rosa C, Tsai W, et al. Ex vivo detection of CD8 T cells specific for H-Y minor histocompatibility antigens in allogeneic hematopoietic stem cell transplant recipients. *Transpl Immunol*. 2014;30(4): 128–135.
5. Socie G, Blazar BR. Acute graft-versus-host disease: from the bench to the bedside. *Blood*. 2009;114:4327–4336.
6. Bleakley M, Riddel SR. Molecules and mechanisms of the graft-versus-leukaemia effect. *Nat Rev Cancer*. 2004;4:371–380.
7. Glucksberg H, Storb R, Fefer A, et al. Clinical manifestations of graft-versus-host disease in human recipients of marrow from HLA-matched sibling donors. *Transplantation*. 1974;18:295–304.
8. Mathewson ND, Jenq R, Mathew AV, et al. Gut microbiome-derived metabolites modulate intestinal epithelial cell damage and mitigate graft-versus-host disease. *Nat Immunol*. 2016;17:505–513.
9. Schwartz R, Dameshek W. Drug-induced immunological tolerance. *Nature*. 1959;183:1682–1683.
10. Berenbaum MC, Brown IN. Prolongation of homograft survival in mice with single doses of cyclophosphamide. *Nature*. 1963;200:84.
11. Robinson TM, O'Donnell PV, Fuchs EJ, et al. Haploidentical bone marrow and stem cell transplantation: experience with post-transplantation cyclophosphamide. *Semin Hematol*. 2016;53:90–97.
12. Martinez C, Gayoso J, Canals C, et al. Post-transplantation cyclophosphamide-based haploidentical transplantation as alternative to matched sibling or unrelated donor transplantation for Hodgkin lymphoma: a registry study of the Lymphoma Working Party of the European Society for Blood and Marrow Transplantation. *J Clin Oncol*. 2017;35:3425–3432.
13. Zeiser R, Bubnoff NV, Butler J, et al. Ruxolitinib for glucocorticoid-refractory acute graft-versus-host disease. *N Engl J Med*. 2020;382:1800–1810.
14. Yeager AM, Vogelsang GB, Jones RJ, et al. Induction of cutaneous graft-versus-host disease by administration of cyclosporine to patients undergoing autologous bone marrow transplantation for acute myeloid leukemia. *Blood*. 1992;79:3031–3035.

15. Keung YK, Beaty MW, Pettenati M, et al. Possible role of engraftment syndrome and autologous graft-versus-host disease in myelodysplastic syndrome after autologous stem cell transplantations: retrospective analysis and review of the literature. *Clin Lymphoma Myeloma Leuk*. 2010;10:129–133.
16. Pidala J, Kim J, Anasetti C, et al. The global severity of chronic graft-versus-host disease, determined by National Institutes of Health consensus criteria, is associated with overall survival and non-relapse mortality. *Haematologica*. 2011;96:1678–1684.
17. Weiden PL, Flournoy N, Thomas ED, et al. Antileukemia effect of graft-versus-host disease in human recipients of allogeneic-marrow grafts. *N Engl J Med*. 1979;300:1068–1073.
18. van den Brink MRM, Porter DL, Giralt S, et al. Relapse after allogeneic hematopoietic cell therapy. *Biol Blood Marrow Transplant*. 2010;16:S138–S145.
19. Davids MS, Kim HT, Costello C, et al. A multicenter phase 1 study of nivolumab for relapsed hematologic malignancies after allogeneic transplantation. *Blood*. 2020;135:2182–2191.
20. Jacoby E, Yang Y, Qin H, et al. Murine allogeneic CD19 CAR T cells harbor potent antileukemic activity but have the potential to mediate lethal GVHD. *Blood*. 2016;127:1361–1370.
21. Schmidt S, Liu Y, Hu ZH, et al. The role of donor lymphocyte infusion (DLI) in post-hematopoietic cell transplant (HCT) relapse for chronic myeloid leukemia (CML) in the tyrosine kinase era. *Biol Blood Marrow Transplant*. 2020;1137–1143.
22. Steinmann J, Bertz H, Wäsch R, et al. 5-Azacitadine and DLI can induce long-term remissions in AML patients relapsed after allograft. *Bone Marrow Transplant*. 2015;50:690–695.
23. Guglielmi C, Arcese W, Dazzi F, et al. Donor lymphocyte infusion for relapsed chronic myelogenous leukemia: prognostic relevance of the initial cell dose. *Blood*. 2002;100:397–405.
24. Ciceri F, Bonini C, Stanghellini MT, et al. Infusion of suicide-gene-engineered donor lymphocytes after family haploidentical haematopoietic stem-cell transplantation for leukaemia (the TK007 trial): a non-randomised phase I-II study. *Lancet Oncol*. 2009;10:489–500.
25. Anderson BE, McNiff J, Yan J, et al. Memory $CD4^+$ T cells do not induce graft-versus-host disease. *J Clin Invest*. 2003;112:101–108.
26. Chen BJ, Deoliveira D, Cui X, et al. Inability of memory T cells to induce graft-versus-host disease is a result of an abortive alloresponse. *Blood*. 2007;109:3115–3123.
27. Rowley SD, Donato ML, Bhattacharyya P. Red blood cell incompatible allogeneic hematopoietic progenitor cell transplantation. *Bone Marrow Transplant*. 2011;46:1167–1185.
28. Holstein S, Lunning MA. CAR T-cell therapy in hematologic malignancies: a voyage in progress. *Clin Pharmacol Ther*. 2020;107:112–122.
29. Bachireddy P, Burkhardt UE, Rajasagi M. Haematological malignancies: at the forefront of immunotherapeutic innovation. *Nat Rev Cancer*. 2015;15:201–215.
30. Ansell SM, Lesokhin AM, Borrello I, et al. PD-1 Blockade with nivolumab in relapsed or refractory Hodgkin's lymphoma. *N Engl J Med*. 2015;372:311–319.
31. Bashey A, Medina B, Corringham S, et al. CTLA4 blockade with ipilimumab to treat relapse of malignancy after allogeneic hematopoietic cell transplantation. *Blood*. 2009;113:1581–1588.
32. Anasetti C, Logan BR, Lee SJ, et al. Peripheral-blood stem cells versus bone marrow from unrelated donors. *N Engl J Med*. 2012;367:1487–1496.
33. Vasconcelos ZF, Santos BM, Costa ES, et al. T-lymphocyte function from peripheral blood stem-cell donors is inhibited by activated granulocytes. *Cytotherapy*. 2003;5:336–345.
34. Offner F, Schoch G, Fisher LD, et al. Mortality hazard functions as related to neutropenia at different times after marrow transplantation. *Blood*. 1998;88:4058–4062.
35. Rowley SD, Zuehlsdorf M, Braine HG, et al. CFU-GM content of bone marrow graft correlates with time to hematologic reconstitution following autologous bone marrow transplantation with 4-hydroperoxycyclophosphamide purged bone marrow. *Blood*. 1987;70:271–275.
36. Ringden O, Barrett AJ, Zhang MJ, et al. Decreased treatment failure in recipients of HLA-identical bone marrow or peripheral blood stem cell transplants with high CD34 cell doses. *Br J Haematol*. 2003;121:874–885.
37. Heimfeld S. HLA identical stem cell transplant: is there an optimal CD34 cell dose? *Bone Marrow Transplant*. 2003;31:839–845.
38. Socie G, Clift RA, Blaise D, et al. Busulfan plus cyclophosphamide compared with total-body irradiation plus cyclophosphamide before marrow transplantation for myeloid leukemia: long-term follow-up of 4 randomized studies. *Blood*. 2001;98:3569–3574.
39. Storb R, Yu C, Wagner JL, et al. Stable mixed hematopoietic chimerism in DLA-identical littermate dogs given sublethal total body irradiation before and pharmacological immunosuppression after marrow transplantation. *Blood*. 1997;89:3048–3054.
40. McSweeney PA, Niederwieser D, Shizuru JA, et al. Hematopoietic cell transplantation in older patients with hematologic malignancies: replacing high-dose cytotoxic therapy with graft-versus-tumor effects. *Blood*. 2001;97:3390–3400.
41. Oliansky DM, Gordon LI, King J, et al. The role of cytotoxic therapy with hematopoietic stem cell transplantation in the treatment of follicular lymphoma: an evidence-based review. *Biol Blood Marrow Transplant*. 2010;16:443–468.
42. Oliansky DM, Antin JH, Bennett JM, et al. The role of cytotoxic therapy with hematopoietic stem cell transplantation in the therapy of myelodysplastic syndromes: an evidence-based review. *Biol Blood Marrow Transplant*. 2009;15:137–172.
43. Gisselbrecht C, Glass B, Mounier N, et al. Salvage regimens with autologous transplantation for relapsed large B-cell lymphoma in the rituximab era. *J Clin Oncol*. 2010;28:4184–4190.
44. Sweetenham JW, Pearce R, Taghipour G, et al. Adult Burkitt's and Burkitt-like non-Hodgkin's lymphoma—outcome for patients treated with high-dose therapy and autologous stem-cell transplantation in first remission or at relapse: results from the European Group for Blood and Marrow Transplantation. *J Clin Oncol*. 1996;14:2465–2472.
45. Schoch LK, Cooke KR, Wagner-Johnston ND, et al. Immune check point inhibitors as a bridge to allogeneic transplantation with posttransplant cyclophosphamide. *Blood Adv*. 2018;2(17):2226–2229. https://doi.org/10.1182/bloodadvances.2018019208.

Flow Cytometry

Cristiane J. Nunes-Santos, Sergio D. Rosenzweig, and Thomas A. Fleisher

Flow cytometry has become a standard laboratory tool in evaluating hematopoietic cells, including identifying leukocyte populations and subpopulations, a method referred to as immunophenotyping. The clinical application of this technology has been facilitated by the development of instruments and data analysis systems suitable for routine use in diagnostic laboratories. In addition, the expanded range of monoclonal antibodies specific for lymphocyte (and other hematopoietic cells) surface antigens directly conjugated to a number of different fluorescent indicators (fluorochromes) provide an extensive panel of reagents that facilitate multicolor (polychromatic) studies.

The clinical needs that pushed this technology date back to the emergence of absolute CD4 T-cell counts as a critical measure for disease assessment and follow-up in managing patients infected with the human immunodeficiency virus (HIV). This was followed by the routine application of cell characterization by flow cytometry to evaluate hematologic malignancies, and more recently, in the study of immunodeficiency and immune dysregulation disorders.

Recent advances in instrumentation and fluorochrome chemistry allow for routine polychromatic flow cytometry studies, with concomitant assessment of cell surface markers and intracellular parameters (intracellular proteins, phosphorylated [phospho] proteins, and cytokines), as well as identification of changes linked to cellular activation and apoptosis. Intracellular flow cytometry also can be applied to evaluate cell cycle status (i.e., G_0-G_1, S, G_2-M) based on DNA staining, useful in evaluating tumor cells and assessing the *in vitro* lymphocyte response to various stimuli. Additionally, evaluation of lymphocyte proliferation can be performed with cell tracking dyes that allow quantitation of the rounds of cell division, and assessment of immune cell-mediated cytotoxicity can be performed. Finally, characterization of antigen-specific T cells following immunization or associated with normal and/or abnormal immune responses in association with disease states can be accomplished using multimer technology and intracellular cytokine detection following *in vitro* antigen exposure.

This chapter focuses on basic concepts of flow cytometry, including instrument characteristics, data management, lymphocyte gating, and directed use of test reagents. In addition, a brief overview of intracellular protein detection, cell activation and cell-mediated cytotoxicity studies, cell cycle analysis, apoptosis detection, and multimer technology is provided, focusing on the appropriate clinical application of these approaches and their limitations.

INSTRUMENTATION

The basic components of a flow cytometer, as shown in Fig. 93.1, include the illumination source, optical bench, fluidic system, electronics, and computer.[1] Briefly, stained cells flow into single-file due to the fluidic system and are interrogated by one or more light sources to generate light signals directed by the optical system to the photodetectors that convert light into electronic signals for storage and subsequent analysis.

The fluidic system consists of isotonic sheath fluid that moves the sample stream containing the cells for analysis. This process is accomplished by injecting the cell sample into flowing sheath fluid, resulting in a hydrodynamically focused single-file flow of cells (or other particles) that move through the analysis point while maintaining the cell stream in a constant, central location.[2] The centrally focused cell stream ensures that the illumination of all cells is virtually equivalent. Thus, the difference in magnitude of the emission signal(s) generated from each cell reflects biologic differences between the cells (rather than reflecting the variation produced by non-focused cells). Hydrodynamic focusing has the additional advantage of producing little or no change in cell shape, although it may affect cell orientation. The consistency in maintaining cell shape facilitates distinguishing "architectural" differences between specific leukocyte types (see Gating section).[3] However, this method limits flow rates to 60 to 100 µL/min, which can lead to long acquisition times for detecting very rare events. To overcome this problem, recently introduced flow cytometry instruments utilize acoustic focusing to align cells through the use of sound waves that allows sample flow rates of up to 1000 µL/min, without impacting signal quality.[4]

Illumination in standard clinical instruments is generated by two or more lasers, each of which provides a specific monochromatic light source (e.g., a sapphire laser generates a 488 nm wavelength [blue] beam). Modern lasers are small and available in multiple wavelengths, including ultraviolet (350 nm), violet (405 nm), blue (488 nm), green (532 nm), yellow (560 nm), orange (610 nm), and red (633 nm), permitting the simultaneous use of multiple fluorochromes having different excitation requirements.[5] The point where the light illuminates the cell in analytical instruments occurs within a flow cell, while in cell sorters, the beam intersects cells flowing as a stream in the air.

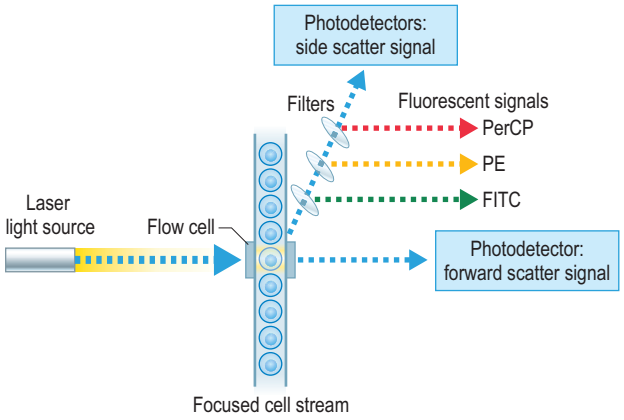


FIG. 93.1 Simplified design of a flow cytometer with one illumination source *(laser)* set up to collect five parameters. These include the two non-fluorescent parameters *(blue light)* forward, and side scatter, as well as three fluorescent parameters, green (FITC), orange (PE), and red (PerCP) light.

The optical bench contains lenses that shape and focus the illumination beam to ensure consistent excitation energy at the analysis point.

The illumination of a cell generates both non-fluorescent and fluorescent signals collected and measured by optically coupling the signal to a detection system consisting of filters linked to photodetectors. The two non-fluorescent signals, forward- and side-scatter, are measured at the same wavelength as the excitation signal (e.g., 488 nm from a blue light source; see Gating section), whereas those for the fluorescence channels utilize specific filters that allow passage of light with wavelengths specific to each fluorochrome (e.g., green, orange or red; see Fluorochrome section). Configurations with increased numbers of photodetectors allow for evaluating an expanded number of colors (parameters).[6]

The internal electronics in the flow cytometer provide the system for converting analog light signals (photoelectrons) into digital signals for acquisition and storage in a computer. The intensity of these converted signals is measured on a relative scale that is generally set in either 256 or 1024 equal increments (channels) for display and analysis. Several specialized analysis programs provide graphic displays of single-parameter histograms or two-color displays (see Data display section below). Most programs enable the operator to evaluate the number and percentage of events, mean and/or median channel fluorescence, and selected statistical measures for each identified cell, and these can be aggregated into specific populations and/or subpopulations of cells. Thus, a flow cytometer provides a platform having the capacity to assess multiple pieces of discrete information (parameters) generated from each individual cell contained within a large number of cells in the sample, and these are typically accrued at rates of 1000–2000 (or more) cells per second.

FLUORESCENCE REAGENTS

Standard monoclonal antibody reagents for clinical use are typically directly conjugated to a fluorochrome, a dye that absorbs and emits light of different wavelengths based on the energy lost during the return of excited electrons to their ground state following illumination by light of a specific wavelength. Thus, the emitted light has a longer wavelength (lower energy) than the wavelength of the excitation beam. The number of commercially available fluorochromes has increased dramatically, and dye conjugates have enhanced the capacity to perform extended polychromatic studies using instruments with three or more lasers.[5] Fluorochromes in clinical immunophenotyping include the organic dyes fluorescein isothiocyanate (FITC), phycoerythrin (PE), peridin chlorophyll protein (PerCP), and allophycocyanin (APC). Conjugations of PE and APC to cyanines (Cy5, Cy5.5, and Cy7) and Alexa Fluor dyes produce tandem dyes with additional emission spectra based on energy transfer. This allows the simultaneous evaluation of 6 to 8 or more colors in most current clinical instruments with two or three lasers.

One recent advance in the field was the development of a new class of inorganic fluorescent semiconductor nanocrystals named quantum dots (QDs).[7] These particles have broad excitation spectra (525 to 800 nm) and sharp, discrete emission spectra that varies depending on their core size. This means that QDs of different sizes (and consequently of different colors) can be excited by the same laser source, allowing simpler multiplexing. In addition, QDs have high quantum yield, high molar extinction coefficients, and extraordinary resistance to photo- and chemical degradation. These qualities make them perfectly suitable for use in biological studies, including intracellular *in vivo* imaging and fluorescence resonance energy transfer (FRET) analysis.[7]

Another advancement involves the more recently developed polymer dyes that not only provide increased sensitivity and precision but also have the potential to increase the number of colors available in polychromatic studies. Polymer dyes harvest light and can be used alone or in a tandem dye arrangement. Further advantages of these include their exceptionally bright emission as well as being photostable and resisting quenching.[8]

Additional dyes are available for functional studies and include calcium-sensitive dyes (e.g., fluo-3), glutathione-sensitive dyes (e.g., monochlorbimane), and H_2O_2-responsive dyes (e.g., dihydrorhodamine 123).[9] Assessment of DNA content can be performed with dyes that intercalate double-stranded DNA and RNA, including propidium iodide and ethidium bromide. In addition, there are ultraviolet-excited dyes that are highly specific for DNA, including Hoechst 33258 and 4,6-diamidino-2-phenylindole (DAPI); acridine orange is used for simultaneous staining of DNA/RNA.[10]

DATA ANALYSIS

Gating

KEY CONCEPTS

Gating

- Method for defining cell population of interest.
- Typically performed using forward and side scatter and lineage specific antibodies.

The proper assessment of specific cell types within a mixture requires initial identification of lineage-specific cells, an approach referred to as gating. In practical terms, immunophenotyping focused on lymphocytes requires minimizing the non-lymphocytes in the evaluation by generating a lymphocyte gate. The standard clinical sample is anticoagulated whole blood, and lymphocyte gating requires eliminating the great

majority of non-lymphocytes from the collected data such that the expression of a percentage for a specific lymphocyte subpopulation is accurate. Without gating, the data can also be negatively impacted by the co-expression of surface antigens on different cell lineages (e.g., CD4 is found on lymphocytes and monocytes at differing density). In addition, nonspecific binding of monoclonal reagents through Fcγ receptors and the level of cytophilic human immunoglobulin varies between cell types, making appropriate gating crucial to generate valid data. Gating is also used to focus the evaluation on other hematopoietic cells: monocytes, granulocytes, eosinophils, erythrocytes, and platelets.

Initial gating of whole blood typically involves using the two non-fluorescent parameters, forward scatter (FSC), and side scatter (SSC) (Fig. 93.2A).[3] FSC reflects cellular cross-sectional area (direct relationship to cell size), whereas SSC is an indication of the cellular granularity. The combination of these two non-fluorescent parameters provides a three-part differential in red blood cell lysed whole blood that distinguishes between normal lymphocytes, monocytes, and granulocytes. As shown in Fig. 93.2A, lymphocytes have the lowest FSC and SSC, monocytes have higher FSC and SSC, and granulocytes have the greatest SSC. This method is effective in distinguishing a relatively pure population of lymphocytes under most circumstances. However, the presence of nucleated red cells, large platelets, basophils, or other particulate debris can produce contaminating events (cells) within the lymphocyte gate. Furthermore, malignant or activated lymphoid cells may not fit into the previously outlined standard light scatter patterns.

A method for confirming the integrity of the light scatter-based lymphocyte gate uses the monoclonal "gating" reagents anti-CD45 and anti-CD14.[11] These two monoclonal antibodies more accurately identify the three-part differential. Lymphocytes have the highest level of CD45 binding but are CD14 negative; granulocytes have a lower level of CD45 binding and are intermediate CD14 positive; while monocytes have high levels of both CD45 and CD14 expression (see Fig. 93.2B). Importantly, non-leukocytes, including erythrocytes and platelets, are negative for these markers. However, malignant leukocytes with characteristics of early precursor cells often have altered CD45 and/or CD14 expression that must be recognized when studying hematologic malignancies. Gating reagents provide a reliable means of checking the light scatter-based lymphocyte gate for the frequency of non-lymphocytes within the gate and the extent of lymphocyte exclusion from the gate. Guidelines for an acceptable degree of contamination within the lymphocyte gate, as well as the level of lymphocyte exclusion, are contained within the US Clinical and Laboratory Standards Institute guideline for lymphocyte immunophenotyping.[12]

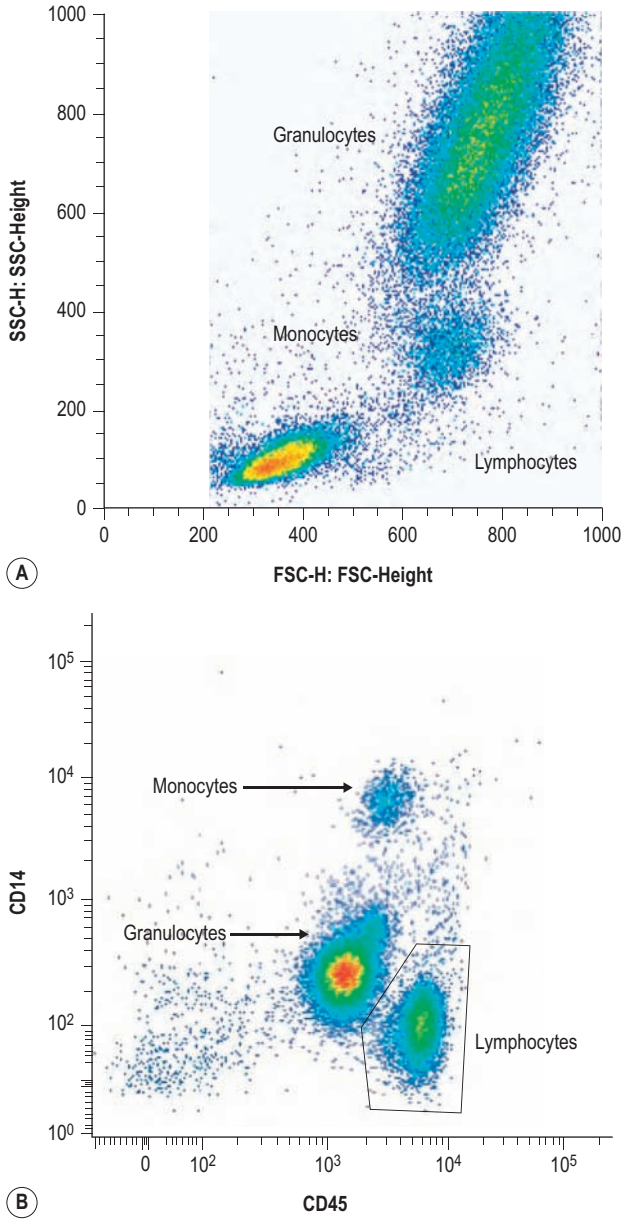


FIG. 93.2 (A) Gating based on FSC (forward scatter) and SSC (side scatter) dot plots on a lysed whole blood sample, demonstrating the basic three-part leukocyte differential with lymphocytes, monocytes, and granulocytes. (B) Dot plot with CD45/CD14 gating reagents showing the fluorescence distribution of all the three leukocyte types identified to include lymphocytes, monocytes, and granulocytes, as well as a small number of non-lysed red blood cells and/or debris.

Data Display

KEY CONCEPTS

Data Presentation

- Fluorescence intensity is plotted versus cell number.
- Can present cumulative data on more than one parameter for each cell.
- Multicolor data presentation can increase cell subpopulation resolution.

The simplest method for demonstrating flow cytometry data is the single-parameter histogram (Fig. 93.3), a graphic presentation of cell number on the *y*-axis versus fluorescence (light) intensity from a single fluorochrome on the *x*-axis. Integration of curve areas provides the number of cells, and often there are two distinct distributions; one referred to as negative identifies cells that are not bound specifically by the monoclonal reagent, and the second represents cells bound by the antibody. Negative reflects low-level fluorescence resulting from cellular autofluorescence and any nonspecific binding of the monoclonal reagent(s), the magnitude of which varies between different cell types. The interpretation of the data is simplified when there are two distinct cell populations (i.e., negative and positive), while the evaluation of two overlapping distributions is more difficult.

Multiparameter data can be evaluated using a series of single-parameter histograms that consider each fluorochrome independently. However, presenting two parameters simultaneously using a correlated display is more informative (Fig. 93.4). This approach enables the simultaneous visualization of four different populations: A^+/B^-, A^-/B^+, A^+/B^+, and A^-/B^-. More recently, these displays evolved to include a mixture of logarithmic (for higher intensity expression) and linear (for lower intensity expression) intensity for each axis to allow for better interpretation of events with very low to absent fluorescence. This combined display approach resolves events being displayed compacted against the axes even with properly compensated samples and will be used in the illustrations throughout this chapter.

The simultaneous use of *n* monoclonal reagents can identify a total of 2^n subpopulations. These different subsets can be identified sequentially by first dividing the cells into those that are positive versus negative with one reagent and then evaluating the defined subpopulations for additional reagents using a two-color approach. Alternatively, modern software can represent multiple populations as polychromatic plots, which can simplify data analysis.[13] The polychromatic approach can provide a means to further resolve subpopulations and has been particularly useful in the evaluation of cellular differentiation, activation, and functional correlates, as well as clarifying overlapping cell subpopulations.

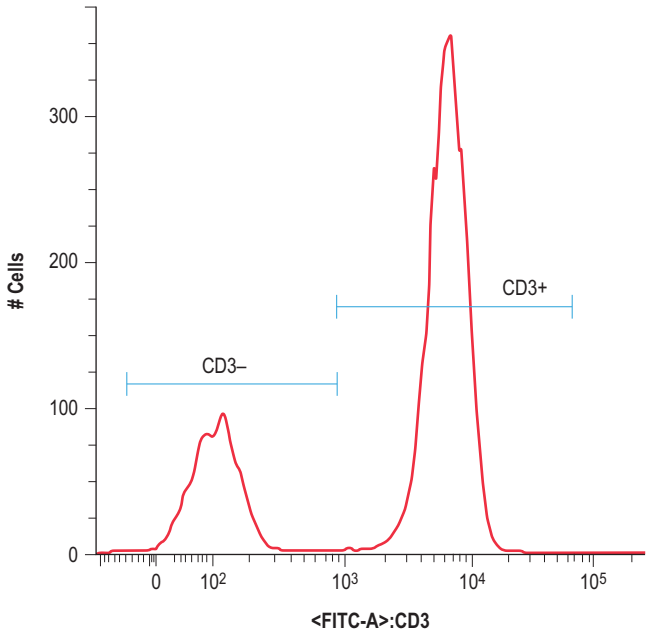


FIG. 93.3 Single-parameter histogram for CD3 expression on lymphocytes demonstrating the negative, non-T-cell population (B cells, NK cells) and a positive T-cell population. Integrating the area under each curve would provide the numbers and percentage of cells present in each of the two subpopulations.

Positive–Negative Discrimination

Evaluating clinical immunophenotyping data requires establishing criteria for the boundaries between negative or nonstained (autofluorescence plus nonspecific stained) cells and positive (specifically stained) cells. A commonly used approach involves using directly conjugated control monoclonal antibodies of the appropriate class or subclass (e.g., IgG1, IgG2a, IgG2b, or IgM) that do not specifically react with human lymphocyte surface antigens (commonly called "isotype controls"). The marker (discriminator) is set at the fluorescence histogram channel number such that it includes 98% to 99% of the negative cells (Fig. 93.5A).

The negative cell population is based on the aggregate of baseline cellular autofluorescence plus nonspecific binding that varies according to cell type. For this reason, the use of isotype controls may not correctly identify the positive-negative threshold for specific cell types, particularly when staining dimly expressed proteins. Perfectly mimicking the specific antibody used would require the isotype control to have the same antibody to fluorochrome ratio and brightness, something not easily accomplished. To overcome these difficulties, an alternative method has been described and is referred to as Fluorescence-Minus-One (FMO).[14]

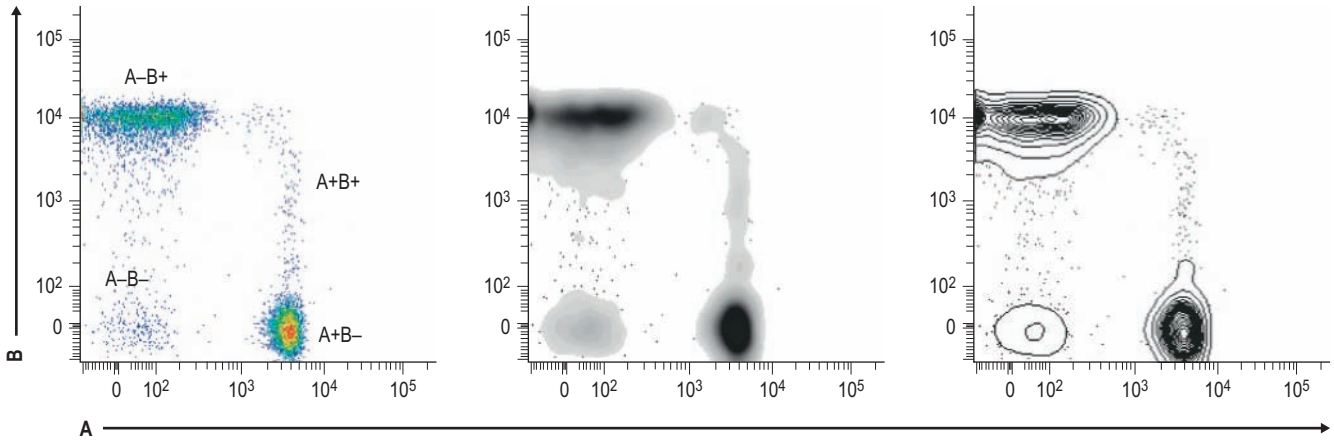


FIG. 93.4 Examples of dot pseudocolor *(left)*, density *(middle)*, and contour *(right)* displays based on the same two-color parameter data. All three techniques enable simultaneous evaluation of both parameters, in this case evaluating the expression of markers A and B. These plots identify four populations of cells, those expressing only A (A^+), only B (B^+), both A and B (A^+/B^+, very few), and expressing neither A nor B (A^-/B^-).

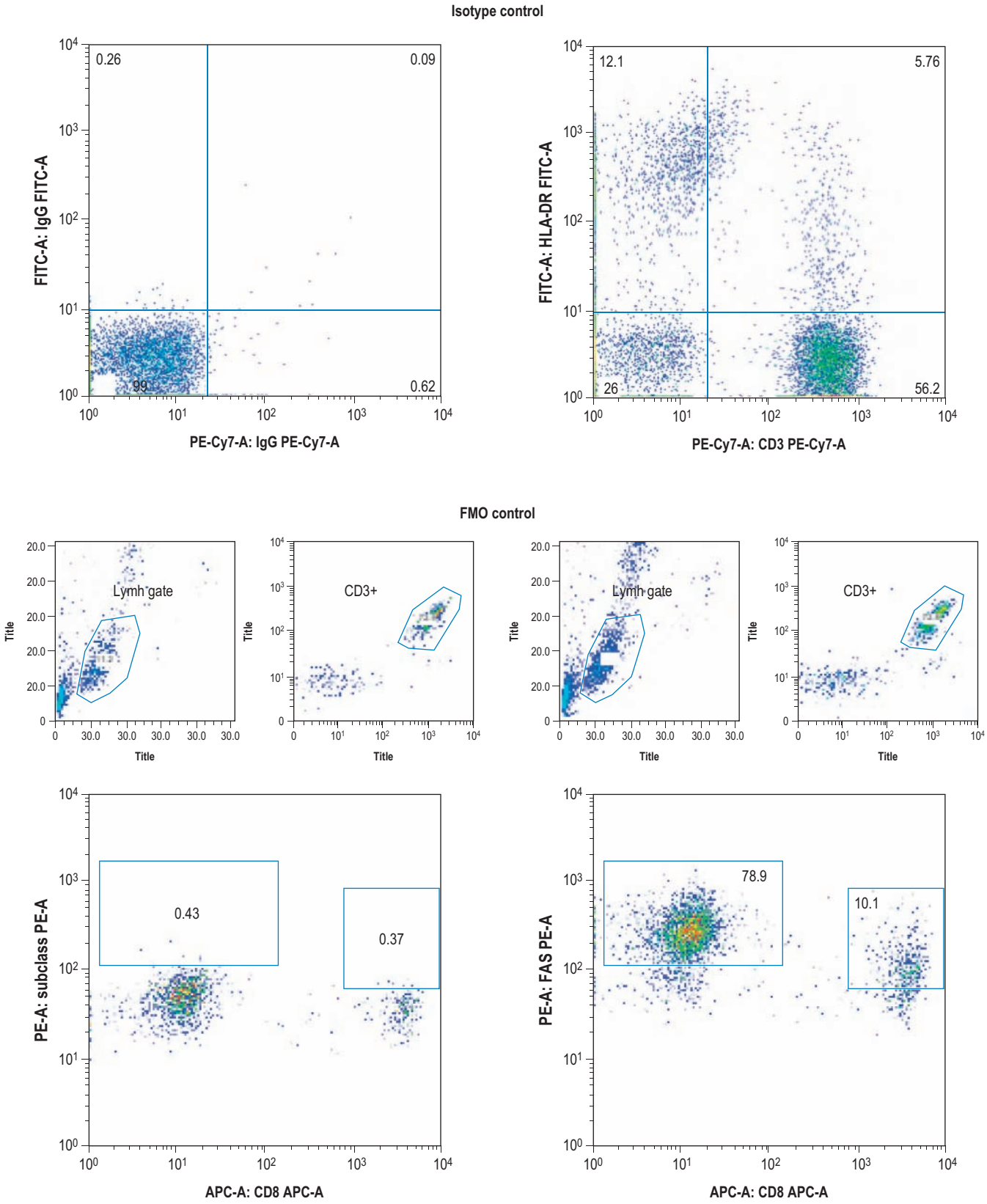


FIG. 93.5 Positive–negative discrimination strategies. (A) Nonspecific antibodies (isotypes) were used to stain the sample shown on left, and the positivity threshold determined and applied to the sample on the right. (B) FMO. The positivity threshold was determined by staining the sample on the left with the population-specific markers (CD3, TCRαβ, and CD8) and omitting antibody to CD95 (FAS). The panel on the right contains antibody to CD95. Note that the positivity threshold is slightly different for $CD8^+$ and $CD8^-$ cells.

FMO involves the staining of the sample with all the antibodies of interest, except the one targeted for positive-negative threshold. As an example, to define the negative threshold for the protein CD95 (FAS) in $CD8^+$ and $CD8^-$ T cells, the FMO control tube would include the cell subset-specific markers (CD3, TCRαβ, and CD8) and omit anti-FAS. After appropriate gating on that population, the threshold can be adequately defined, and it is different for the two exemplified populations (see Fig. 93.5B, right panel). This method is more costly due to the multiple control tubes needed for each sample.

Compensation

The fluorescence signals emitted by different fluorochromes are not completely separated by the filters. This can lead to signal overlap that is corrected by electronic subtraction of the overlapping signal, a process referred to as compensation. The overlap is particularly significant when using multiple fluorochromes, each with different spectral properties.[14] The compensation process involves subtraction of the "spillover" signal detected by the photodetector that is generated by samples stained with only one fluorochrome. Currently, most flow cytometry analysis software allows for off-line compensation, where the single reagent stained tubes are used to create a compensation matrix that is then applied to all the tubes in the experiment. This allows for much-simplified compensation procedures without the need for any hardware compensation during data collection.

Quality Control

Quality control is a critical component of clinical flow cytometry to ensure optimal results.[15] This includes monitoring instrument setup and performance, optimizing sample preparation and reagents, and standardizing controls and data interpretation. Quantitative flow cytometry based on a fluorescence standard curve provides quantitative data in units referred to as molecular equivalents of soluble fluorochrome (MESF). Finally, participation in interlaboratory proficiency testing surveys, such as the triannual samples provided by the College of American Pathologists (CAP), is an important additional measure to monitor laboratory performance, and this is mandated in US clinical laboratories by the Clinical Laboratory Improvement Amendment of 1988 (CLIA 88).

METHODS

Whole-blood lysis represents the most common technique and consists of mixing a fixed volume of anticoagulated whole blood (or bone marrow) with one or more directly conjugated monoclonal antibodies, followed by incubation at a designated temperature and time. Next, the red blood cells are lysed and the sample is washed and fixed to reduce infectious risk before being run into the flow cytometer. The non-lysed cells remaining include all peripheral blood leukocytes and any non-lysed red cells, platelets, and debris. The heterogeneity of the sample necessitates careful lymphocyte gating (see section above) to generate accurate immunophenotyping data. The advantages of the whole-blood lysis method include fewer preparation steps, less sample handling, and a lower likelihood of differential lymphocyte loss. Flow cytometry can also evaluate alternative sources of cells (e.g., bone marrow, bronchial alveolar lavage fluid, fine-needle aspirates). Patient studies must be determined with the same methods and reagents as used to determine the control ranges to ensure comparability. The number of events (cells) analyzed typically ranges from 10,000 to 20,000 in routine clinical studies but must be increased when evaluating very small subpopulations of cells to produce statistically relevant data.

TABLE 93.1 Selected Lymphocyte Surface Antigens for Immunophenotyping

T cells
Pan-T cell: CD3, CD2, CD7, CD5
Major T-cell subset: CD4, CD8
Surface antigens associated with function: CD28, CD31, CD38, CD45RA, CD45RO, CD62L, CCR7
Activation antigens: CD25, CD40L, CD69, CD71, HLA-DR

B cells
Pan-B cell: CD19, CD20, surface immunoglobulin
Major B-cell subset: CD5, CD21
Surface antigens associated with function: CD27, CD40, CD80, CD86
Activation antigens: CD23, CD25

NK cells
Pan-NK cell: CD16, CD56
NK subset: CD2, CD8, CD57

The application of control ranges must consider the fact that significant changes occur in lymphocyte distribution and development during childhood, as well as changes in lymphocytes that occur among the very elderly.[16] Other factors can also impact lymphocyte distribution, including race, gender, diurnal variation, recent or intercurrent infection.

The choice of immunophenotyping reagents depends on the cells being targeted for study and the question being asked. However, regardless of the specific set-up, the inclusion of a tube with gating reagents (anti-CD45 and anti-CD14) to confirm the integrity of the standard lymphocyte gate is recommended.[11] In addition, control reagents should be included to establish the fluorescence intensity of negative cells. Important internal controls include a pan-T-, B-, and NK-cell marker for every sample (Table 93.1), based on the principle that the whole is the sum of its parts. Thus, the total percentage of lymphocytes in the gate determined by the gating reagents should approximately equal the sum of the percentages for T, B, and NK cells. A technical or biological explanation must be identified when this relationship does not hold. These would include the presence of immature or malignant cells that were not identified by standard pan-T-, B-, and NK-cell reagents. In addition, if the gating reagents (CD45/CD14) had not been included, contaminating cells (e.g., myeloid precursors, nucleated red blood cells, large platelets) with forward- and side-scatter characteristics similar to those of lymphocytes could not be ruled out. Potential technical problems include reagent or fluorochrome degradation, failure to add a reagent, and a host of others. Evidence for any major technical error requires repeating the study.

Additional data that can be used for controls depend on the setup. For example, the availability of multiple antibodies that identify a similar cell sub-population can serve as a useful check (e.g., total T cells by comparing CD3 and CD5 or CD2; total B cells by comparing CD19 and CD20). In addition, the use of specific reagents in more than one tube enables comparison between the repeat values. The application of internal checks should be performed by the flow operator as a simple means of confirming the validity of the data. Insights regarding unusual biological findings may also be uncovered through this type of evaluation (e.g., the presence of an increased population of $CD4^-/CD8^-$ double-negative T cells).

The challenge in performing immunophenotyping is to accurately identify cells with specific surface characteristics (antigens). As previously noted, the capacity to discriminate cell subpopulations is often enhanced through the directed use of antibody combinations. The typical data generated consist of the percentage of negative versus positive cells when using one reagent and multiple subpopulations when using more than one reagent. Regardless of the experimental design, it is important to consider not only the percentage of cells within each subpopulation but also the absolute number of cells. This is most commonly obtained by multiplying the relevant percentage from the flow cytometer by the absolute lymphocyte count obtained using a white blood count and differential. For example, when assaying for CD4 T-cell counts, the percentage of $CD4^+$ cells is multiplied by the absolute lymphocyte count to yield the absolute CD4 count. However, using two separate procedures (i.e., dual platforms) to generate the final result introduces the possibility of an additive error based on the inherent errors of the two different methods. This fueled a search for approaches performing both tasks by flow cytometry (i.e., a single platform). One alternative involves the inclusion of a fixed number of fluorescent beads (in a defined volume) in each tube as a reference standard to generate absolute numbers without requiring the use of an absolute lymphocyte count. An alternative approach involves the use of impedance-based cell counting in the flow cytometer to generate an absolute lymphocyte count (dependent on a fixed volume of sample being run) to generate both percentage and absolute number for each specific population or subpopulation. Regardless of the approach, the reporting of both percentages and absolute numbers is necessary when immunophenotyping peripheral lymphocytes.

The objective of evaluating malignant cells is often to characterize the lineage and differentiation level of the abnormal cells rather than quantifying subpopulations. The pattern of reactivity combined with fluorescence intensity is often useful in identifying leukemic patterns, whereas absolute numbers are usually not required. However, flow cytometric detection and quantitation of rare abnormal cells can be useful in evaluating for minimal residual disease post therapy in lymphoproliferative disorders.

PRACTICAL APPLICATIONS OF FLOW CYTOMETRY

Immunophenotyping Studies

CLINICAL RELEVANCE

Immunophenotyping Studies

- Can be used to identify cell subsets, lineage, stage of cell differentiation, state of cell activation, and clonality.
- Lymphocyte results should be checked with T cells + B cells + NK cells = 100%.
- Immunophenotyping studies are not the equivalent of lymphocyte function studies.

The majority of immunophenotyping studies are directed to enumerate specific cell subpopulations, evaluating for the presence or absence of particular surface antigens, identifying the differentiation level of specific cells, determining cell lineage, evaluating for functional correlates based on specific antigen expression, examining for evidence of cell activation, and/or establishing evidence of monoclonality.

Quantification of a particular cell subpopulation can be readily accomplished by flow cytometry generating cell numbers based on the percentages of the respective cell types (and an absolute lymphocyte count). The evaluation of CD4 T-cell counts has formed the basis for monitoring patients infected with HIV.[17] The quantitation of $CD34^+$ hematopoietic stem cells in donor peripheral blood or bone marrow is used in many cellular reconstitution protocols.[18] Subpopulation characterization can also be useful in the evaluation of patients with clinical history and laboratory findings suggestive of immune deficiency.[19] This has taken on a higher level of significance in the setting of newborn screening for severe T-cell immune deficiency via the T-cell receptor excision circle (TREC) assay that, when abnormal, is typically followed by immunophenotyping to evaluate for naive T cells.

When evaluating for the presence or absence of surface proteins associated with specific functional attributes, it is important to realize that this approach does not assess the actual functional status of the cells. This point is clearly illustrated by the finding of normal B-cell numbers in most patients with common variable immune deficiency (CVID) despite the fact that these patients fail to produce immunoglobulins normally (see Chapter 33). However, changes in the characteristics of the B cells, particularly relative to memory B cells, provide potential insight into different phenotypes of this disorder and provide additional support for heterogeneity of patients with CVID. Due to the limitations of immunophenotyping, it is common practice when evaluating the status of the immune system to perform cell function testing in parallel.

Flow cytometry can be used to test for the presence or absence of a specific cell surface antigen. An example of this type of application is in the evaluation of a patient with a history of recurrent skin infections, delayed wound healing, and persistent granulocytosis, which suggests a diagnosis of leukocyte adhesion deficiency type 1. This disorder results from a defect in the gene encoding CD18, preventing the expression of three different heterodimeric adhesion molecules (β2 integrins), each containing CD18 (see Chapter 39). This disorder can usually be diagnosed by studying granulocytes (and lymphocytes) for the expression of CD18 (as well as the three isoforms of CD11). Patients often have decreased rather than absent CD18 expression, and further confirmation of the diagnosis can be accomplished by demonstrating a failure of CD18 upregulation following granulocyte activation.

Immunophenotyping can also help address questions regarding the level of cell differentiation. Antibodies specific to proteins expressed by early (precursor) cells represent one approach and would include evaluating for the thymocyte marker CD1 or the pre-B-cell marker CD10 (CALLA), to cite two. However, defining the developmental level of a particular cell population or subpopulation is best accomplished using a panel of reagents that span the natural history of the cell lineage. This approach represents the standard for testing leukemias and lymphomas (see Chapters 77 and 78). In addition to defining the presence or absence of specific antigens, evaluating their level of expression is also valuable, which may be altered in the abnormal cells. Malignant cells may also express antigens associated with different lineages and have altered forward- and side-scatter characteristics, as well as diminished or absent CD45 expression, requiring modified gating approaches.

Issues of monoclonality can be dealt with using flow cytometry when analyzing B cells, and in some circumstances, when studying T cells. Normally B cells are a heterogeneous mixture of mutually exclusive κ or λ light-chain-positive cells. Measuring the distribution of κ or λ light-chain expressing B-cells or plasmacytes can be informative as to the presence or absence of monoclonality (see Chapter 79). The capacity to evaluate T-cell

monoclonality by flow cytometry is less definitive and consists of using T-cell antigen receptor β-variable (Vβ) chain-specific reagents to look for evidence of significant over representation of one Vβ chain family. This approach currently consists of setting up a series of tubes, each with three different Vβ family-specific monoclonal antibodies, one conjugated with FITC, one with PE, and the third with FITC plus PE. This combination enables distinguishing the frequency of each of the three different Vβ families per tube (green$^+$, orange$^+$, green$^+$/orange$^+$) and represents a flow cytometric method to complement PCR-based spectratyping.[20]

The state of lymphocyte activation can be addressed by evaluating for the presence of surface antigens that are found only on activated cells or are upregulated following activation. These include receptors for specific growth factors (e.g., IL-2 receptor α chain, CD25), receptors for critical elements required for cell growth (e.g., transferrin receptor, CD71), ligands for cell–cell communication following activation (CD152 [CD40 ligand] on activated CD4 T cells), and surface antigens that are upregulated as a result of activation (e.g., adhesion molecules, HLA-DR, CD69). In addition, the memory status of both T cells and B cells can be assessed based on differential surface molecule expression associated with prior antigen encounter. This enables a distinction to be made between naïve T cells that express CD45RA, CD31 (recent thymic emigrants), CD62L, and CCR7 from memory T cells that express the alternative CD45 isoform, CD45RO (and varied CD62L or CCR7, depending on whether the cells are central or effector memory cells).[21] In addition, memory B cells can be detected by the expression of CD27 and be further divided into isotype-switched and non-switched memory cells based on their pattern of surface immunoglobulin expression.[22]

Defects associated with familial lymphohistiocytosis (FLH) are generally associated with abnormal NK cell function. Many of the FHL-causing defects can be determined by flow cytometry. For example, SAP and XIAP intracellular staining (see next section) can be used to evaluate for X-linked lymphoproliferative (XLP) disorders type 1 and 2, respectively. Likewise, lack of intracellular perforin expression in NK cells would be indicative of hemophagocytic lymphohistiocytosis (HLH) type 2. Additionally, the evaluation of CD107a surface expression, which is normally expressed on cytoplasmic granules and upon degranulation in response to incubation with specific target cells (e.g., K562 cells), is expressed on the surface of NK cells and is useful in determining the underlying genetic defect causing FHL.[23] Specifically, lack of surface CD107a is indicative of syntaxin-11 or MUNC-13.4 defects.

Intracellular Evaluation

Cellular Activation

CLINICAL RELEVANCE

Intracellular Flow Cytometry

- Activation-directed studies:
  - Cell (lymphocyte) proliferation
  - Calcium flux
  - Intracellular protein phosphorylation
  - Oxidative burst: neutrophils
- Intracellular cytokine studies:
  - Clarify the Th1/Th2/Th17 status of an immune response
  - Can be assessed in an *in vitro* antigen-specific response
  - Can be combined with evaluation of cell surface studies

Ligand binding and transmembrane signal transduction resulting in cellular activation can be evaluated using flow cytometry. Changes in intracellular ionic calcium concentration (Ca^{2+}) are frequently used to monitor cell activation after ligand binding. These changes are associated with the activation of phospholipase C and protein kinase C. In general, three reagents have been used to measure Ca^{2+}: quin 2, indo-1, and fluo-3. Quin 2 has a low excitation coefficient and is not useful for flow cytometry; indo-1 requires ultraviolet excitation; fluo-3 can be excited by 488 nm but does not permit ratiometric analysis. Nevertheless, because of its ease of use fluo-3 is currently the most widely used probe for intracytoplasmic Ca^{2+} evaluation by flow cytometry. Strict attention must be paid to loading conditions, the presence or absence of free Ca^{2+} in the medium, experimental temperature, baseline measurements, and calibration. This approach can be combined with cell surface marker or cell cycle evaluation.[24]

Intracellular pH changes related to cellular activation also can be evaluated. The most useful probe for pH is SNARF-1.[24] This probe can be excited at 488 nm and allows for ratiometric analysis with detection wavelengths set for 575 and 640 nm. Glutathione (glutamylcysteinylglycine [GSH]) is an important antioxidant generated during cell activation that can be measured by flow cytometry.[24] The fluorescent probe monochlorobimane is commonly used for this measurement, but it is complicated by the need to determine GSH by high-performance liquid chromatography (HPLC).

Additional approaches to evaluate cellular activation include assessment of intranuclear markers (Ki-67, PCNA) as well as surface proteins that are upregulated following cellular activation (e.g., CD69, CD25, CD71).[25] Actual cell division can be evaluated using cell tracking dyes (e.g., PKH26, CFSE, Cell Trace™ Violet) that lose 50% of their fluorescence with each round of cell division.[26] This approach has become more common in the clinical assessment of lymphocyte function due to the capacity to evaluate specific lymphocyte subpopulations responding to mitogenic and antigenic stimuli. Cell labeling dyes also can be used to label target cells in cell-based cytotoxicity assays.[27] Recently, an approach to evaluate lymphocyte proliferation following cell stimulation has been described using the thymidine analog, EdU. Detection of DNA synthesis induced by the different activating agents is measured using a copper-catalyzed click chemistry, which results in EdU being covalently bonded to a fluorescent azide.[28] This approach allows the assessment of cell proliferation at the cell population or subpopulation (e.g., CD3, CD4, CD8) level and can be used in association with mitogen and recall-antigen stimulation.

An alternative approach to the functional evaluation of cell activation is based on the detection of phosphorylated intracellular proteins associated with specific activation signals using flow cytometry. An example of this is the detection of phosphorylated signal transducer and activator of transcription 1 (STAT1) following interferon-γ (IFN-γ) stimulation of monocytes, which has been found to be equal to or more sensitive than immunoblotting.[29] Using this same principle, a more comprehensive evaluation of multiple signaling pathways using different cytokines (e.g., IL-2, IL-4, IL-6, IL-7, IL-21, and IFNα) and combinations of JAK-STAT proteins in CD4$^+$ T cells, or alternative activation pathway in macrophages (through IL-4), could also be determined (Fig. 93.6). This type of assay requires fixation and permeabilization optimized to the phosphoepitope

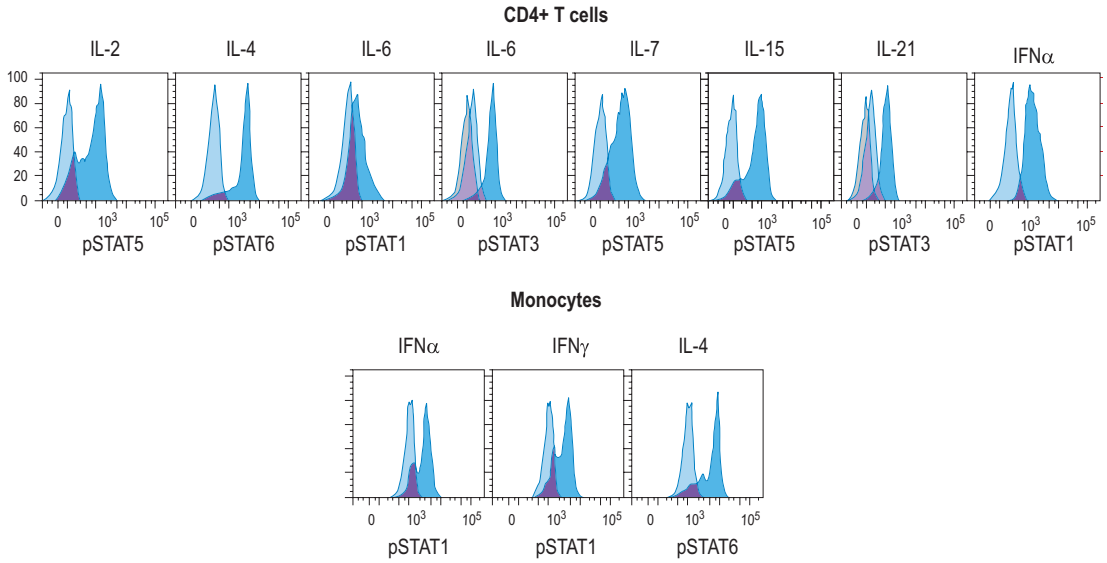


FIG. 93.6 Phosphorylation of STAT proteins in CD4⁺ T cells (upper panel) and monocytes (lower panel) upon cytokine stimulation. Isotype control is represented in gray and baseline levels in light blue.

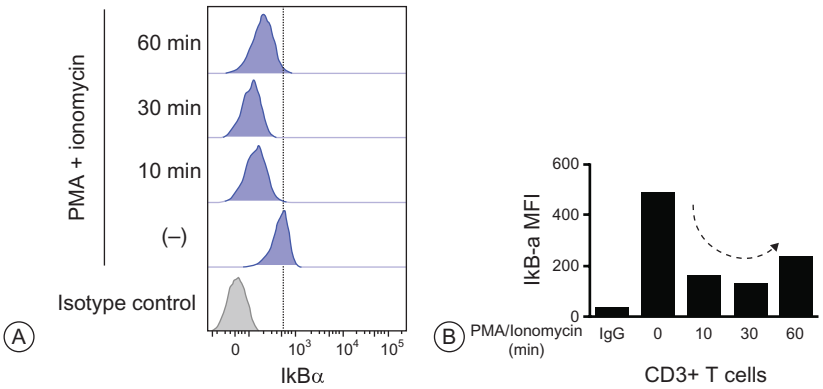


FIG. 93.7 (A) Histograms of intracellular IkBa in CD3⁺ lymphocytes at baseline (–) and upon stimulation with phorbol myristate acetate (PMA) + ionomycin. Degradation of IkBa followed by re-synthesis is seen over time. (B) Median fluorescence intensity (MFI) of IkBa at baseline (0) and three time points following stimulation.

of interest to allow the entry of the specific reagent. Similarly, the compatibility of surface epitopes and fluorochromes with certain fixation and permeabilization protocols needs to be considered. Currently, several intracellular signaling proteins that undergo phosphorylation following a specific activation signal can be assessed using commercially available reagents.

The NF-κB pathway has an important role in regulating lymphocyte development, immune responses, inflammation, cell proliferation, and cell death. Two signaling pathways have been described to be associated with NF-κB: canonical and noncanonical. Signaling through the NF-κB canonical pathway depends on IκBα phosphorylation, ubiquitination, and proteasome degradation upon cell stimulation. Once IκBα is degraded, it releases c-Rel/p65 and p50 to be translocated to the nucleus in order to initiate gene transcription. Evaluation of IκBα phosphorylation (phospho-IκBα), as well as IκBα degradation and re-synthesis kinetics using flow cytometry, and intracellular staining allows for the determination of the NF-κB canonical pathway signaling integrity (Fig. 93.7).

The assessment of oxidative burst following cell stimulation plays a central role in neutrophil function testing using the hydrogen peroxide-sensitive dye dihydrorhodamine 123 (DHR123). This procedure involves loading granulocytes with the dye, stimulating with phorbol myristate acetate (PMA), and evaluating for fluorescence by flow cytometry.[9] This test has proved to be extremely accurate in diagnosing patients with chronic granulomatous disease (CGD) and carriers of X-linked CGD.

An important advantage is its sensitivity, which allows the detection of one normal cell in a population of 1000 abnormal cells. This makes the assessment of oxidative burst a useful tool in following allogeneic granulocyte survival after transfusion into patients with CGD, as well as a means of following donor chimerism in the setting of allogeneic stem cell transplantation. It also provides a method to identify corrected cells following gene therapy in CGD and has utility in predicting disease outcome in CGD.[30]

Intracellular Cytokine Detection

Flow cytometry affords a platform to evaluate cytokine production at a single-cell level using cytokine-specific directly conjugated monoclonal antibodies following fixation and permeabilization of cells. This approach allows for the simultaneous

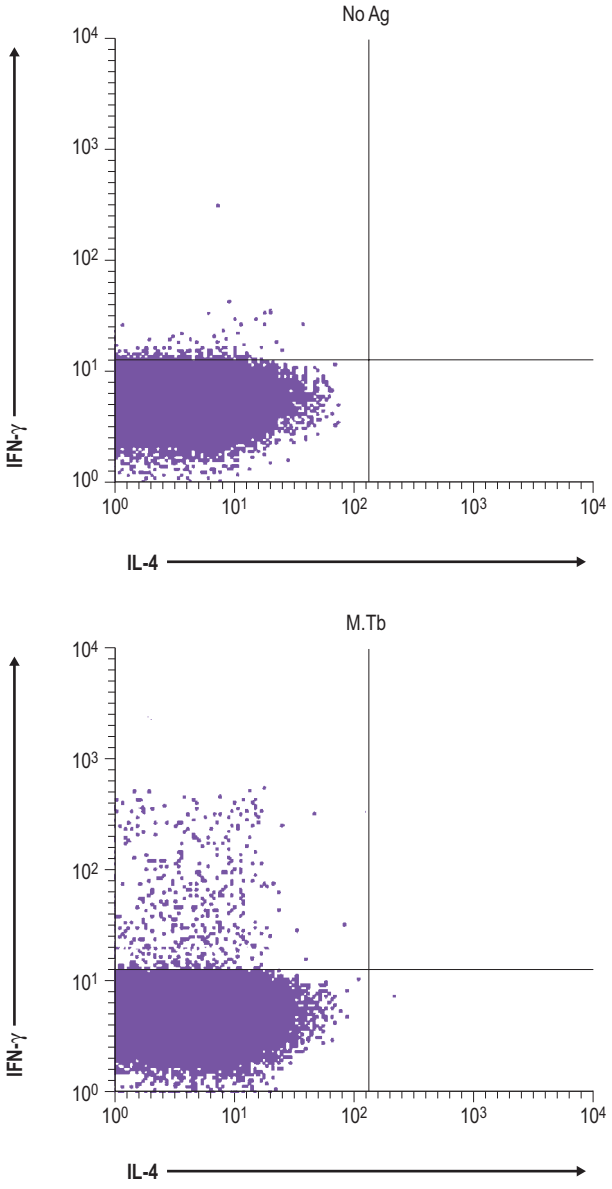


FIG. 93.8 Two-color dot plots of CD3+ T cells evaluated for intracytoplasmic interferon-γ and IL-4 expression. The donor had a positive skin test to PPD and demonstrated a Th1 pattern of cytokine expression (interferon-γ) in response to *Mycobacterium tuberculosis* antigen, with an absence of a Th2 cytokine pattern (IL-4). Courtesy of Calman Prussin, MD.

detection of two or more intracellular cytokines in combination with cell surface markers and/or other intracellular markers. Important aspects of intracellular cytokine detection include the use of a protein transport inhibitor during activation, the use of proper controls, and the choice of antibodies. As there is little or no spontaneous cytokine production in circulating human lymphocytes, intracellular cytokine detection requires *in vitro* activation. Initial experience was based on supraphysiological stimulation using PMA and ionomycin, but antigen-specific activation systems have also proven to be feasible. It should be emphasized that regardless of the activation method, the duration of activation is an important variable, as individual cells reach maximum cytokine production at different times. In addition, different cytokines have different optimal periods of activation. It is recommended that a proper kinetic profile be established for the biological system or clinical condition being studied.[31]

To increase the level of intracellular cytokines, inhibitors of intracellular protein transport and secretion (e.g., monensin or brefeldin) are commonly used, which lead to the accumulation of proteins within the cell. Nonspecific binding of the antibody reagents is an issue, as permeabilization allows access not only to the cytokine of interest but also to other proteins present in much greater quantities than on the cell surface. In addition, fixation further increases nonspecific binding and the use of both a negative-control sample, which contains an excess of unlabeled or "cold" anti-cytokine antibody, and an isotype-matched or FMO-control sample provide the optimal control. When the conjugated anti-cytokine is added to the negative-control sample, it can only bind to other proteins in a nonspecific manner, thereby providing a measure to discriminate between specific and nonspecific binding. The use of directly conjugated anti-cytokine antibodies not only simplifies the staining procedure but also provides the best distinction between specific and nonspecific binding. Because the fixation agent may change the native state of certain epitopes, it is also important to use antibodies that recognize antigens after fixation when combining cell surface characterization with intracellular cytokine evaluation.

One of the main applications of intracellular cytokine detection by flow cytometry has been the study and refinement of the Th1/Th2/Th17 paradigms. It has recently become clear that the regulated secretion of cytokines can be used to study the response of individual T cells to both polyclonal stimuli and specific antigens. Measuring antigen-specific T-cell cytokine expression in response to specific antigen offers a useful alternative to the tetramer-based approach (discussed below) to quantify the frequency of antigen-specific T cells (Fig. 93.8).[31]

Cell Cycle Analysis

CLINICAL RELEVANCE

Cell Cycle Analysis

- Useful for screening percentage of S phase and aneuploidy
- Can be combined with cell surface studies
- Can be combined with markers of apoptosis

In addition to surface immunophenotyping and cytoplasmic characterization, flow cytometry is also used in cell cycle analysis. Propidium iodide (PI) is the most used fluorochrome, owing to its optimal linear DNA-binding capacity in a variety of different cell types. Thus, a single-parameter histogram of DNA content using PI readily permits the determination of cell cycle compartments, expressed as the percentage of cells in G_0–G_1, S, and G_2-M (Fig. 93.9A). In addition to these conventional parameters, the presence or absence of aneuploidy can be determined by inspection of the G_0–G_1 peak and/or use of a DNA index (ratio of abnormal DNA content to a diploid DNA standard). Also, elevation in the S and/or G_2-M phase can be detected. The optimal display of these data uses a combination of side scatter versus DNA content. Cells observed on the histogram in the area below the level of G_0–G_1 may be undergoing apoptosis.[32] When dealing with DNA staining, a consistent cellular source of DNA (e.g., chicken erythrocytes) should be used as an internal reference for evaluating DNA content and evaluating the cell cycle distribution.

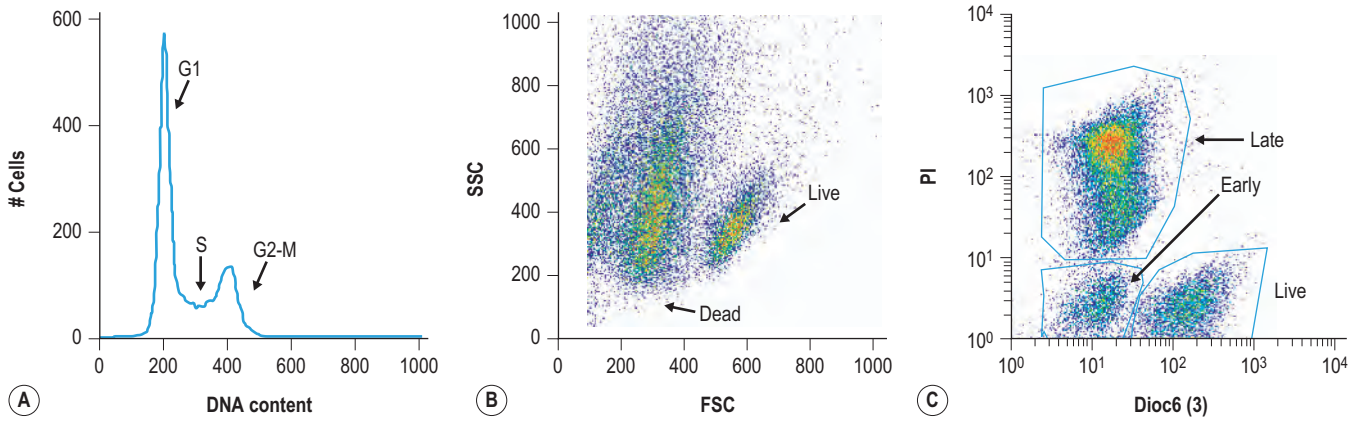


FIG. 93.9 (A) Assessment of DNA content as a reflection of cell cycle demonstrating cells in G_1, S, and G_2/M phases. (B) Assessment of live versus dead cells based on forward- and side-scatter characteristics. (C) Assessment of cell apoptosis using PI and Dioc6 (3) identifying cells that recently initiated apoptosis (early), cells that are dead (late), and cells that are alive (live).

It should be noted that several different computer algorithms have been developed to determine the relative proportion of each cell cycle compartment, and the selection of a software program is not a trivial process. The major instrument manufacturers supply cell cycle analysis programs, and there are also third-party programs available. Generally, the optimal program should be capable of modeling two or more aneuploid populations, subtracting debris (particularly if formalin-fixed paraffin-embedded archival material is used), and accurately estimating S-phase cells.[32] The combination of a surface marker and cell cycle has been very useful in differentiating normal cell populations from tumor cell populations. One example is the use of anti-κ, anti-λ, or B-cell reagents to separate the aneuploid B-cell clone from the remaining normal, reactive B cells in a lymphoid cell mixture. Another uses cytokeratin as a marker to distinguish between the tumor cells and the inflammatory cells that are present.

The other major event that has occurred in cell cycle analysis has been the development of technology using the incorporation of bromodeoxyuridine.[33] This thymidine analog is used to directly determine the percentage of S-phase cells. Also, when used in kinetic studies, it permits a determination of the individual times for the components of the cell cycle and a determination of the growth fraction.

Apoptosis Detection

Flow cytometry has become the method of choice for the detection and quantification of cellular apoptosis.[34] This is due in part to its capacity for rapid assessment of a large number of cells and samples. Many distinct features of an apoptotic cell can be evaluated by flow cytometry based on light scatter, plasma membrane changes, mitochondrial transmembrane potential, DNA content, and DNA integrity.

The light-scattering properties of a cell undergoing programmed cell death are the simplest attributes that can be assessed by flow cytometry. Dying cells typically shrink, producing a loss in FSC and, despite an initial transient SSC increase, there is ultimately a decrease in SSC (see Fig. 93.9B). The use of light scatter can be combined with cell surface staining to help characterize the dying cells. However, scatter changes alone are not specific to apoptosis and should be accompanied by an additional characteristic associated with cell death. Live cells have phospholipids asymmetrically distributed in the inner and outer plasma membrane, with phosphatidylcholine and sphingomyelin on the outer surface and phosphatidylserine (PS) on the inner side. Early during apoptosis, cells lose asymmetry, exposing PS on the outside. Annexin V is a protein that binds preferentially to negatively charged phospholipids such as PS, and directly conjugated annexin V is a useful reagent for the specific detection of apoptotic cells.[34]

Another characteristic of plasma membranes associated with live cells is that they exclude charged cationic dyes such as propidium iodide (PI) and 7-amino-actinomycin-D (7-AAD). Consequently, only cells in a late stage of apoptosis, with ruptured cell membranes, will take up these dyes. Thus, the combined use of cationic dyes (e.g., PI) with annexin V allows the discrimination between live cells (annexin V negative/PI negative), early apoptotic cells (annexin V positive/PI negative), and late apoptotic cells (annexin V positive/PI positive).

Assessment of mitochondrial transmembrane potential ($\Delta\psi m$) is yet another technique used to identify apoptotic cells. Cells decrease $\Delta\psi m$ very early in the apoptotic process before rupture of the plasma membrane, losing the ability to accumulate potential-dependent dyes such as rhodamine 123, JC-1, or 3,3″-dihexyloxacarbocyanine iodide ($Dioc6^3$). These dyes can also be used with PI to detect cells in the different stages of apoptosis (see Fig. 93.9C).

Measurement of DNA content can also be employed to distinguish live from dead cells, as described above (see Cell Cycle Analysis). This kind of analysis has to be done using a linear scale, not logarithmic, in order to discriminate dying cells from debris. DNA cleavage also exposes -OH termini associated with the DNA breaks, and these can be detected via the attachment of fluorochrome-conjugated deoxynucleotides in a reaction catalyzed by exogenous TdT, a technique called TUNEL.

Peptide–Major Histocompatibility Complex Multimers

KEY CONCEPTS

Peptide–Major Histocompatibility Complex Tetramers

- Useful for assessing the number of antigen-specific T cells
- Can be directed at both $CD4^+$ and $CD8^+$ T cells
- Requires information about the antigenic peptide and HLA (MHC) restriction

In contrast to B cells, direct visualization of antigen-specific T cells *ex vivo* has, until recently, been unsuccessful. In 1996, Altman et al. introduced a novel flow-cytometry-based methodology that enables the direct visualization and quantification of antigen-specific T cells.[35] The generation of soluble peptide-major histocompatibility complex (MHC) multimers allows for multiple TCRs being engaged at the same time, which results in markedly greater avidity of these multimeric ligands for their peptide-specific TCR. The most common multimer methodology involves engineering a biotinylation recognition sequence on the –COOH terminus of the extracellular domain of one chain of the MHC molecule that, after combining with a specific antigenic peptide, is bound by avidin or streptavidin. As both avidin and streptavidin have four biotin-binding sites, the result is a tetrameric peptide–MHC complex that serves as a ligand for T cells specific for both the peptide and MHC. Flow-cytometric detection is achieved by labeling streptavidin with a fluorochrome. The major pitfall to this approach is the need to know the antigen-derived peptide and its HLA-restrictions, as well as the HLA type of each subject studied. Since the initial report, an increasing number of tetramer-based studies have appeared. Most have focused on the MHC class I-mediated immune response in both mice and humans to a variety of infectious agents, including cytomegalovirus (CMV), HIV, Epstein–Barr virus, and others. Since the initial description with class I-restricted recognition, detection of antigen-specific CD4 T cells with tetramers of soluble MHC class II molecules and covalently linked peptide has also been reported. New methodologies now allow higher valency multimers such as pentamers, octamers, dodecamers, and dextran-based multimers.[36]

In addition to demonstrating the feasibility of this approach, the published studies have provided several new insights into the MHC class I-mediated immune response (see Chapter 5). For example, it has become clear that the extent of the MHC class I-mediated cellular response is much greater than previously estimated. Furthermore, the extensive proliferation of CD8 T cells during an acute infection is not the result of bystander activation but represents an expansion of antigen-specific CD8 T cells. Peptide–MHC -multimer assays have also shown promise in the study of the kinetics of primary and secondary immune responses, as well as in a better understanding of concepts such as immunodominance and clonal exhaustion.

An obviously attractive aspect of this technology is that tetramer staining can be combined with a variety of cell surface and intracellular phenotypic and functional markers. Already there are indications that the phenotype of antigen-specific T cells varies between individuals and between different phases of the immune response. In addition, tetramer-positive T cells can be sorted for further analysis, such as cytotoxicity assays or *in vitro* expansion. The multimer-based technology has not only proved useful for the study of the immune response to infectious agents, but it has also been applied to the study of oral tolerance, autoimmune conditions, and tumor immunology. It is likely that this highly sensitive and specific technology and other approaches that define antigen-specific response will find many more applications and will lead to new discoveries and a reassessment of certain existing concepts.[37]

CONCLUSION

ON THE HORIZON

- The combination of flow cytometry and mass spectrometry expands the number of cell markers analyzed simultaneously (>35 currently performed), including extracellular and intracellular targets to provide an extensive evaluation of functional markers. However, this technology has not yet entered the majority of clinical laboratories.
- Flow cytometry utilizing spectral rather than conventional light analysis has the potential to allow more "colors" to be accurately evaluated without the need for complex compensation schemes with conventional fluorochromes. This could facilitate more extensive polychromatic studies to be performed easily but is not currently offered by most clinical laboratories.

Flow cytometry has become readily available in clinical laboratories, and the application of this technology has moved forward in parallel to significant improvements in instrumentation and the availability of an array of monoclonal reagents. Properly performed flow cytometry can provide rapid and accurate lymphocyte subpopulation identification. The primary clinical indications of immunophenotyping remain quantifying CD4 T-cell counts in HIV infection, lineage assignment in leukemias and lymphomas, evaluation of other hematologic cell types, and assessing CD34 expression to identify stem cells for transplantation. Additional uses include characterizing immune deficiency disorders, evaluating immune-mediated inflammatory diseases, and assessing patients following organ transplantation. Generally, immunophenotyping does not represent a diagnostic procedure but rather plays a part in the evaluation and understanding of complex disorders and the longitudinal evaluation of immunomodulatory therapy.

It is critical to recognize that immunophenotyping is a means of identifying cells, but it is not directed at cell function. The expansion of flow-cytometric techniques to evaluate intracellular characteristics, assess intracellular changes associated with activation, characterize apoptosis and identify antigen-specific T cells has moved this technology into the cell function arena. These newer approaches expand the utility of flow cytometry as a valuable method for the characterization of various aspects of immune function.

ACKNOWLEDGMENT

This work was supported in part by the Intramural Research Program of the National Institutes of Health Clinical Center

REFERENCES

1. Shapiro H. *How flow cytometers work*. 4th ed. Hoboken: Wiley; 2003.
2. Kachel V, Fellner-Feldegg H, Menke E. *Hydrodynamic Properties of Flow Cytometry Instruments*. New York: Wiley; 1990.
3. Thompson JM, Gralow JR, Levy R, Miller RA. The optimal application of forward and ninety-degree light scatter in flow cytometry for the gating of mononuclear cells. *Cytometry*. 1985;6(5):401–406.
4. Ward MD, Kaduchak G. Fundamentals of Acoustic Cytometry. *Curr Protoc Cytom*. 2018;84(1):e36.

5. Chattopadhyay PK, Hogerkorp CM, Roederer M. A chromatic explosion: the development and future of multiparameter flow cytometry. *Immunology*. 2008;125(4):441–449.
6. Liechti T, Roederer M. OMIP-060: 30-Parameter Flow Cytometry Panel to Assess T Cell Effector Functions and Regulatory T Cells. *Cytometry A*. 2019;95(11):1129–1134.
7. Chattopadhyay PK, Yu J, Roederer M. Application of quantum dots to multicolor flow cytometry. *Methods Mol Biol*. 2007;374:175–184.
8. Chattopadhyay PK, Gaylord B, Palmer A, Jiang N, Raven MA, Lewis G, et al. Brilliant violet fluorophores: a new class of ultrabright fluorescent compounds for immunofluorescence experiments. *Cytometry A*. 2012;81(6):456–466.
9. Vowells SJ, Sekhsaria S, Malech HL, Shalit M, Fleisher TA. Flow cytometric analysis of the granulocyte respiratory burst: a comparison study of fluorescent probes. *J Immunol Methods*. 1995;178(1):89–97.
10. Darzynkiewicz Z, Halicka HD, Zhao H. Analysis of cellular DNA content by flow and laser scanning cytometry. *Adv Exp Med Biol*. 2010;676: 137–147.
11. Loken MR, Brosnan JM, Bach BA, Ault KA. Establishing optimal lymphocyte gates for immunophenotyping by flow cytometry. *Cytometry*. 1990;11(4):453–459.
12. Gratama J, Kraan J, Keeney M, Mandy M, Sutherland D, Wood B. Enumeration of immunologically defined cell populations by flow cytometry. *Clinical and Laboratory Standards Institute Approved Guideline*. 2007;27:26–28.
13. Roederer M, Moody MA. Polychromatic plots: graphical display of multidimensional data. *Cytometry A*. 2008;73(9):868–874.
14. Roederer M. Spectral compensation for flow cytometry: visualization artifacts, limitations, and caveats. *Cytometry*. 2001;45(3):194–205.
15. Owens MA, Vall HG, Hurley AA, Wormsley SB. Validation and quality control of immunophenotyping in clinical flow cytometry. *J Immunol Methods*. 2000;243(1-2):33–50.
16. Shearer WT, Rosenblatt HM, Gelman RS, Oyomopito R, Plaeger S, Stiehm ER, et al. Lymphocyte subsets in healthy children from birth through 18 years of age: the Pediatric AIDS Clinical Trials Group P1009 study. *J Allergy Clin Immunol*. 2003;112(5):973–980.
17. Yarchoan R, Venzon DJ, Pluda JM, Lietzau J, Wyvill KM, Tsiatis AA, et al. CD4 count and the risk for death in patients infected with HIV receiving antiretroviral therapy. *Ann Intern Med*. 1991;115(3):184–189.
18. Sekhsaria S, Fleisher TA, Vowells S, Brown M, Miller J, Gordon I, et al. Granulocyte colony-stimulating factor recruitment of CD34+ progenitors to peripheral blood: impaired mobilization in chronic granulomatous disease and adenosine deaminase--deficient severe combined immunodeficiency disease patients. *Blood*. 1996;88(3):1104–1112.
19. Delmonte OM, Fleisher TA. Flow cytometry: Surface markers and beyond. *J Allergy Clin Immunol*. 2019;143(2):528–537.
20. Pilch H, Höhn H, Freitag K, Neukirch C, Necker A, Haddad P, et al. Improved assessment of T-cell receptor (TCR) VB repertoire in clinical specimens: combination of TCR-CDR3 spectratyping with flow cytometry-based TCR VB frequency analysis. *Clin Diagn Lab Immunol*. 2002;9(2):257–266.
21. Sallusto F, Geginat J, Lanzavecchia A. Central memory and effector memory T cell subsets: function, generation, and maintenance. *Annu Rev Immunol*. 2004;22:745–763.
22. Avery DT, Ellyard JI, Mackay F, Corcoran LM, Hodgkin PD, Tangye SG. Increased expression of CD27 on activated human memory B cells correlates with their commitment to the plasma cell lineage. *J Immunol*. 2005;174(7):4034–4042.
23. Marcenaro S, Gallo F, Martini S, Santoro A, Griffiths GM, Aricó M, et al. Analysis of natural killer-cell function in familial hemophagocytic lymphohistiocytosis (FHL): defective CD107a surface expression heralds Munc13-4 defect and discriminates between genetic subtypes of the disease. *Blood*. 2006;108(7):2316–2323.
24. Rabinovitch PS, June CH, Kavanagh TJ. Introduction to functional cell assays. *Ann N Y Acad Sci*. 1993;677:252–264.
25. Mardiney M, Brown MR, Fleisher TA. Measurement of T-cell CD69 expression: a rapid and efficient means to assess mitogen- or antigen-induced proliferative capacity in normals. *Cytometry*. 1996;26(4):305–310.
26. Allsopp CE, Langhorne J. Assessing antigen-specific proliferation and cytokine responses using flow cytometry. *Methods Mol Med*. 2002;72: 409–421.
27. Slezak SE, Horan PK. Cell-mediated cytotoxicity. A highly sensitive and informative flow cytometric assay. *J Immunol Methods*. 1989;117(2): 205–214.
28. Yu Y, Arora A, Min W, Roifman CM, Grunebaum E. EdU incorporation is an alternative non-radioactive assay to [(3)H]thymidine uptake for in vitro measurement of mice T-cell proliferations. *J Immunol Methods*. 2009;350(1-2):29–35.
29. Fleisher TA, Dorman SE, Anderson JA, Vail M, Brown MR, Holland SM. Detection of intracellular phosphorylated STAT-1 by flow cytometry. *Clin Immunol*. 1999;90(3):425–430.
30. Kuhns DB, Alvord WG, Heller T, Feld JJ, Pike KM, Marciano BE, et al. Residual NADPH oxidase and survival in chronic granulomatous disease. *N Engl J Med*. 2010;363(27):2600–2610.
31. Foster B, Prussin C, Liu F, Whitmire JK, Whitton JL. Detection of intracellular cytokines by flow cytometry. *Curr Protoc Immunol*. 2007 Chapter 6:Unit 6.24.
32. Shankey TV, Rabinovitch PS, Bagwell B, Bauer KD, Duque RE, Hedley DW, et al. Guidelines for implementation of clinical DNA cytometry. *International Society for Analytical Cytology. Cytometry*. 1993;14(5): 472–477.
33. Rothaeusler K, Baumgarth N. Assessment of cell proliferation by 5-bromodeoxyuridine (BrdU) labeling for multicolor flow cytometry. *Curr Protoc Cytom*. 2007 Chapter 7:Unit7.31.
34. Telford WG. Multiparametric Analysis of Apoptosis by Flow Cytometry. *Methods Mol Biol*. 2018;1678:167–202.
35. Altman JD, Moss PA, Goulder PJ, Barouch DH, McHeyzer-Williams MG, Bell JI, et al. Phenotypic analysis of antigen-specific T lymphocytes. *Science*. 1996;274(5284):94–96.
36. Kong YY, Kwok WW. Identification of Human Antigen-Specific CD4+ T-Cells with Peptide-MHC Multimer. *Methods Mol Biol*. 2019;1988: 375–386.
37. Kern F, LiPira G, Gratama JW, Manca F, Roederer M. Measuring Ag-specific immune responses: understanding immunopathogenesis and improving diagnostics in infectious disease, autoimmunity and cancer. *Trends Immunol*. 2005;26(9):477–484.

94

Assessment of Functional Immune Responses in Lymphocytes

Roshini Sarah Abraham

The immune response is mediated by a complex network of cells, soluble and membrane-bound biological mediators, and receptors interacting within the context of specific tissues and organs to protect against pathogens. T cells recognize antigen primarily in the context of antigen-specific T-cell receptors (TCRs) and specific peptides presented by molecules of the major histocompatibility complex (MHC), either class I or class II (Chapters 5 and 6). T cells are also capable of responding nonspecifically to polyclonal stimuli, such as mitogens (in vitro) or superantigens (in vivo) that initiate a broader, non-antigen-specific proliferative response. T cells upon activation also produce cytokines that are crucial to their effector functions including activating B cells and inducing antibody production, promoting differentiation of cytotoxic T cells, activating macrophages, and promoting activation and migration of inflammatory cells. Cytokines play a role in the induction of the T-cell response when produced by antigen-presenting cells (APCs) at the time of antigen recognition. B cells recognize antigen directly via the B-cell receptor (BCR; involving surface immunoglobulin) and produce antibodies with help from T cells (T-dependent antigens) or the innate immune system (T-independent antigens). More recently, the granularity of the antibody response has been further delineated, and T-dependent antibody responses are divided into type 1, with help provided by T-follicular helper cells (Tfh), and type 2, with help provided by NKTFh cells (derived from invariant NKT cells [iNKT] expressing the Vα24 Jα18 antigen receptor). Similarly, T-independent responses are classified into three groups: type 1, induced by recognition of microbial antigens by the Toll-like receptors (TLRs) in the absence of Bruton tyrosine kinase (BTK); type 2, which requires the presence of BTK; and type 3, which has neutrophil B-helper cells.[1–3] The antibody responses to vaccination depend on a variety of factors,[4] including the mobilization of particular B-cell subsets in response to specific antigens, and subsequent antibody production. B1 B cells are considered a unique population of B cells enriched in certain secondary lymphoid tissue, which are capable of self-renewal and producing natural antibodies, specifically immunoglobulin M (IgM), and priming T cells.[5]

Regulatory (Treg) cells have been shown to play a critical role in the control of both physiological and pathological immune responses in a variety of contexts (Chapter 13). Treg cells exert a direct inhibitory effect on the development of autoimmune disease because their absence leads to the development of a severe phenotype, as manifested by the FOXP3 deficiency, IPEX (immune dysfunction/polyendocrinopathy/enteropathy/X-linked) (Chapter 34).[6] Treg cells can suppress the proliferation of antigen-stimulated naïve T cells, as demonstrated by in vitro studies, and induction of FOXP3 in conventional naïve T cells imparts suppressive function in vitro and in vivo, resulting in the generation of induced-regulatory (iTreg) T cells.[7] Besides producing antibodies to neutralize pathogens as well as presenting antigen to T cells, B cells also exert immunomodulatory control of the immune response, primarily via interleukin (IL)-10 production. IL-10-producing B cells, classified as regulatory B (Breg) cells, play an important role in immune homeostasis and protection against autoimmune responses and inflammatory damage.[8] Breg cells, in addition to producing IL-10, secrete other cytokines that act on other effector T-cell subsets, Tregs, APCs such as dendritic cells (DCs), and macrophages.

NK cells are considered to be innate immune effector cells, but they straddle the threshold of innate and adaptive immunity. NK cells are directly involved in cytotoxicity and cytokine secretion upon activation, but they also function indirectly by regulating APC and the effector T cells response. NK cell activation is controlled by synergistic signals from activating and inhibitory receptors.[9] The characteristics, responses, and assays to measure the activity and function of each of these subsets are described in detail in the sections below.

T-CELL RESPONSE

T cells form the cellular arm of the adaptive immune response, and each T cell has unique specificity derived from the presence of a functional, antigen-recognizing receptor (TCR) on the cell surface. On the surface of the T cell, the TCR associates with the CD3 complex, composed of four distinct subunits (γ, δ, ε, and ζ); the cytosolic components of the CD3 complex are involved in the intracellular propagation of signals after ligation of the TCR. Besides the CD3 complex, the TCR also clusters with a CD4 or CD8 coreceptor, depending on the type of T cell; CD4 T cells recognize antigen (Ag) in the context of MHC class II molecules, whereas CD8 T cells recognize antigen presented on MHC class I. The first signal or "cognate" signal, which is recognition of peptide–MHC complex by the TCR, results in actin-mediated reorganization of the cytoskeleton in both the T cell and APC to form the immunological synapse (Fig. 94.1). The synapse consists of the TCR along with other costimulatory molecules and adhesion receptors, which results in the formation of a large macromolecular structure called the supramolecular activation complex (SMAC). The unique organization of the SMAC results in prolonged and stronger intracellular interactions and appropriate downstream signaling activity, including phosphorylation of the CD3 receptor components. The complete culmination of the TCR-induced signals results in IL-2 production and secretion, which drives T-cell proliferation in an autocrine and paracrine manner.

It is important to recognize that various T-cell processes and stages of differentiation are specifically regulated by metabolic

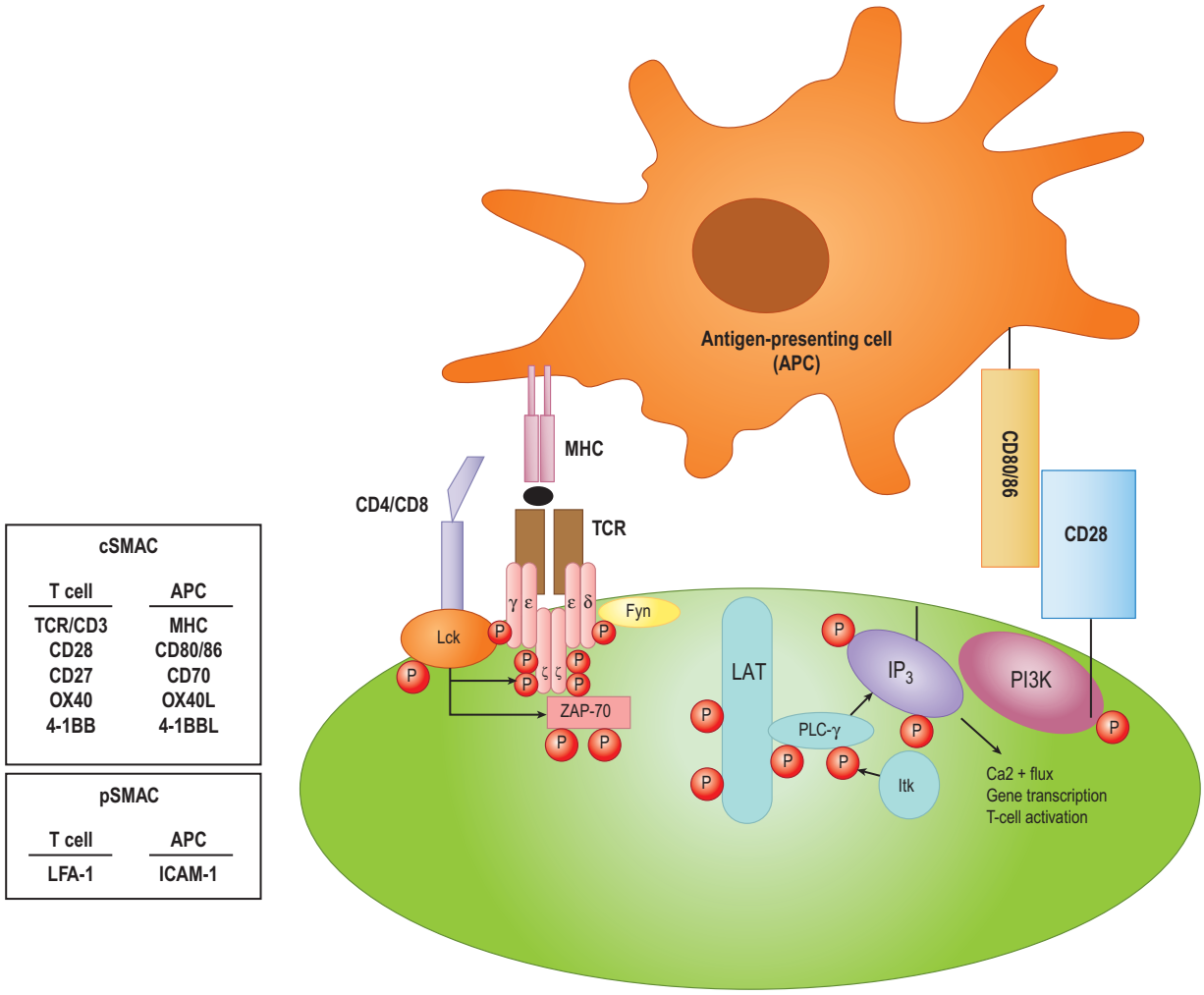


FIG. 94.1 Schematic of T-Cell Activation. T cells recognize antigen in the context of major histocompatibility complex (*MHC*) molecules on antigen-presenting cells (*APCs*). Sustained interaction between a T cell and APC results in the formation of the immunological synapse with membrane reorganization and ordering of key molecules on both cells into a supramolecular activation complex (SMAC) composed of c (central) *cSMAC* and p (peripheral) *pSMAC*. Cognate interaction with appropriate costimulatory signals results in intracellular signaling events that ultimately result in T-cell activation. *LAT*, linker for activation of T cells; *PLC-γ*, phospholipase C-gamma; *TCR*, T-cell receptor; *IP_3*, inositol 1, 4, 5 triphosphate; *P*, phosphorylation; *PI3K*, phosphatidylinositol 3′-hydroxyl kinase; *Lck* and *Fyn* are Src-family nonreceptor protein tyrosine kinases; *Itk*, IL-2–inducible T-cell kinase; *ZAP-70*, ζ-chain–associated protein kinase 70. (Adapted from Fig. 1, Pennock ND, et al. T cell responses: naive to memory and everything in between. *Adv Physiol Educ*. 2013;37:273–83.

pathways.[10,11] For example, naïve T cells rely on oxidative phosphorylation for their metabolism, with additional glycolytic metabolism during times of proliferation. On the other hand, activated effector T cells demonstrate aerobic glycolysis and increased oxidative phosphorylation. T-helper cell-1 (Th1), Th2, and Th17 CD4 effector T cells use glycolytic metabolism, whereas regulatory T-cell and memory T-cell development is enhanced by fatty acid oxidation and catabolic metabolism. In particular, memory T cells are quiescent and mainly use oxidative phosphorylation; however, on antigenic rechallenge, the use of oxidative phosphorylation and glycolysis is rapidly facilitated.[10]

CLINICAL RELEVANCE

Lymphocyte Function Evaluation

- Assessment of lymphocyte function typically refers to the measurement of T-cell function via cellular proliferation to nonspecific and specific stimuli; however, it is much broader and comprehensive, and refers to any method that assesses any of the various effector or regulatory aspects of any lymphocyte subset.
- The robustness of clinical interpretation of lymphocyte function assays is dependent on the analytical procedure, sample type, timing of collection, and clinical context, among other factors. Assays performed in clinical diagnostic immunology laboratories are standardized to minimize inconsistencies in these variables.

MEASUREMENT OF T-CELL FUNCTION VIA ACTIVATION MARKERS

Activation of T cells results in the expression of several induced markers, which can be used to ascertain the competence of T cells participating in the immune response. These include CD69, CD154 (CD40L), MHC class II, and CD25, which are expressed sequentially and can be assessed by flow cytometry.[12] In contrast to in vivo activation of T cells and subsequent expression of these

markers, in vitro assessment of T-cell function involves activation of T cells with nonspecific and polyclonal stimulants such as phorbol myristate acetate (PMA) or phytohemagglutinin (PHA), which results in expression of these markers in a kinetically regulated fashion.[12] The expression of CD40L on activated T cells, besides being useful as a global marker for T-cell activation, is used more specifically for the diagnosis of a primary immunodeficiency, X-linked hyper-IgM syndrome or CD40L deficiency (Chapter 34). CD40L on activated CD4 T cells interacts with CD40, expressed constitutively on B cells, and participates in isotype (class)-switching as well as providing costimulatory help to T cells. The expression of CD40L on activated CD4 T cells can easily be ascertained in the laboratory using an in vitro T-cell stimulation protocol. In this assay, CD69 is used as a control establishing early T-cell activation. The majority of patients (~80%) with mutations in *CD40LG* have absent protein expression on activated CD4 T cells; however, ~20% of mutations may remain permissive for protein expression, but with aberrant function. These patients can be identified by incorporating an additional component in the assay using a soluble form of the receptor CD40-muIg to assess binding of the ligand. The flow cytometric assay for CD40L expression and function offers a rapid diagnostic test for XL-hyper-IgM syndrome.[12]

ASSESSMENT OF CELLULAR VIABILITY IN LYMPHOCYTES

The flow cytometry–based assays also allow determination of viable cells in the sample, which is particularly important in the interpretation of results after peripheral blood mononuclear cell (PBMC) isolation from anticoagulated (heparinized) whole blood. The use of annexin V in a flow cytometry assay enables the visualization of apoptotic cells. In addition to visualizing apoptotic cells, dead cells can be assessed by simultaneously using 7-amino-actinomycin-D (7-AAD) (Chapter 93), which is a membrane-impermeable dye excluded from viable cells. When internalized into dying or dead cells, it binds to double-stranded DNA by intercalating between base pairs in guanine (G)-cytosine (C)-rich regions. These two dyes can be combined to provide information on the proportion of viable cells in the starting cell mixture for proliferation assays.

Measurement of T-Cell Competence Via Proliferation

Mitogens are very potent stimulators of T-cell activation and induce polyclonal T-cell proliferation. It has been suggested that mitogens can induce T-cell proliferative responses even if the lymphocytes are incapable of responding adequately to antigenic (physiological) stimuli. Therefore, abnormal T-cell responses to mitogens are considered a diagnostically less sensitive but more specific test of aberrant T-cell function. Lectin mitogens have been shown to bind to the TCR, thereby activating quiescent T cells. Mitogenic stimulation induces increased intracellular calcium (Ca^{2+}) in T cells, which is essential for T-cell proliferation. Whereas PHA is a strong T-cell mitogen (Fig. 94.2), pokeweed mitogen (PWM) is a weak T-cell mitogen that also induces B-cell activation and proliferation with a different timeline for maximal stimulation. Mitogens such as PHA activate T cells by binding to cell membrane glycoproteins, including the TCR-CD3 complex. In addition, there are many mitogenic or comitogenic antibodies, including those directed against the CD3 coreceptor that can stimulate T-cell proliferation. Typically, anti-CD3 antibodies provide an initial activation signal and provide a variable proliferative response. Addition of a costimulatory antibody (anti-CD28) to anti-CD3 results in enhanced proliferation. An exogenous T-cell growth factor, such as IL-2, can also be used as an alternate to anti-CD28 costimulation, and in patients with suspected IL-2 receptor (IL-2R)–associated signaling defects, it may be more helpful from a diagnostic perspective than the use of anti-CD28. IL-2, an autocrine cytokine, has been demonstrated to be critical in T-cell proliferation and the regulation of T-cell growth through binding to a heterotrimeric receptor complex consisting of three chains—α, β, and γ (IL-2Rα, IL-2Rβ, and IL-2Rγ)—on the surface of T cells. Triggering of the TCR leads to the synthesis of IL-2 in certain T-cell subsets with induction of high-affinity IL-2Rs on antigen- or mitogen-activated T cells, the binding of IL-2 to the IL-2R ultimately leads to T-cell proliferation. The use of exogenous IL-2 in association with anti-CD3 allows discrimination of T cells that cannot proliferate to other mitogenic signals but can respond to a potent growth factor such as IL-2.

Antigens, including *candida* antigen (CA) and tetanus toxoid (TT), have been widely used to measure antigen-specific recall (anamnestic) T-cell responses when assessing cellular immunity. This may be more revealing about cellular immune compromise than assessing the response of lymphocytes to mitogens because the latter can induce T-cell proliferative responses even if those T cells are incapable of responding adequately to antigenic (physiological) stimuli. Therefore, abnormal T-cell responses to antigens are considered a diagnostically more sensitive, but less specific, test of aberrant T-cell function. Antigens used in recall assays measure the ability of T cells bearing specific TCRs to respond to antigenic peptides presented by APCs. The antigens used for assessment of the cellular immune response are selected to represent antigens, seen by a majority of the population either through natural exposure (CA) or because of vaccination (TT) using a longer in vitro culture period (6 to 7 days) before assessing the response.

In addition to measuring cellular proliferation as a readout for T-cell function, the production of cytokines by activated T cells is another important component for evaluating T-cell functional activity. T cells typically are not monofunctional, producing a single cytokine on activation; rather, they are multi- or polyfunctional, and the range of cytokines are produced sequentially rather than simultaneously. Although a population of stimulated T cells will have individual T cells

KEY CONCEPTS

T-Cell Activation and Function

- Cognate recognition of peptide major histocompatibility complex (MHC) on the antigen-presenting cells (APCs) by the T-cell receptor results in formation of the immunological synapse.
- Activated T cells express early activation markers, such as CD69 and CD40L. Other T-cell activation markers include CD25 and MHC class II (human leukocyte antigen–D related [HLA-DR]) molecules.
- T cells can be stimulated to proliferate using nonspecific stimulants such as plant lectins (mitogens) and cross-linking of the CD3 coreceptor, along with other costimulatory molecules (e.g., anti-CD28) or in the presence of exogenous interleukin-2 (IL-2).
- The magnitude of antigen-specific T-cell proliferation is dependent on the starting frequency of antigen-specific T cells.
- Flow cytometry–based assays of T-cell proliferation offer the advantages of high resolution, accounting for cellular dilution due to T-cell lymphopenia and single-cell analysis.

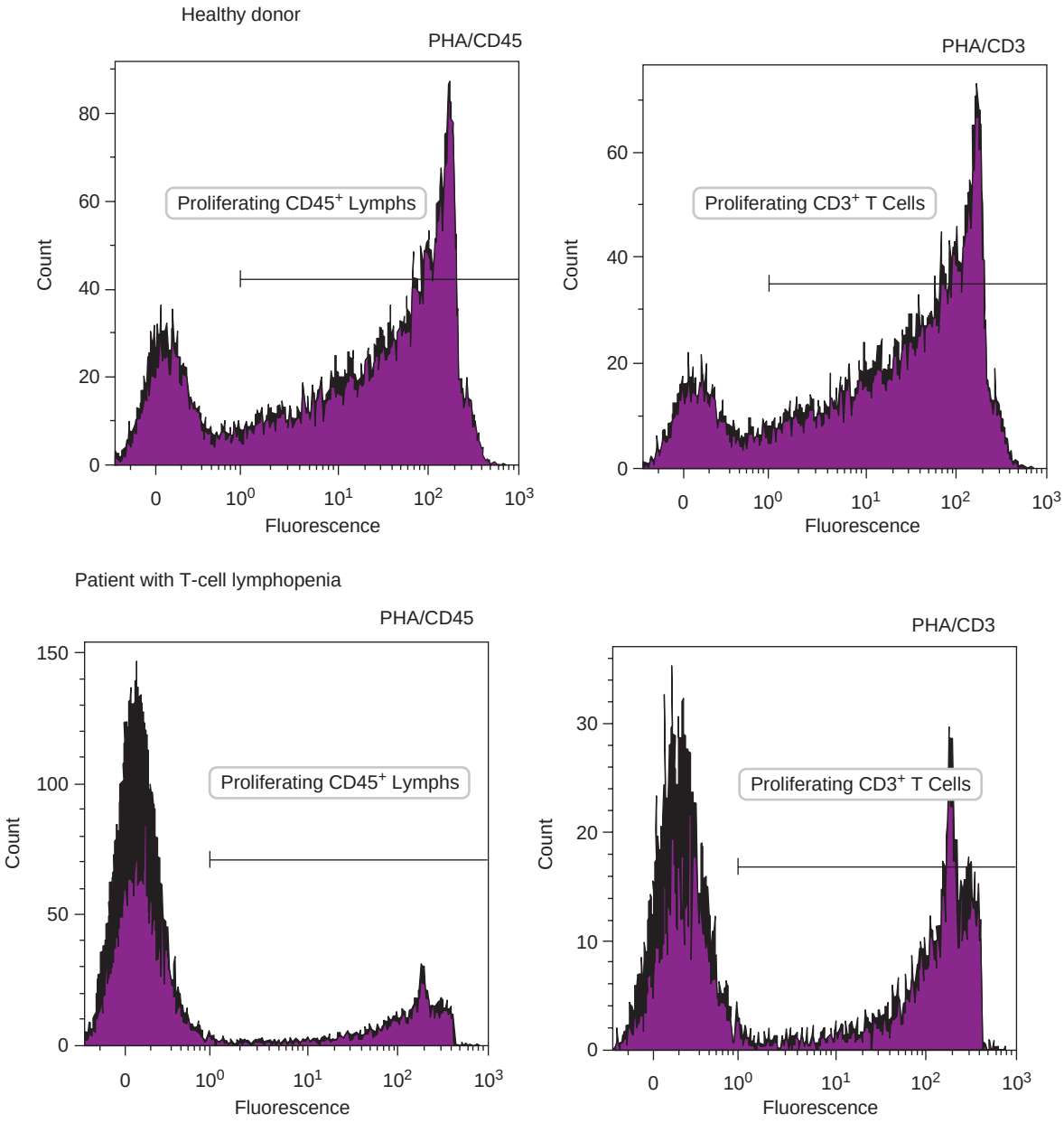


FIG. 94.2 T-Cell Proliferation to Mitogenic Stimulants. T cells can respond polyclonally to in vitro stimulation with various mitogens. T-cell proliferation assessment by flow cytometry allows analysis in total lymphocytes (*CD45*$^+$) and T cells (*CD3*$^+$). The *top panel* demonstrates T-cell proliferation to phytohemagglutinin (*PHA*) in both cell subsets. In patients with T-cell lymphopenia, where cellular dilution may be a concern, single-cell analysis by flow cytometry allows discrimination of functional versus nonfunctional T cells. In the *lower panel*, a patient with T-cell lymphopenia has abnormal proliferation when total lymphocytes are assessed; however, the response is normal when the T-cell compartment is specifically evaluated. Therefore, this patient, who would have been classified as abnormal based only on the CD45$^+$ lymphocyte response, can be reclassified as having normal T-cell proliferation based on the CD3$^+$ T-cell response, but with significant T-cell lymphopenia.

producing different cytokines in a temporally regulated manner. These cytokines can be measured by intracellular flow cytometry after T-cell activation with mitogens, in both CD4 and CD8 T-cell subsets, or in the culture supernatant of activated cells. Typically, the cytokines measured in in vitro stimulation assays include IL-2, interferon (IFN)-γ, IL-4, IL-5, IL-6, IL-17, and tumor necrosis factor (TNF)-α.

The commonly used method for assessing T-cell proliferation for decades involved the measurement of ^{3}H-thymidine (^{3}H-T) incorporated into the DNA of proliferating cells. The results are expressed as either counts per minute (cpm) or disintegrations per minute (dpm) of both activated and nonactivated cells (background) cultured for a fixed period of time, usually 72 hours for mitogens. A reference range based on proliferation results from a group of control subjects should be provided along with the patients' results (both background and post-stimulation results). Although this method remains widely employed, several disadvantages exist, including but not limited to the use of radioactive material (^{3}H-T); its inability to discriminate responder cell subsets or to account for cellular

dilution, which is particularly relevant in the context of T-cell lymphopenia; and the lack of information on the contribution of cell death after stimulation and its impact on the final result. To overcome the intrinsic shortcomings of the ^{3}H-T method, newer flow cytometry assays are currently being used in the clinical diagnostic setting and are gaining popularity because of the additional information they provide.

The flow cytometric methods for measuring cell proliferation include the use of fluorescent dyes to identify proliferating cells. One of the more commonly used dyes, carboxyfluorescein diacetate succinimidyl ester (CFSE), must be carefully used when measuring lymphocyte proliferation in vitro since it can be toxic to cells and nonoptimal labeling conditions can influence the measurement of cell proliferation and interpretation of results. CFSE is a fluorescent cell-membrane permeable dye similar in physical properties to the commonly used fluorochrome, fluorescein isothiocyanate (FITC). During cell proliferation, the intensity of staining in daughter cells is half that of parent cells, allowing visualization of the number of rounds of cell divisions associated with the successive decrease in fluorescence. The disadvantages to using CFSE in the clinical laboratory are its photo-instability, limitation in the number of cell divisions being identified (≤7 cell divisions), and interpretation that requires measuring a loss of signal rather than a gain of signal. In addition, as noted above, CFSE at concentrations of 37 nM to 10 μM can be toxic to cells, resulting in increased cell death. Furthermore, it can modulate the expression of activation markers, resulting in a decrease in CD69, human leukocyte antigen–D related (HLA-DR), and CD25 expression. Finally, it has been reported that there is an increase in the number of false-positive results with CFSE, making it suboptimal for measuring lymphocyte proliferation in patients with severe cellular immunodeficiencies. Alternatives to CFSE include related compounds, such as CellTrace Violet (CTV) and Cell Proliferation Dye eFluor 670 (CPD), which have different excitation and emission spectra compared with CFSE. These dyes also can be used for tracking lymphocyte proliferation status in vivo, in animal models, permitting measurement of up to 11 cell divisions.[1]

Another alternative to tritiated thymidine is the use of a thymidine analogue 5-ethynyl-2′-deoxyuridine (EdU), which can combine with a fluorescent azide in a copper-catalyzed cycloaddition reaction (referred to as "Click" chemistry) and permits flow cytometric evaluation of lymphocyte proliferative responses by assessing its incorporation into cellular DNA. EdU labeling is a fast and sensitive method for measuring cell proliferation and facilitates the identification of dividing cells. It is relatively more photostable than CFSE and is added to cells after completion of the stimulation period with the measurement being comparable to the thymidine method in regards to a gain of signal as the endpoint. The EdU assay has not shown the limitations of the CFSE method in the clinical laboratory and has been used to evaluate lymphocyte proliferation in a spectrum of patients, including those with severe combined immune deficiency (SCID). In fact, this assay has been particularly useful in discriminating between functional T cells and nonfunctional T cells in the context of severe T-cell lymphopenia, which cannot be achieved with the standard thymidine assay. All the proliferation data shown in this chapter utilize this EdU-based measurement of T-cell proliferation, and results are typically provided for both CD45 bright positive (total lymphocytes) and CD3 T cells, with the former being more representative of the data generated with the standard thymidine assay.* In recent years, there has been an explosion of a variety of kits meant to evaluate T-cell activation and proliferation. Each of these has pros and cons, and it is beyond the scope of this chapter to delve into the details of these kits. Very recently, an alternative flow cytometric method to assess T-cell proliferation was described, which uses STAT5A phosphorylation as the read-out for this specific cellular function.[13]

[1]This section have been reproduced in part with permission from ASM Press (Abraham RS, Lymphocyte Activation) (copyright permission obtained for 5th edition).

MEASUREMENT OF CELL-MEDIATED CYTOTOXICITY

CD8 T cells are considered the representative cytotoxic T-cell in the immune system, and cellular cytotoxicity is a mechanism to eliminate cells infected with intracellular pathogens, allogeneic cells, or tumor cells. CD8 T cells, like CD4 T cells, recognize antigen via the TCR and kill target cells via cytotoxic protein granule exocytosis and/or cytokine production. Over the past few decades, the cytotoxic potential of CD4 T cells has been described, particularly in viral infections. Although cytotoxic CD4 T cells are rare in the circulation of healthy individuals (<2%), they can account for substantial proportions of total CD4 T cells in certain viral infections, including but not limited to human immunodeficiency virus (HIV).

KEY CONCEPTS

Cellular Cytotoxicity

- CD8 T cells and natural killer (NK) cells are involved in the killing of cellular targets and contain intracellular granules with cytotoxic proteins, such as perforin and granzymes.
- Cytotoxic CD4 T cells are a subset of memory T cells with cytolytic potential and are usually observed in circulation in the context of chronic viral infections, e.g., cytomegalovirus (CMV), human immunodeficiency virus (HIV).
- Tetramer-based assays have been used to quantify and delineate the function of antigen-specific CD8 T cells (also CD4 T cells).
- Cellular degranulation results in the expression of CD107a on the cell surface of CD8 T cells and NK cells and is often used as a surrogate of cytotoxic activity.
- NK cells recognize target cells that lack major histocompatibility complex (MHC) class I molecules, e.g., viral cells or tumor cells that have downregulated MHC class I.
- The majority of circulating mature NK cells are cytotoxic ($CD3^{-}CD16^{++}CD56^{+/-}$), while a minority are immature cytokine-producing ($CD3^{-}CD56^{++}$).
- Interleukin (IL)-2 augments NK cell-mediated cytotoxicity (lymphokine-activated killing) and induces interferon (IFN)- γ secretion by NK cells.
- Regulatory T cells control NK cell activation and cytotoxic function by limiting the availability of IL-2.

Cellular cytotoxic activity has been conventionally measured by release of chromium (^{51}Cr) from labeled target (T) cells cultured with various ratios of effector (E) cells (varied E:T ratio). Alternative assays have been developed, including flow cytometry and ELISPOT-based methods. More recently, a method of assessing human CD8 T-cell cytotoxicity has been described

in which both the effector and target cells are primary cells (in contrast to using cell lines as target cells).[14] In this particular assay, autologous B cells isolated from PBMC are used as target cells and cocultured with antigen-specific CD8 effector cells. The killing of the fluorescently labeled target B cells (see NK cell section) is used to estimate cytotoxic activity. Antigen-specific effector cells are quantified in the total CD8 T-cell pool using MHC-peptide tetramers.[14]

A tetramer-based approach to quantifying and measuring the function of cytomegalovirus (CMV)-specific CD8 T cells has been available in the clinical diagnostic immunology laboratory for over a decade (Fig. 94.3). It has shown to be predictive in solid organ transplant patients, using an approach that includes both quantification of CMV-specific CD8 T cells along with function (CD107a expression and IFNγ production).[15] However, the limitation of the tetramer approach in the clinical diagnostic setting is that it is constrained by the number of HLA-peptide tetramer (multimer) combinations that are available for use with a particular antigen (e.g., CMV, Epstein-Barr virus [EBV]). It also requires a priori knowledge of the patient's HLA genotype, and very often it does not incorporate a comprehensive assessment of the CD8 and CD4 cytotoxic T-cell response. Although some laboratories use only a quantitative approach to determine immune competence to pathogens, such as CMV with the tetramer assay, more comprehensive assays are available that also provide a functional assessment of these antigen-specific CD8 T cells (see Fig. 94.3).

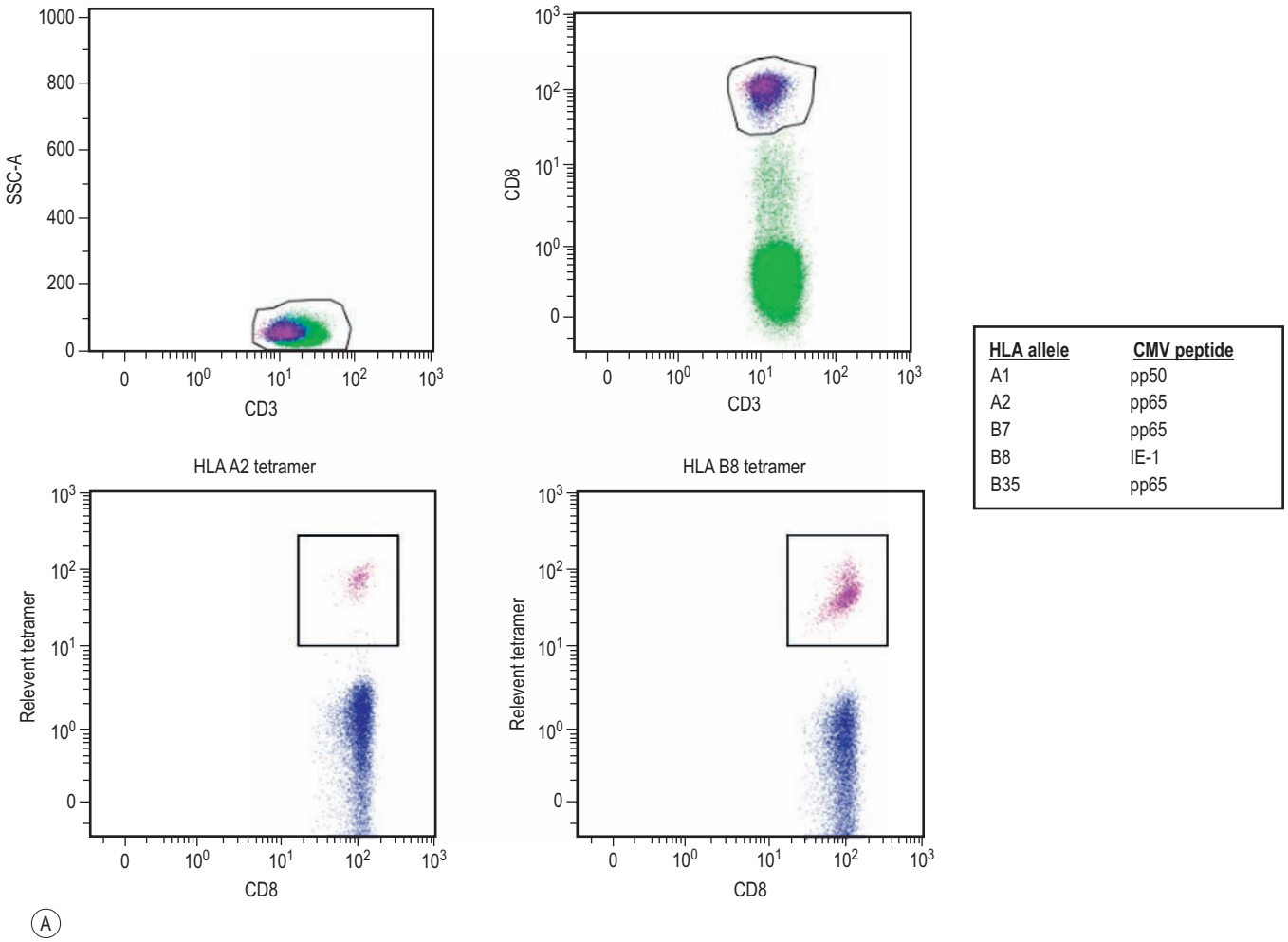


FIG. 94.3 Quantitation of Antigen-Specific CD8 T Cells and Functional Evaluation Using Major Histocompatibility Complex (MHC) Class I Tetramers. The tetramer (multimer) technology has been useful for accurate quantitation of antigen-specific CD8 (or CD4 T cells, if MHC class II tetramers are used). The *top panels* show the identification of $CD3^+$ T cells from total lymphocytes and subsequent segregation into $CD3^+CD8^+$ T cells. In this assay, cytomegalovirus (*CMV*)-specific CD8 T cells are quantitated using five MHC class I tetramers (human leukocyte antigen [*HLA*] *A1, A2, B7, B8,* and *B35*), each recognizing a unique CMV peptide from three major CMV antigenic proteins (*pp50, pp65*, and *IE-1*) as listed in the *box*. Based on the HLA class I haplotype of the individual, one or more tetramers are used for stimulation. The bottom panels show an example of a patient with HLA A2-specific CMV-CD8 T cells and another patient with HLA B8-specific CMV-CD8 T cells. These CMV-specific CD8 T cells can be assessed for functional capacity by gating on the tetramer-positive CD8 T cells and then measuring degranulation (CD107a expression) and interferon-gamma (*IFN-γ*) production after in vitro stimulation with the specific CMV peptide (*lower panel*). The data are shown for CD107a expression and IFN-γ production in unstimulated CMV-CD8 T cells and peptide-stimulated CMV-CD8 T cells.

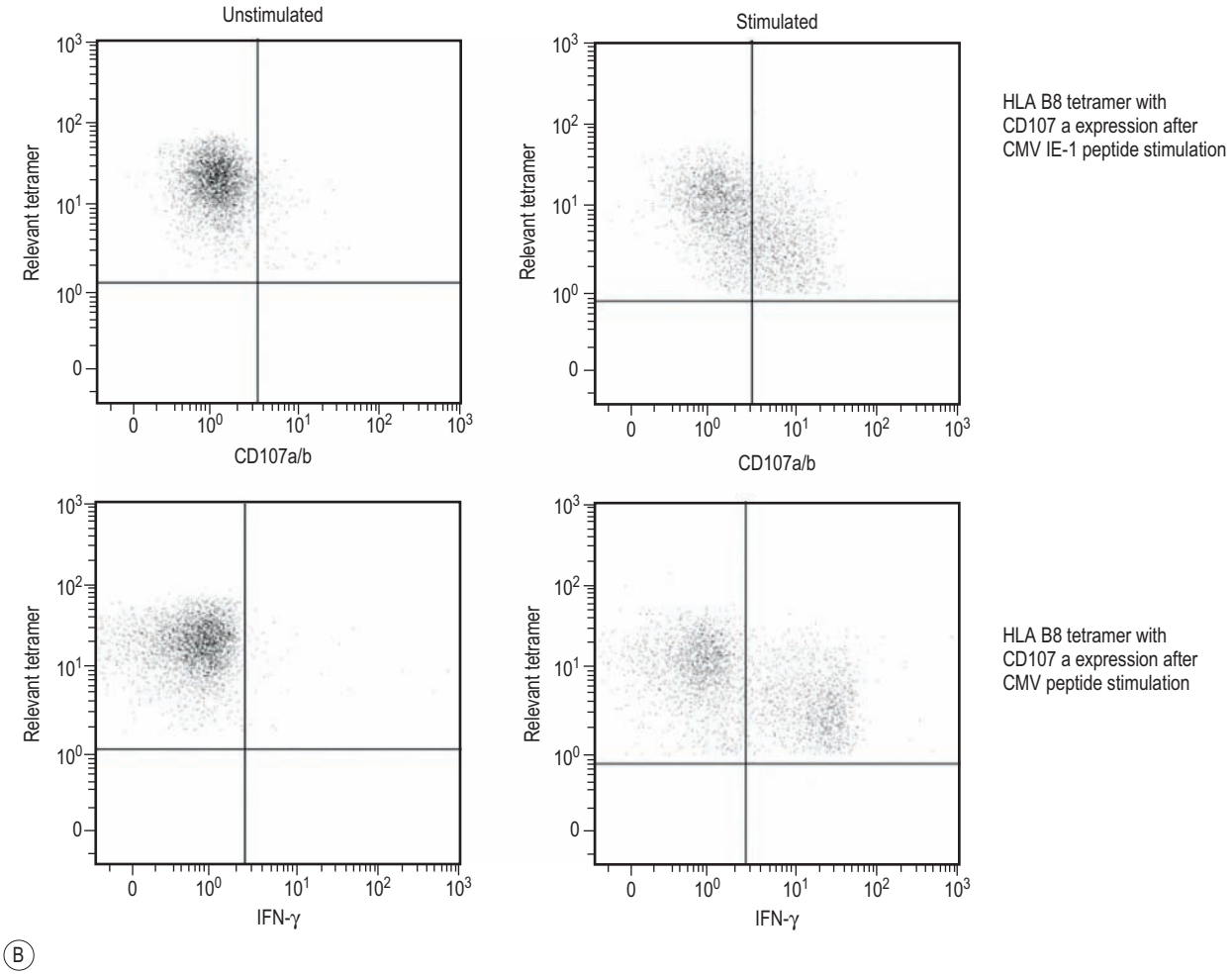


FIG. 94.3—cont'd.

An alternative approach is to evaluate for cytotoxic function, by assessing degranulation by evaluating for CD107a expression by the effector T cell (described in further detail in the NK cell function section) and/or evaluating IFN-γ production by activated cytotoxic CD8 T cells. As a positive control, CD8 T cells are polyclonally stimulated with PMA and ionomycin, and the same markers (CD107a expression and IFN-γ production) are measured. Approximately 10% to 60% of CD8 T cells in healthy adults are activated to express CD107a and to secrete IFN-γ under these conditions. An approach that avoids the use of tetramers involves the stimulation of patient-specific PBMC with overlapping peptide pools of the specific antigen and, subsequently, the evaluation of the CD8 and CD4 T cells that proliferate in response to antigenic stimulation. These T cells can also be assessed for cytotoxic potential by measuring intracellular perforin and granzyme expression as well as degranulation (CD107a expression) after stimulation.[12]

A CMV-specific assay has been developed that measures IFN-γ production in response to CD8 T-cell stimulation with a pool of MHC class I–restricted CMV peptides, the QuantiFERON-CMV (Cellestis Ltd., Melbourne, Australia).[16] This commercial assay is simpler to perform than the tetramer-based approach for quantification and functional analysis of antigen-specific CD8 T cells and does not require knowing the subjects' HLA type (Fig. 94.4). It has been applied in the context of risk prediction for CMV in organ transplant and allogeneic hematopoietic cell transplant (HCT) patients. However, this assay has theoretical limitations in that assessing the production of a single cytokine is unlikely to represent the breadth of the immune response to a complex specific antigen such as CMV.

Assessing Function of Exhausted T Cells

Previous sections have discussed the role of T-cell activation in the immune response to pathogens. An optimal immune response results in normal activation of T cells. But persistent antigenic stimulation alters the differentiation of memory T cells and can result in a state of T-cell "exhaustion" with failure of the immune response. The term "T-cell exhaustion" is broad and can have variable meanings. Nevertheless, it generally refers to effector T cells with impaired cytokine secretion and expression of molecules associated with inhibition of a response as well as expression of specific markers, such as PD-1. Exhausted T cells lose their ability to normally respond to antigenic stimulation via cytokine production and proliferation and are functionally distinct from anergic T cells. T-cell exhaustion often is represented in the context of chronic viral infection but has also been described in the setting of neoplastic disease. Exhausted T cells express a variety of cellular markers, including PD-1, CTLA4,

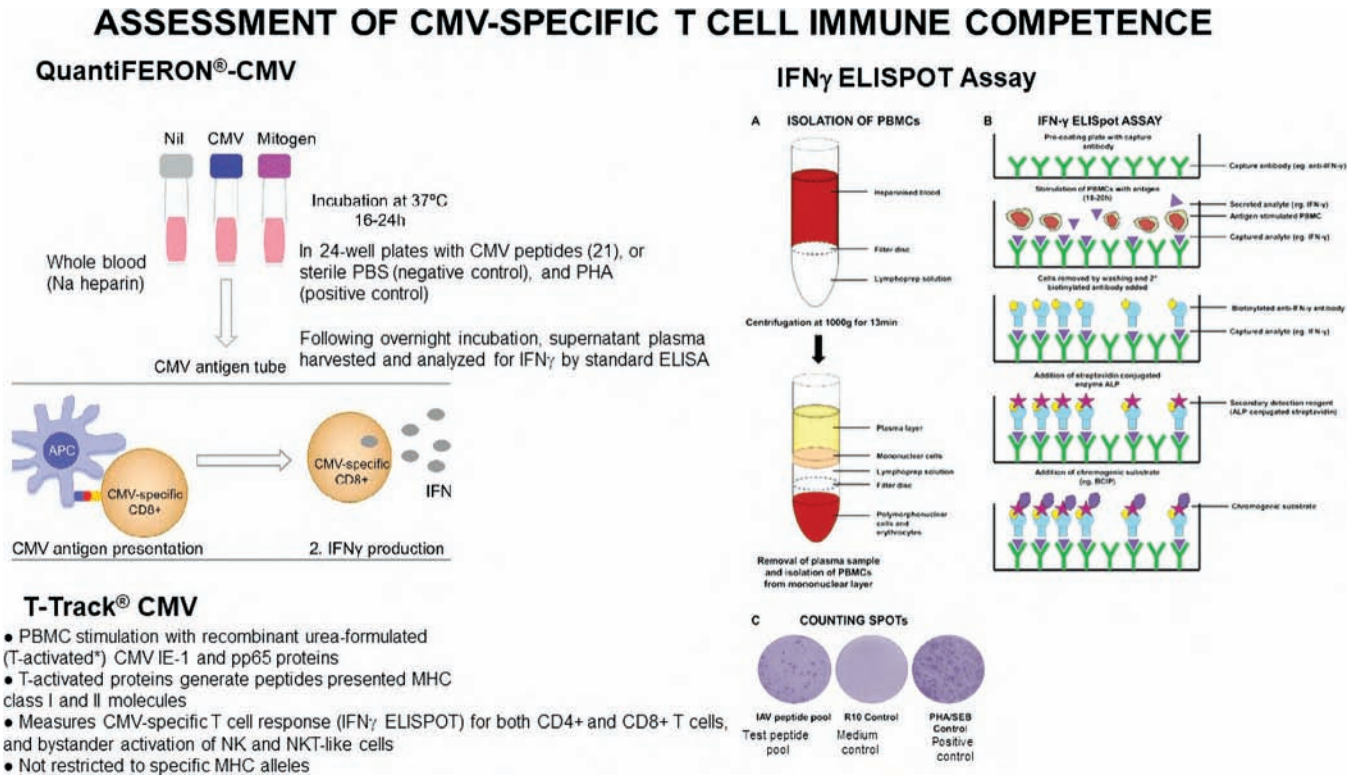


FIG. 94.4 Using Interferon-γ Secretion After Cytomegalovirus-Specific Stimulation to Measure Antigen-specific CD8 T-Cell Response. The QuantiFERON assay for assessing cytomegalovirus (*CMV*)-specific CD8 T-cell function measures interferon-γ (*IFN-γ*) produced by CD8 T cells in response to stimulation to a pool of 21 CMV peptides. The T-track assay, on the other hand, utilizes a special preparation of 2 CMV proteins (IE-1 and pp65) and measures the T (CD8, and CD4) and NK cell response using an IFN-γ ELISPOT. IFN-γ secretion assays use peripheral blood mononuclear cells (*PBMCs*), which are stimulated by CMV (or other antigen) peptide pools, with unstimulated and positive (phytohemagglutinin [*PHA*]) controls, and the secreted cytokine is detected by specific capture antibodies, and revealed by a secondary labeled antibody using either a chromogenic substrate or counting spots created on a membrane (ELISPOT).

TIM-3, and LAG-3, among others, and the pattern of expression and number of receptors can reflect the degree of T-cell impairment. Several of these markers can be assessed on T cells by flow cytometry; however, it is important to recognize that expression of these inhibitory molecules does not always imply that healthy T cells expressing these are exhausted or aberrant in function. In fact, in healthy people, the presence of circulating PD-1^{+}CD8^{+} T cells may represent effector memory T cells rather than exhausted T cells. T-cell exhaustion can be assessed in vitro using fresh human PBMCs stimulated with a superantigen, such as staphylococcal enterotoxin B (SEB) for 72 hours. The culture supernatant is harvested for cytokine analysis, specifically IL-2 and IFN-γ. Following this stimulation, the cells are washed and re-suspended into fresh medium, in the presence or absence of various therapeutic molecules, and the T-cell activation response is assessed. There is maximal production of IFN-γ and IL-2 at 72 hours post-stimulation with varying concentrations of SEB. However, the amount of cytokines produced is decreased 24 and 48 hours after withdrawal of the SEB. This assay can be used to assess the effect of therapeutic agents in vitro to reverse the state of T-cell exhaustion. This assay is not available within the clinical diagnostic laboratory for this purpose. But many laboratories have the ability to assess the expression of inhibitory receptors on T-cell subsets by flow cytometry. However, the results of these assays need to be interpreted cautiously and in context of the reasons discussed above.

T-Cell ELISPOT for Measuring T-Cell Function Through Cytokine Production

Assessing T-cell function at a single-cell level can be achieved using the ELISPOT method. The previously described CMV QuantiFERON assay (see Fig. 94.4) has been used in pediatric HCT patients and was shown to be useful for identifying patients at high risk for developing CMV viremia.[17] A different version of this ELISPOT assay, called T-Track assay[18,19] uses proprietary CMV antigens, which are urea-formulated recombinant CMV proteins, IE-1, and pp65 (called T-activated antigens), for use in an IFN-γ ELISPOT assay (see Fig. 94.4). IFN-γ ELISPOT assays have also been validated and standardized for assessing alloreactive responses in renal allograft recipients to ascertain allograft rejection.[20] While cytokine production by T cells can also be assessed by intracellular flow cytometry, the sensitivity of the ELISPOT assay, particularly at a single-cell level, appears to be higher. However, this approach requires careful standardization and validation as well as the use of appropriate tools to read and interpret the data.

NATURAL KILLER CELL ACTIVATION AND FUNCTION

Unlike T and B cells of the adaptive immune system, innate NK cells do not have antigen-specific recognition and killing

of target cells. NK cells participate directly in effector functions via cytotoxicity and production of cytokines (e.g., IFN-γ) upon activation. Proinflammatory cytokines such as IL-12, IL-15, and IL-18 trigger NK cell proliferation as well as cytotoxicity and production of IFN-γ. The negative regulation of NK cells is controlled by receptors that recognize MHC class I molecules preventing NK cell–mediated cytotoxicity. In contrast, virally infected or tumor cells typically downregulate MHC class I, making them appropriate targets of NK cell–mediated cytotoxicity in the presence of relevant ligands expressed by the target cell. NK cells, like cytotoxic T cells, contain granules with cytotoxic proteins including perforin and granzymes, which are serine proteases that recognize different substrates.

NATURAL KILLER CELL CYTOTOXICITY

The majority of NK cells can be identified by a lack of CD3 on the cell surface in conjunction with the expression of CD56 (NCAM [neural cell adhesion molecule]) and CD16 (FcγRIII). NK cells can be subdivided into two major subsets based on their relative expression of CD56 and CD16: $CD16^{+++(bright)}$ $CD56^{+/-(dim)}$ NK cells, referred to as cytotoxic (mature) NK cells, and $CD56^{+++(bright)}$ CD16 or $CD16^{+/-}$ NK cells, known as regulatory or cytokine-producing (immature) NK cells. The majority (~90%) of circulating human NK cells belong to the cytotoxic category, while a minority (10%) represent cytokine-producing NK cells. Cytotoxicity can be subdivided into natural or spontaneous cytotoxicity directed largely toward virally infected cells or tumor cells, in the absence of prior stimulation or immunization, and antibody-dependent cellular cytotoxicity (ADCC) directed against antibody-coated target cells. NK cells go through a process of education or "licensing" whereby NK cells that express inhibitory receptors to self-MHC class I molecules are called licensed, which means they are more functionally responsive to stimulation, whereas unlicensed NK cells lack receptors for self-MHC class I and are hyporesponsive (Fig. 94.5).

NK cell function is measured in the clinical laboratory by assessment of spontaneous (natural) NK cell cytotoxicity using an MHC class I–deficient myelogenous leukemia cell line, K562. Traditional methods for measuring NK cell cytotoxicity are similar to those used for assessing cytotoxic T-cell (CTL) function based on varying effector:target cell ratios in a 4- to 16-hour ^{51}Cr-release assay compared with the no-lysis and 100% lysis conditions as described for the T-cell cytotoxicity assay. However, there is interest in the use of flow cytometry to measure spontaneous or IL-2–activated (lymphokine-activated killer [LAK]) NK cell cytotoxicity in the clinical laboratory to avoid using radionuclides (Figs. 94.6A and Fig. 94.6B). One flow cytometric assay employed in the clinical laboratory involves the use of fluorescently labeled (CellTracker dyes) target cells (K562) incubated with effector cells (donor or patient PBMC) in the absence (spontaneous) or presence of IL-2 (LAK). Following co-incubation, the lysis of target cells is measured by using 7-AAD to assess for cell death (see Figs. 94.6A and Fig. 94.6B). IL-2 enhances cytotoxic function with increased lytic potential against a broad range of target cells. IL-2 has also been shown to induce IFN-γ secretion by NK cells with upregulation of activation markers, such as CD25 and CD69. Treg cells control NK cell activation and cytotoxic function by limiting access to IL-2. Other methods for measuring NK cell cytotoxic function that are primarily used in the research setting include the use of

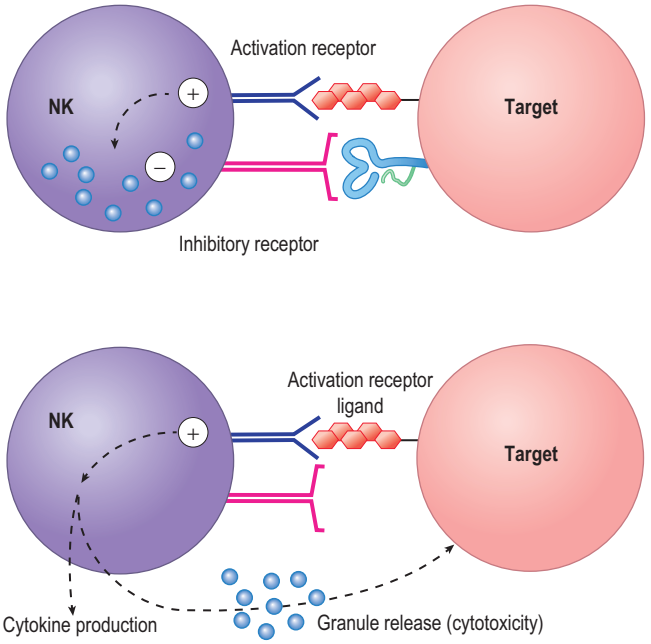


FIG. 94.5 Natural Killer Cell Recognition of Target for Cytotoxic Function. Natural killer (*NK*) cells have inhibitory receptors for self-major histocompatibility complex (MHC) class I. Therefore, NK cells do not kill target cells expressing MHC class I *(upper panel)*. However, when MHC class I expression is downregulated (viral infections, tumors), the target cell is primed for NK cell cytotoxicity, which is mediated by granule exocytosis and release of cytotoxic proteins, including perforin and granzymes *(lower panel)*. (Adapted from figure in French AR, Yokoyama WM. Natural killer cells and autoimmunity. *Arthritis Res Ther.* 2004; 6:8–14.)

image cytometry, microchip screening, and flow-based assays using other dyes, such as calcein AM.

NK cell cytotoxicity assays exhibit significant biological and analytical variability, and, therefore, are of limited utility, especially on shipped samples. Other parameters are widely used as surrogates of cytotoxicity, including flow cytometric measurement of granule exocytosis/degranulation. The membrane of cytotoxic granules in both NK cells and CD8 T cells is composed of several proteins, including CD107a (lysosomal-associated membrane protein 1 [LAMP-1]). On stimulation of NK cells and CD8 cytotoxic T cells, CD107a is upregulated and expressed on the cell surface concomitantly with cytokine secretion and target cell lysis; therefore this study has been frequently used to extrapolate the magnitude of cytotoxic activity. Degranulation (CD107a expression) assays have gained traction in the assessment of familial/primary hemophagocytic lymphohistiocytosis (FHL).[21] The exception is in patients with mutations in the gene encoding perforin (*PRF1;* FHL type 2) where degranulation (CD107a expression) of cytotoxic cells is normal while cytotoxicity is abnormal. Therefore, while degranulation assays may provide relevant information, they cannot substitute for direct measurement of cytotoxicity in all settings.

NK cells also mediate the killing of target cells, via recognition of surface-bound immunoglobulin (IgG) through the Fcγ receptors (Ig-Fc receptors, specifically CD16 or FcγRIIIa), through ADCC. ADCC assays are particularly relevant in the assessment of antitumor responses, especially of new biological

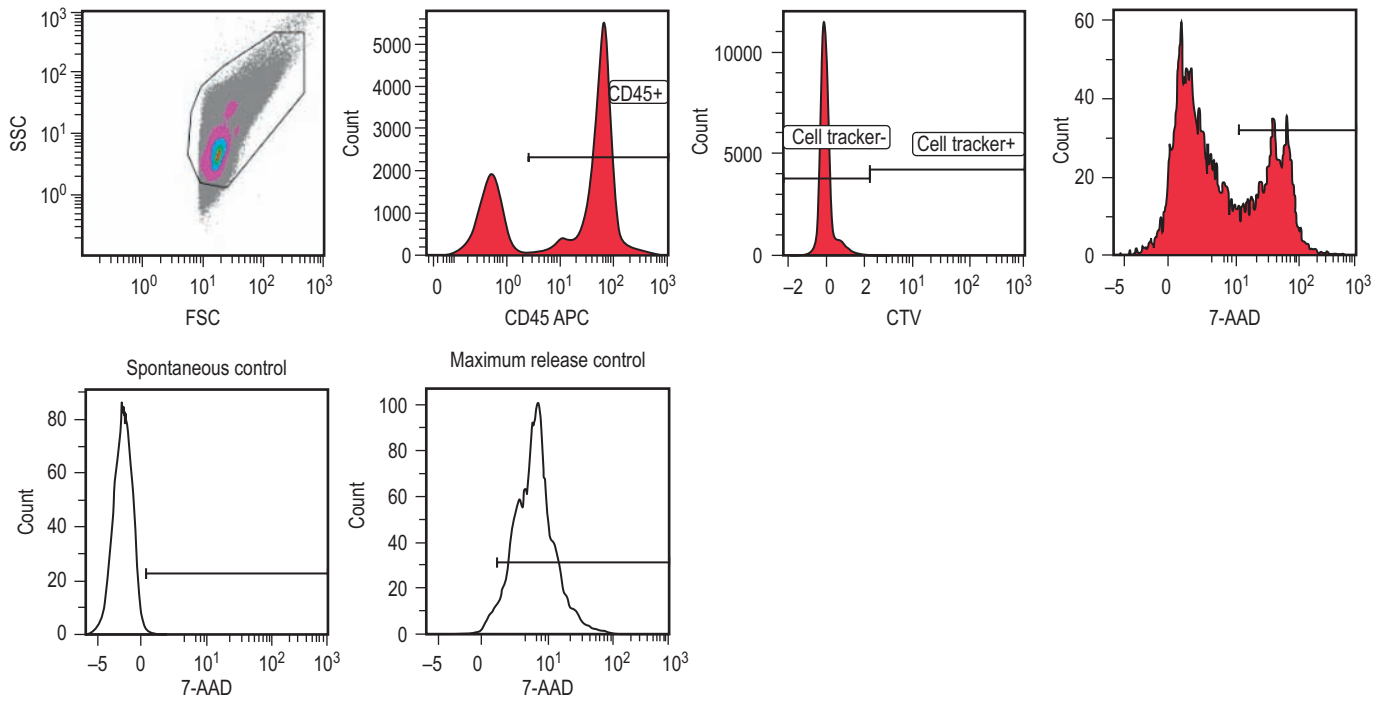


FIG. 94.6A Flow Cytometric Assessment of Natural Killer Cell Cytotoxic Function. The "gold standard" for measuring natural killer cell cytotoxicity (NKC) uses a radioactive label for target cells. In the flow-based assay, target cells (K562, a major histocompatibility complex [MHC] class I–negative erythroleukemia cell line derived from a patient with chronic myelogenous leukemia [CML] in blast crisis) are labeled with a fluorescent marker, CellTracker (*CTV*). Peripheral blood mononuclear cells (PBMC) from healthy controls or patients are cocultured with labeled target cells, and the two cell populations are identified by forward scatter (*FSC*) and side scatter (*SSC*) parameters. Afterward, $CD45^+$ lymphocytes are identified because NK cells (effector cells) are within the lymphocyte subset. CTV^+ and CTV^- cells are separated; CTV^+ cells that are also 7-amino-actinomycin D (*7-AAD*)$^+$ are identified as killed target cells, and their frequency is estimated. Negative control for spontaneous cytotoxicity is achieved by incubation of labeled target cells without effector cells added (called spontaneous control). Positive control for maximal cytotoxicity is achieved by incubating labeled target cells for the same duration but using a permeabilization/fixation method to obtain maximal cell death and fluorescent-positive cells (called maximum release control). This figure depicts spontaneous NK cell cytotoxicity with different concentrations of effector cells (PBMC; E) to a single concentration of target cells (T) starting at 30:1. The data are expressed as delta (Δ)% cytotoxicity after the spontaneous control signal is subtracted from the E:T signal for each concentration.

immunomodulatory agents, as well as of the function of alloantibodies in allograft rejection.

EVALUATION OF REGULATORY T-CELL FUNCTION

Regulatory T cells ($FOXP3^+$Treg) have been well described over the past several years as a distinct subset of T cells that are both developmentally and functionally unique as well as essential to maintaining immune homeostasis and self-tolerance. The major subpopulations of Treg cells include natural Treg (nTreg) cells that are produced in the thymus and induced Treg (iTreg) cells that are generated in the periphery from conventional $FOXP3^-CD4^+$ T cells. The dysfunction of Treg cells results in severe autoimmunity, with IPEX serving as the classic prototype (Chapter 34). Besides their role in controlling the development of autoimmunity, lack of Treg cells or abnormal Treg function has been implicated in the etiopathogenesis of graft-versus-host disease (GvHD) and allograft rejection, while its presence and normal function have been shown to promote allograft tolerance in solid-organ transplantation.

KEY CONCEPTS

Regulatory T Cells and B Cells

- Regulatory T cells (Treg) and B cells (Breg) are distinct subsets of cells with immune regulatory potential and importance in maintaining immune homeostasis and self-tolerance, preventing autoimmunity, and limiting inflammatory damage.
- FOXP3 is a transcription factor and a marker for natural regulatory T cells; while there are no specific cellular markers that define Breg, the production of interleukin (IL)-10 is considered a hallmark.
- Both Treg and Breg have been shown to play important roles in controlling autoimmunity, and graft-versus-host disease (GvHD) as well as in mediating transplant tolerance.
- Treg function is measured in vitro through different types of suppression assays, while Breg are characterized by their ability to produce IL-10 when stimulated via Toll-like receptor (TLR9)/B-cell receptor (BCR) or CD40L cross-linking.
- Natural, thymic-derived Treg are $CD4^+25^+FOXP3^+$ while Breg are $CD19^+CD24^{hi}CD38^{hi}IL\text{-}10^+$.
- Treg can also be induced in the periphery from conventional T cells through cytokine signals (iTreg).

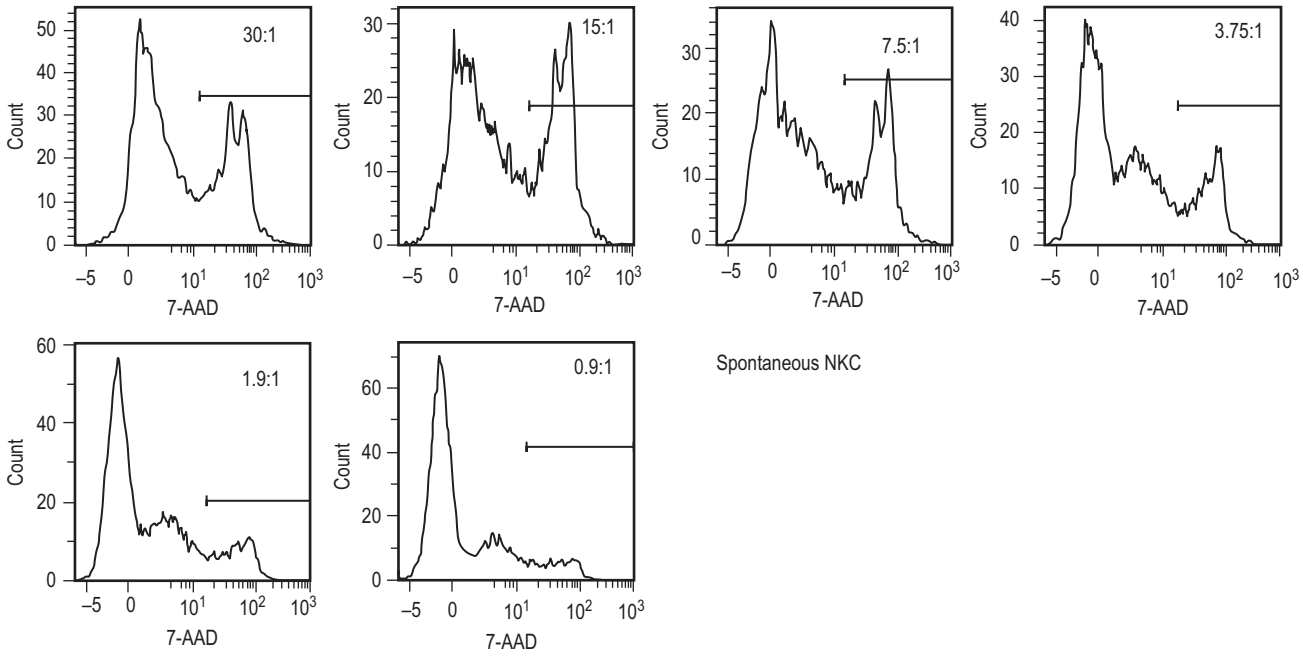


FIG. 94.6B Lymphokine-Activated Killing by Natural Killer Cells. This graph depicts IL-2-stimulated NK cell cytotoxicity (*NKC*) or lymphokine-assisted killing. The methodology is similar to what is described for Fig. 94.6A. The only difference is that a fixed concentration of IL-2 is used with each E:T ratio to activate NK cells and mediate cytotoxicity, and the mechanism of killing is different from that used in spontaneous NK cell cytotoxicity. IL-2 promotes NK cell killing to targets that are resistant to resting NK cell killing. IL-2 stimulation of NK cells results in activation of WAVE2 (WASp family verprolin-homologous 2), which allows F-actin reorganization independent of WASp (Wiskott-Aldrich syndrome protein), and associated NK cell function (Orange JS et al. IL-2 induces a WAVE2-dependent pathway for actin reorganization that enables WASp-independent human NK cell function. *J Clin Invest.* 2011;121(4):1535–1548). This is notably observed in patients with Wiskott-Aldrich syndrome (WAS) who demonstrate (IL-2-stimulated NK cell cytotoxicity but not spontaneous NK cell cytotoxicity). *7-AAD*, 7-Amino-actinomycin D.

Treg cell function is measured in vitro with Treg suppression assays, which utilize sorted $CD4^+CD25^+$ Treg cells in a coculture system with conventional effector T cells to assess suppression of proliferation. However, there are limitations to this approach, including the issue of whether the Treg suppression assays in vitro reflect the biological process in vivo. In addition, antigen-specific Treg suppression cannot be adequately assessed due to the technical difficulties in obtaining sufficient numbers of antigen-specific cells, so the use of a polyclonal activation model with bulk PBMC represents the standard to study Treg suppression. Rapid tests for Treg cell function have been described using short-term (7 to 20 hours) flow-based assays to measure suppression of T-cell activation marker expression (CD40L and CD69). These assays utilize effector T cells ($CD4^+25^-$) activated with anti-CD3/anti-CD28 beads with and without the addition of freshly isolated Treg cells or *ex vivo* expanded Treg cells. When implementing Treg suppression assays in the clinical diagnostic laboratory, a number of technical considerations can confound the interpretation of results, and these are well described by McMurchy et al.[22] In addition, it remains unclear whether standard in vitro Treg suppression assays can adequately account for the significant complexity of Treg subsets that may have different functional properties in various clinical contexts. Therefore the newer techniques of mass cytometry may be required to address the phenotypic diversity and to eventually be harnessed either to evolving applications or other technology to analyze the functional complexity of these populations.

ASSESSMENT OF SIGNALING AND DNA REPAIR PATHWAYS IN LYMPHOCYTES VIA PHOSPHOFLOW CYTOMETRY

A key aspect of studying lymphocyte responses is to assess signaling in appropriate lymphocyte subsets and its alteration in

KEY CONCEPTS

Assessment of DNA Repair Pathways by Phosphoflow

- DNA double-strand break (DSB) repair defects are relevant to VDJ recombination, isotype class switching, and lymphocyte maturation.
- Defects in this process can lead to immunodeficiency with susceptibility to infection and malignancy.
- Homologous recombination (HR), which is error-free, involves RAD50, RAD51, RAD52, and Mre11, while nonhomologous end-joining (NHEJ) is error-prone and utilizes Ku70/80, DNA-PKcs, Artemis, DNA Ligase IV, XRCC4, XLF/Cernunnos.
- ATM is a key regulator of cell-cycle checkpoints following irradiation-induced DSB and coordinates timing of phosphorylation of checkpoint proteins.
- Cell cycle analysis can help identify checkpoint defects and can be easily assessed by flow cytometry.
- Phosphorylation of H2AX (γH2AX) is a useful marker of DNA damage related to irradiation-induced DSB.
- However, several kinases phosphorylate H2AX, which can confound identification of specific defects; therefore assessing multiple proteins in the pathway allows identification of a broader group of DNA repair defects.

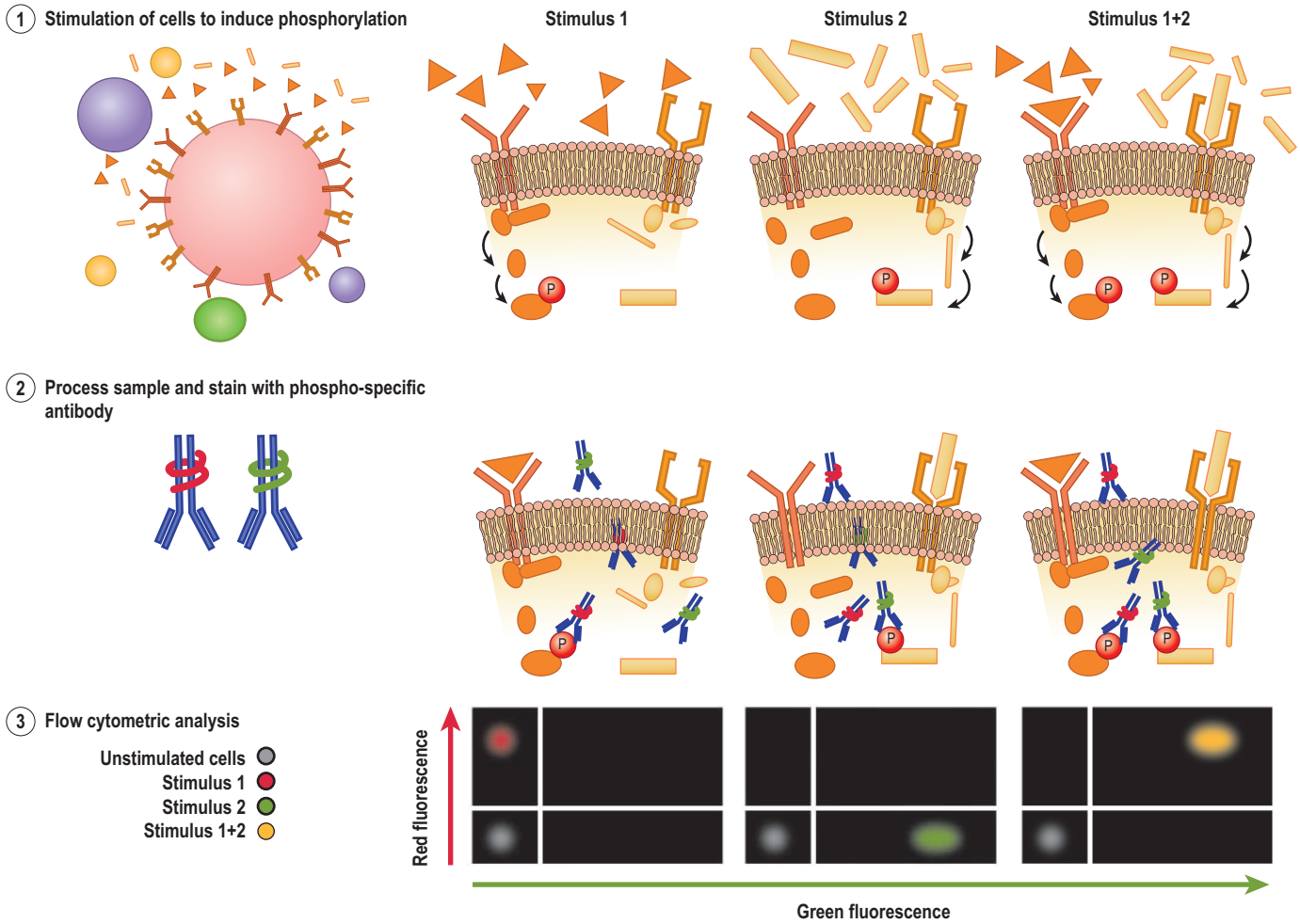


FIG. 94.7 Analysis of Phosphoproteins in Cells by Flow Cytometry. Cells activated by specific stimuli can be assessed for activation and induction of specific signaling pathways by measuring phosphoproteins. A single stimulus or multiple stimuli can be used to phosphorylate different proteins, and intracellular staining can be performed with phospho-specific antibodies, which is subsequently analyzed by multicolor flow cytometry. See text for detailed explanation. (Figure modified from Fig. 2 by Krutzik PO, et al. Analysis of protein phosphorylation and cellular signaling events by flow cytometry: techniques and clinical applications. *Clin Immunol* 2004; 110:206–2210.

pathological conditions. A versatile tool called phosphoflow is unique in its ability to allow the study of multiple intracellular signaling molecules in specific lymphocyte populations at a single-cell level. Phosphoflow assays for the signal transducer and activator of transcription (STAT) molecules (Fig. 94.7) have been well described and are discussed in Chapter 93. The use of phosphoflow assays to assess Bruton tyrosine kinase (BTK) phosphorylation in X-linked agammaglobulinemia (XLA) with leaky (hypomorphic) defects, as well as to assess radiosensitivity and the DNA repair pathway, will be covered briefly.

Mutations in *BTK* impair B-cell maturation and function, and patients with XLA can have either no peripheral B cells (null mutations) or reduced B cells (hypomorphic/leaky mutations), depending on the specific genetic defect (Chapter 33). Two regulatory tyrosine residues in BTK undergo rapid phosphorylation upon BCR cross-linking (Y551 in the SH1 domain and Y223 in the SH3 domain). In the flow assay, Y223 phosphorylation is measured after an anti-IgM antibody is used to cross-link the BCR (for 3 minutes) because Y551 phosphorylation could not be detected in the time interval the assay was performed. In addition to PBMC, Ramos cell line (a B-cell line derived from a patient with Burkitt lymphoma) is used as a control for B cells, while pervanadate (complex of vanadate with hydrogen peroxide), an irreversible protein tyrosine-phosphatase inhibitor, is used in this assay as a positive control (Fig. 94.8). In the Ramos cell line, it is possible to visualize the Y551 phosphorylation (see Fig. 94.8) with pervanadate treatment.

A number of genetic disorders, collectively classified as XCIND (x-ray [irradiation] sensitivity, cancer susceptibility, immunodeficiency, neurological involvement, and double-strand DNA breakage), cause impairment in cellular ability to repair DNA double-strand breaks (DSBs). These defects have a significant impact on the ability of cells to grow, differentiate, and function normally; they include ataxia-telangiectasia (AT) due to *ATM* gene mutations and radiosensitive severe combined immunodeficiencies (rs-SCIDs) due to mutations in the *DCLRE1C, LIG4, NHEJ1*, and *PRKDC* genes. These disorders make it very desirable to have an assay capable of rapidly assessing DNA repair in lymphocytes in response to radiation damage. A flow cytometry assay has been developed that is capable of measuring the function of several proteins in the DNA repair pathway after induction of DSBs via irradiation (Fig. 94.9A). Following

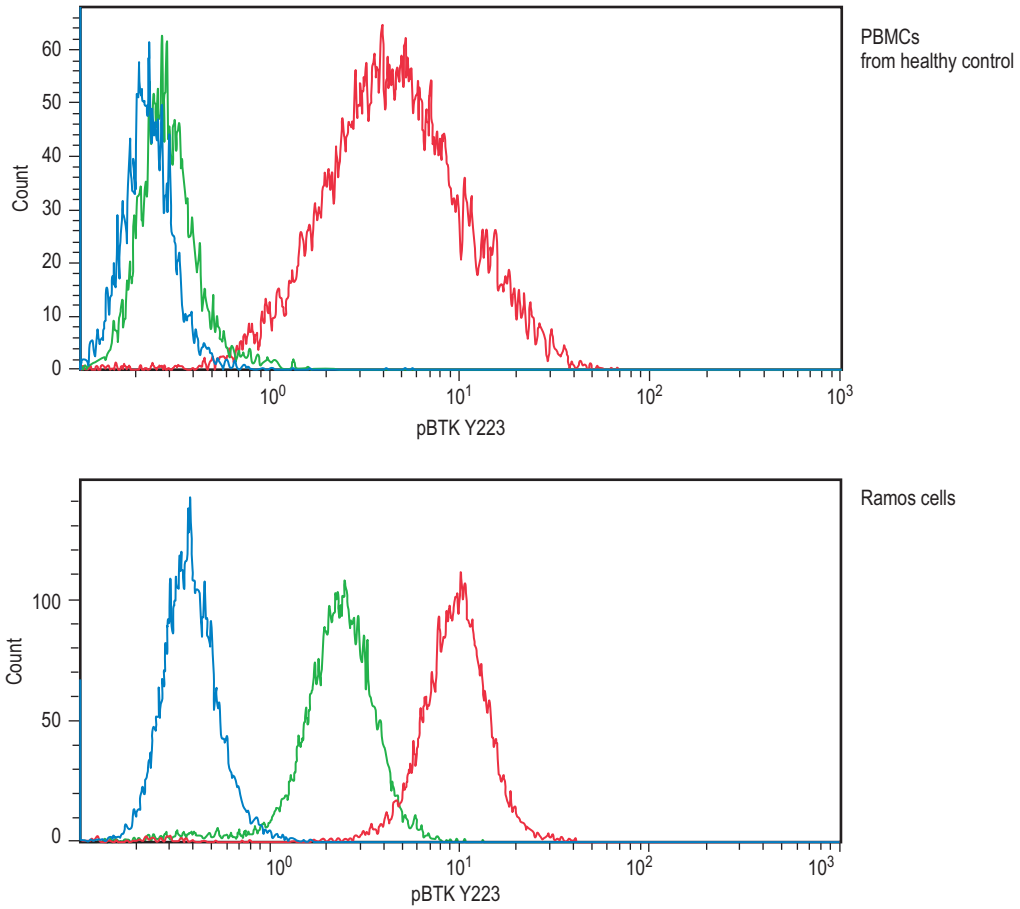


FIG. 94.8 Analysis of Bruton Tyrosine Kinase Phosphorylation in Peripheral Blood Mononuclear Cell and Ramos B-Cell Line. Bruton tyrosine kinase (*BTK*) phosphorylation can be assessed in B cells in X-linked agammaglobulinemia (XLA) patients with leaky defects who have some preserved B cells in the blood. B cells are stimulated with either anti-immunoglobulin M (IgM) (B-cell receptor [BCR] cross-linking; *green line*) or pervanadate (positive control; *red line*). The example shown here is for BTK phosphorylation at the Y223 residue in peripheral blood mononuclear cells (*PBMCs*) from a healthy donor *(top panel)* and a B-cell line, Ramos cells. The *blue line* represents the unstimulated control. In the bottom panel, BTK phosphorylation of the Y551 residue is shown for PBMCs from a healthy donor *(left)* and Ramos cells *(right)*. The phosphorylation of the Y551 residue cannot be visualized for anti-IgM stimulation, and therefore data are shown only for the positive control, pervanadate.

low-dose irradiation, the function of ATM and ATR (ATM-Rad3–related kinase) pathways are assessed by analyzing the autophosphorylation of ATM at serine 1981 (see Fig. 94.9A). This step is required for ATM activation and is followed by phosphorylation of downstream targets SMC1 and H2AX. H2AX is a histone that belongs to the H2A family and is a component of the histone octamer in nucleosomes. It is phosphorylated by both ATM and ATR and is the first step in the recruitment and localization of DNA repair proteins. In patients with AT, there is a complete absence of phosphorylation. The frequency (%) of lymphocyte subsets that phosphorylates H2AX (γH2AX) appears normal (Fig. 94.9B), as the phosphorylation by ATR and other kinases is intact; however, the magnitude of phosphorylation, as measured by the normalized median fluorescence intensity (MFI ratio), is substantially decreased (Fig. 94.9C). The kinetics of phosphorylation indicates that maximal phosphorylation occurs 1 hour after induction of DSB and that there is dephosphorylation by 24-hour post-irradiation (not shown) in healthy controls, but in AT patients, since there is a lack of ATM phosphorylation on induction of DNA DSBs, there is no kinetic regulation. This flow-based assay enables a rapid assessment of radiation sensitivity in various clinical contexts[23,24] as well as the ability to visualize the function of multiple proteins of the DNA repair pathway. It can also be used to characterize the DNA repair function of known and unknown genetic defects, and it can identify functional phenotypes in patients with atypical presentations that may be missed if only genetic information is used.

ASSESSMENT OF B-CELL FUNCTION

While T cells and NK cells form the foundation of the cellular immune response, B cells are the main driver of humoral immunity. B cells are multifaceted in their function; they produce antibodies via differentiation into plasma cells, they act as APC for T cells, and they secrete potent immunomodulatory cytokines while downregulating immune responses via IL-10 production. B cells can proliferate in response to polyclonal mitogenic stimuli such as PWM, albeit with a much weaker response than seen with T cells stimulated with PHA (see Fig. 94.2). The starting point for any evaluation of B-cell function involves measuring

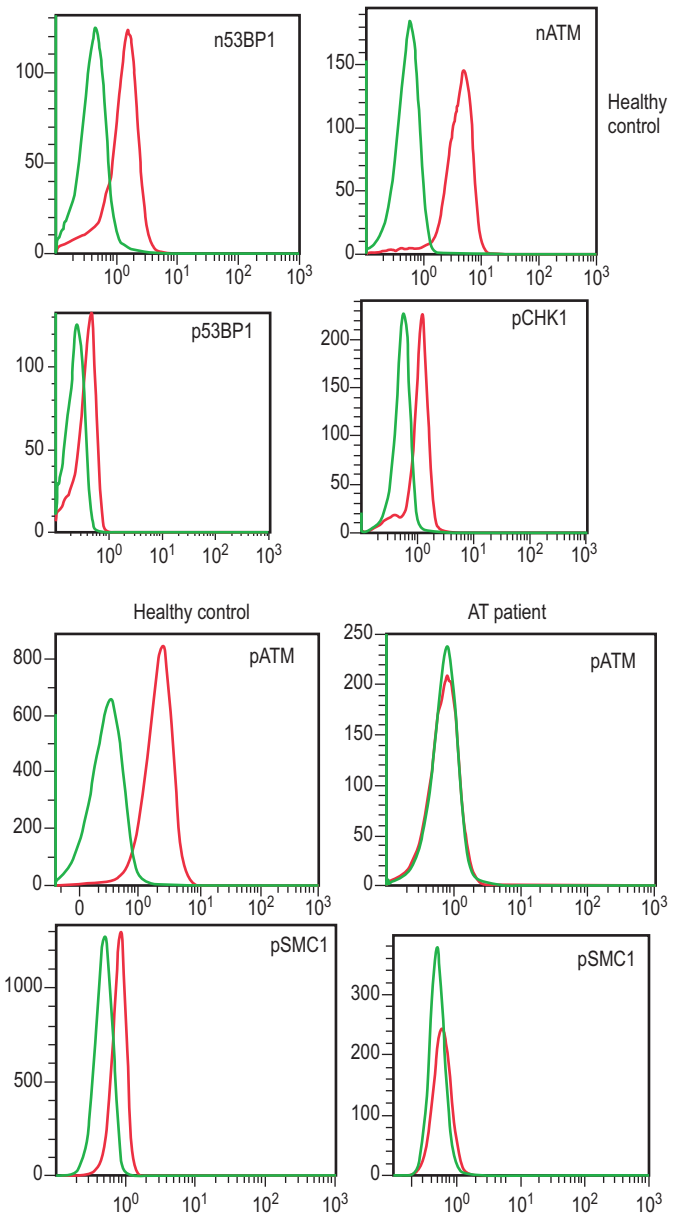


FIG. 94.9A Assessment of Radiosensitivity in Lymphocytes by Flow Cytometry. DNA repair pathway defects can be rapidly and sensitively analyzed by flow cytometry. The *left set of panels* shows ATM and SMC1 phosphorylation in T cells of a healthy control (similar data in B and natural killer (NK) cells; data not shown). The *right set of panels* shows absent phosphorylation of ATM and SMC1 in a patient with ataxia-telangiectasia (AT). The *green line* represents the unirradiated sample, and the *red line* represents the data post-irradiation with 2 Gy (low-dose) radiation. AT patients also show an inability of ATM to phosphorylate not only itself but also downstream targets, such as SMC1 in T cells. This is also true for other lymphocyte subsets (B cells and NK cells).

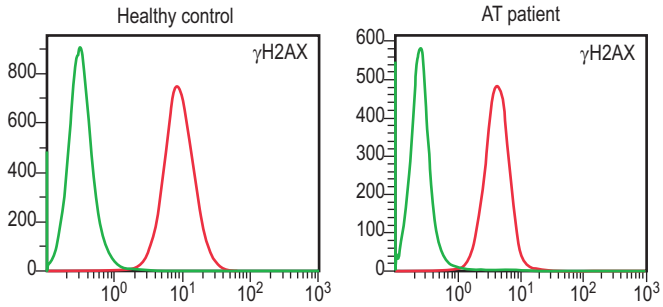
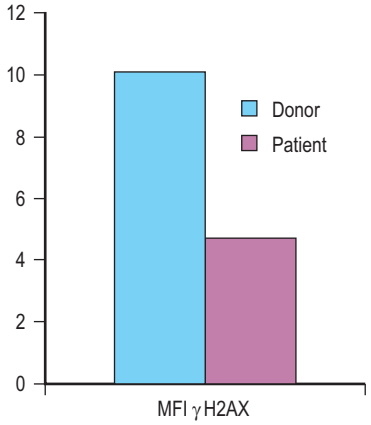


FIG. 94.9B Role of γH2AX in DNA Repair. Phosphorylated histone H2AX (*γH2AX*) is one of the earliest proteins to be mobilized in the presence of DNA double-strand breaks (DSBs), which in this case was caused by exposure to radiation. Once DNA DSBs are repaired, γH2AX is dephosphorylated and returns to baseline. In ataxia-telangiectasia (*AT*) patients, there does not appear to be any defect in phosphorylation of H2AX (γH2AX) and mobilization based on the proportion (%) of T, B, or NK cells expressing this protein.

serum immunoglobulin levels followed by in vivo antibody responses to vaccination with both protein (e.g., tetanus toxoid) and carbohydrate (e.g., Pneumovax 23) antigens.

A more recent area of focus in the laboratory evaluation of B cells has been the regulatory function of B cells, specifically concerning the subset classified as regulatory B (Breg) cells. B cells that secrete IL-10 have been described as Breg cells and are now recognized to be an important component of the host immune response that protects against autoimmunity and also limits inflammatory damage.[8] In humans, the phenotype of Breg cells has been described as $CD19^{+}CD27^{-}CD24^{hi}CD38^{hi}CD5^{+}CD1d^{hi}$. However, the ability to produce IL-10 appears to be the defining feature rather than any particular constellation of cell-surface markers. IL-10 exerts potent anti-inflammatory effects and enhances survival, proliferation, differentiation, and isotype class-switching of B cells. Both naïve and memory human B cells have the ability to produce IL-10 in response to stimulation via TLR9 and the BCR, but only ~15% of B cells can produce IL-10 in response to stimulation. In the laboratory, Breg cells can be assessed by isolating PBMC from blood and culturing with CpG-B (TLR9 stimulation) or CpG-B plus recombinant CD40L for 3 days in vitro, followed by the addition of PMA, ionomycin, and brefeldin A (BFA) for the last 5 hours of culture. Cells can then be harvested, washed, and stained with CD19 and IL-10 antibodies (the latter requires intracellular staining), and then analyzed with a flow cytometer. To assess the presence of Breg cells in blood without in vitro differentiation, blood or PBMC can be analyzed for $CD19^{+}CD24^{hi}CD38^{hi}$ B cells that are positive for intracellular IL-10 using a similar flow cytometry protocol, but the levels of these cells can be highly variable, and accuracy of detection and quantitation is likely to be dependent on the analytical method and clinical context.[2]

[2]Portions of this section have been reproduced with permission from ASM Press (Abraham RS, Lymphocyte Activation) (copyright permission obtained, 5th edition).



	MFI (γH2AX)
Donor	10.13
Patient-AT	4.72

FIG. 94.9C Relevance of Median Fluorescence Intensity Assessment for DNA Repair Defect Measurements. Median fluorescence intensity (*MFI*) can provide additional valuable information on DNA repair defects and should be used along with frequency analysis. In ataxia-telangiectasia (AT) patients, although the proportion (%) of lymphocyte subsets expressing phosphorylated histone H2AX (*γH2AX*) appears normal, it is significantly decreased (by >50% in AT (*purple bar*) compared with healthy controls (*blue bar*). The MFI is represented as normalized to the unirradiated sample, as a ratio.

SUMMARY

In conclusion, lymphocyte responses in humans can be assessed using a variety of analytical tools as described in this chapter and applied to many clinical contexts, including but not limited to primary immune deficiencies, autoimmunity, transplantation, and immune dysregulation disorders. Most of these measurements use cells from blood and not from other lymphoid tissues, due to accessibility and ease of generating control reference ranges. This chapter is not an exhaustive treatise on all aspects of lymphocyte function and the immune response; nor does it cover all areas of normal and abnormal pathology in this context. Rather, it is meant to take the reader on a tour of the immune landscape and to provide salient highlights of the diversity and relevance of lymphocyte function.

ON THE HORIZON

- The advent of newer assays, including flow cytometry-based as described herein, have largely but not completely replaced traditional and often less sensitive radioactive or other cumbersome methods for assessing lymphocyte function.
- Standardization of flow cytometry assays can improve quality of data and reporting of results across laboratories performing similar tests (Optimized Multicolor Immunofluorescence Panels-OMIPs[25]; MiSet RFC Standards[26]; MIFlowCyt.[27]
- Microchip and advanced mass cytometry techniques, along with imaging cytometry, may offer newer, multiplex approaches to the assessment of lymphocyte subsets, individually and in cellular interactions.
- Functional and phenotypic data need to be ideally correlated with relevant genomic, transcriptomic, proteomic, metabolomic, epigenomic, and microbiome analyses for effective characterization of the complex interactions that govern the immune response in the normal and dysregulated state.
- This will necessitate further refinements of "big data" analysis, which can include but is not limited to, experimental studies in addition to computational modeling. Examples include antigen-specific characterization of specific B-cell and T-cell receptors in an individual as an extension of current "deep-sequencing" repertoire analysis or transcriptomic analysis during immune quiescence or activation on a multidimensional scale.

REFERENCES

1. Lanzavecchia A. Dissecting human antibody responses: useful, basic and surprising findings. *EMBO Mol Med*. 2018;10(3):e8879.
2. Puga I, Cols M, Barra CM, et al. B cell-helper neutrophils stimulate the diversification and production of immunoglobulin in the marginal zone of the spleen. *Nat Immunol*. 2012;13(2):170–180.
3. Vinuesa CG, Chang PP. Innate B cell helpers reveal novel types of antibody responses. *Nat Immunol*. 2013;14(2):119–126.
4. Zimmermann P, Curtis N. Factors that influence the immune response to vaccination. *Clin Microbiol Rev*. 2019;32(2):e00084.
5. Cunningham AF, Flores-Langarica A, Bobat S, et al. B1b cells recognize protective antigens after natural infection and vaccination. *Front Immunol*. 2014;5:535.
6. Park JH, Lee KH, Jeon B, et al. Immune dysregulation, polyendocrinopathy, enteropathy, X-linked (IPEX) syndrome: a systematic review. *Autoimmun Rev*. 2020;19(6):102526.
7. Kunicki MA, Amaya Hernandez LC, Davis KL, et al. Identity and diversity of human peripheral Th and T regulatory cells defined by single-cell mass cytometry. *J Immunol*. 2018;200(1):336–346.
8. Mauri C, Menon M. Human regulatory B cells in health and disease: therapeutic potential. *J Clin Invest*. 2017;127(3):772–779.
9. Abel AM, Yang C, Thakar MS, Malarkannan S. Natural killer cells: Development, maturation, and clinical utilization. *Front Immunol*. 2018;9:1869.
10. Buck MD, O'Sullivan D, Pearce EL. T cell metabolism drives immunity. *J Exp Med*. 2015;212(9):1345–1360.
11. MacIver NJ, Michalek RD, Rathmell JC. Metabolic regulation of T lymphocytes. *Annu Rev Immunol*. 2013;31:259–283.
12. Abraham RS. Lymphocyte a Activation. In: Detrick B, Schmitz JL, Hamilton RG, eds. *Manual of Molecular and Clinical Laboratory Immunology*. 8th ed. Washington, DC: ASM Press; 2016:269–279.
13. Bitar M, Boldt A, Freitag MT, et al. Evaluating STAT5 phosphorylation as a mean to assess T cell proliferation. *Front Immunol*. 2019;10:722.
14. Mbitikon-Kobo FM, Bonneville M, Sekaly RP, Trautmann L. Ex vivo measurement of the cytotoxic capacity of human primary antigen-specific CD8 T cells. *J Immunol Methods*. 2012;375(1–2):252–257.
15. Meesing A, Abraham RS, Razonable RR. Clinical correlation of cytomegalovirus infection with CMV-specific CD8⁺ T-cell immune competence score and lymphocyte subsets in solid organ transplant recipients. *Transplantation*. 2019;103(4):832–838.
16. Walker S, Fazou C, Crough T, et al. Ex vivo monitoring of human cytomegalovirus-specific CD8⁺ T-cell responses using QuantiFERON-CMV. *Transpl Infect Dis*. 2007;9(2):165–170.
17. Paouri B, Soldatou A, Petrakou E, et al. Quantiferon-Cytomegalovirus assay: a potentially useful tool in the evaluation of CMV-specific CD8⁺ T-cell reconstitution in pediatric hematopoietic stem cell transplant patients. *Pediatr Transplant*. 2018;22(5):e13220.
18. Banas B, Boger CA, Luckhoff G, et al. Validation of T-Track(R) CMV to assess the functionality of cytomegalovirus-reactive cell-mediated immunity in hemodialysis patients. *BMC Immunol*. 2017;18(1):15.
19. Barabas S, Spindler T, Kiener R, et al. An optimized IFN-gamma ELISpot assay for the sensitive and standardized monitoring of CMV protein-reactive effector cells of cell-mediated immunity. *BMC Immunol*. 2017;18(1):14.

20. Bestard O, Lucia M, Crespo E, et al. Pretransplant immediately early-1-specific T cell responses provide protection for CMV infection after kidney transplantation. *Am J Transplant*. 2013;13(7):1793–1805.
21. Bryceson YT, Pende D, Maul-Pavicic A, et al. A prospective evaluation of degranulation assays in the rapid diagnosis of familial hemophagocytic syndromes. *Blood*. 2012;119(12):2754–2763.
22. McMurchy AN, Levings MK. Suppression assays with human T regulatory cells: a technical guide. *Eur J Immunol*. 2012;42(1):27–34.
23. Buchbinder D, Smith MJ, Kawahara M, et al. Application of a radiosensitivity flow assay in a patient with DNA ligase 4 deficiency. *Blood Adv*. 2018;2(15):1828–1832.
24. Cousin MA, Smith MJ, Sigafoos AN, et al. Utility of DNA, RNA, protein, and functional approaches to solve cryptic immunodeficiencies. *J Clin Immunol*. 2018;38(3):307–319.
25. Mahnke Y, Chattopadhyay P, Roederer M. Publication of optimized multicolor immunofluorescence panels. *Cytometry A*. 2010;77(9):814–818.
26. Lucas F, Gil-Pulido J, LaMacchia J, et al. MiSet RFC standards: defining a universal minimum set of standards required for reproducibility and rigor in research flow cytometry experiments. *Cytometry A*. 2020;97(2):148–155.
27. Lee JA, Spidlen J, Boyce K, et al. MIFlowCyt: the minimum information about a Flow Cytometry Experiment. *Cytometry A*. 2008;73(10):926–930.

Assessment of Neutrophil Function

Debra Long Priel and Douglas B. Kuhns

Neutrophils, also known as polymorphonuclear neutrophils (PMNs; because of their multilobed nucleus) or granulocytes (because of the numerous granules found in the cytoplasm) are major contributors to innate host defense against invading microorganisms (particularly bacteria and fungi). Neutrophils are bone marrow–derived, terminally differentiated cells incapable of further cellular division but could be derived from progenitor populations in the spleen.[1] Early studies indicated that neutrophils have a short life span in the circulation ($t_{1/2}$ = 6–8 hours) and they survive an additional 1 to 2 days in surrounding tissue;[2] more recent findings reveal that neutrophils may survive 10 times longer in circulation, up to 5.4 days.[3]

Neutrophils, with a diameter of 10 to 15 μm and a volume of 346 μm^3, have a unique morphology. The nucleus of a mature neutrophil is segmented into 3 to 5 lobes with chromosomes randomly distributed among the lobes. Neutrophils also have an extensive array of storage granules prepackaged with specific proteins defined by the differentiation stage during maturation.[4] Granules are classified into four distinct groups: azurophilic, specific, gelatinase, and secretory granules. Azurophilic granules contain myeloperoxidase, lysozyme, antimicrobial peptides, defensins, proteases, and lysosomal acid hydrolases. The specific granules contain lactoferrin, lysozyme, and vitamin B_{12}–binding protein serving as storage pools for CD11b/CD18 and cytochrome b_{558} of the superoxide anion radical ($O_2^{\bullet}$)–generating enzyme, nicotinamide adenine dinucleotide phosphate (NADPH) oxidase, or NOX2. The gelatinase granules are a subset of the specific granules that have a high content of gelatinase. The secretory granules are highly mobilizable intracellular vesicles that contain alkaline phosphatase and other surface antigens.

The primary function of neutrophils is the ingestion (phagocytosis) and subsequent killing of microorganisms. This process requires the assembly of NOX2, the phagocyte oxidase (phox) enzyme complex consisting of at least four cytosolic components—$p47^{phox}$, $p67^{phox}$ [5], $p40^{phox}$ [6] and Rac2[7]—and two membrane components—$p22^{phox}$ and $gp91^{phox}$—that constitute cytochrome b_{558}.[8,9] This enzyme reduces molecular O_2 to $O_2^{\bullet}$ using NADPH generated by the oxidation of glucose through the pentose phosphate pathway; $O_2^{\bullet}$ either spontaneously or enzymatically converts to hydrogen peroxide (H_2O_2). In the presence of a metal, such as iron (Fe^{2+}), H_2O_2 and $O_2^{\bullet}$ can react to form the highly reactive hydroxyl radical, $OH^{\bullet}$. Alternatively, the azurophilic granule constituent (myeloperoxidase) catalyzes the formation of hypochlorous acid from H_2O_2 and chloride (Cl^-). The combined activities of these reactive O_2 species (ROS), antimicrobial peptides, and lysosomal hydrolases result in the ultimate destruction of the ingested microorganism. Excess production of ROS and release of lysosomal hydrolases into the extracellular milieu can lead to tissue damage and inflammation.

Neutrophils can also exude extracellular microbicidal activity through the formation of neutrophil extracellular traps (NETs), a matrix of DNA and granular enzymes that is purported to entrap bacteria and promote their killing.[10] During NET formation (NETosis), the nucleus loses its lobular shape, the nuclear membrane disintegrates into a chain of vesicles surrounding the DNA, and the cell loses granular integrity. The nuclear material fills most of the cell, mixing with the granular contents. The cells round up and DNA is forcibly extruded from the cell, conveying its granular enzymes trapped within the DNA matrix. Mitochondria in neutrophils (few in number with relatively little oxidative phosphorylation) play an important role in NETosis. Activation of neutrophils by immune complexes results in marked depolarization of the mitochondria, increased mitochondrial ROS production, and redistribution of the mitochondria to the periphery of the neutrophil where they contribute their mitochondrial DNA to the extruded NET.[11] A subset of neutrophils that co-sediment with mononuclear cells on a Ficoll-Paque discontinuous gradient cushion (low-density neutrophils) appears to be more prone to NET formation.[12] Many techniques measure NETosis, ranging from assays that measure extracellular elastase activity bound to DNA and releasable by DNase treatment to quantitative immunofluorescence assays that monitor the release of extracellular DNA.[13] Multicolor fluorescence-activated cell sorting assays have been described that monitor extracellular DNA, citrullinated histone h3, and myeloperoxidase (MPO) as surrogate markers for cells undergoing NETosis in vivo.[14] NETs are thought to play a role in enhancement of the inflammation seen in autoimmune diseases such as psoriasis, rheumatoid arthritis, and systemic lupus erythematosus. Recently, hypotheses reveal that unregulated NET formation contributes to the inflammatory and microvascular thrombotic complications in the lungs of patients with COVID-19.[15]

Neutrophils display a diverse array of cellular functions. Abnormalities in these functions can severely compromise host defense, leading to recurrent bacterial and fungal infections. To localize specific deficiencies of neutrophil function, assays have been developed to mimic these functions both in vivo and in vitro. Often, a preliminary screening of several neutrophil functions is performed to localize deficits and more vigorous testing of specific function is performed. Assays to assess neutrophil function should address several limitations—the number of cells required for the assay, the type of cell preparation needed (isolated neutrophils vs. whole blood), the overall incubation time for the assay, the complexity of the assay, and the rapidity of data collection. These issues become more critical if multiple functional assays are planned concurrently. Since neutrophils cannot be stored or frozen and maintain viability, neutrophils

from normal individuals are generally assayed in parallel to validate the results, doubling the number of assays to be performed. Additionally, isolation of neutrophils can take 1 to 2 hours, limiting the time available for functional assays. Fluorescent probes have increased the sensitivity of many assays and eliminated the need for radioactive probes. The use of multiwell microplates and microplate readers has reduced the number of cells required and has facilitated the collection of data. Experience in handling neutrophils and the time constraints of assays can limit the availability of this testing to laboratories that specialize in assessment of neutrophil function.

KEY CONCEPTS

Criteria to Assess Neutrophil Function

Because of (1) the time required to isolate neutrophils and (2) the shortened life span of neutrophils after isolation, assays of neutrophil function should have minimal complexity and enable rapid data collection.

ISOLATION OF NEUTROPHILS

Clinical Indications and Implications

Assays that avoid neutrophil isolation are preferred because of their artificial priming during isolation.[16] However, most assays require isolated neutrophils to eliminate any possible contributions of other leukocytes and blood components. In general, blood should be drawn using either citrate or heparin as anticoagulant and maintained at 20°C to 25°C in polypropylene containers. Most isolation protocols require 1 to 2 hours to obtain purified neutrophils.

Principles and Interpretation of Laboratory Assessment

Most neutrophil isolation protocols use differences in the cell density as the basis for the separation. The relative densities of blood cells are as follows: erythrocytes > neutrophils and eosinophils > monocytes, lymphocytes, and basophils > platelets. Ficoll-Paque is a solution of sodium diatrizoate (a dense, triiodinated compound), and Ficoll (a polysaccharide) with a density (1.077 g/cm^3) that falls between the density of neutrophils and that of the mononuclear cells. To isolate neutrophils,[17] whole blood is diluted with saline and underlaid with Ficoll-Paque solution. After centrifugation for 30 minutes at 500*g*, the less-dense monocytes, lymphocytes, basophils, and platelets remain at the upper interface of the Ficoll-Paque solution, whereas the denser erythrocytes and neutrophils pass through the solution and pellet at the bottom. The mononuclear cells are carefully harvested and the remaining Ficoll-Paque solution aspirated. The erythrocyte/neutrophil pellet is resuspended with saline and mixed with 3% dextran. Dextran promotes rouleaux formation of erythrocytes, causing them to sediment more rapidly than the neutrophils at 1*g*. The neutrophil-enriched supernatant fluid is harvested from the bulk of the (sedimented) erythrocytes. Contaminating erythrocytes are removed by a brief (30-second) hypotonic lysis with 0.2% saline. The isotonicity is quickly restored with an equal volume of 1.6% saline. A second hypotonic lysis removes many of the red blood cell (RBC) debris. In general, 1–2 × 10^6 neutrophils can be isolated per milliliter of whole blood from a normal subject with a normal white blood cell (WBC) count. All procedures are performed at room temperature, and the isolated cells are maintained in a balanced salt solution without divalent cations. The most common cell contaminants of the neutrophil preparation are eosinophils. Further purification of a standard neutrophil preparation with anti-CD16 magnetic immunobeads results in a neutrophil preparation that is generally ≥99% neutrophils. A second neutrophil isolation protocol that uses a discontinuous gradient of plasma/Percoll has often been used to minimize exposure of neutrophils to trace contamination by bacterial lipopolysaccharide (LPS) and reduce neutrophil priming.[18]

Isolated neutrophils are routinely frozen in aliquots of 5 × 10^6 cells/vial. For Western blot studies, neutrophils (1 × 10^6 cells/mL of buffer) are pretreated for 20 min with the cell permeant, irreversible serine protease inhibitor, diisopropylfluorophosphate (DFP, 1–5 mM). DFP is a volatile, potent neurotoxin that can irreversibly bind to and inactivate acetylcholinesterase and should be used with extreme caution. The cell suspension is then spun, and the supernatant fluid removed from the cell pellet before freezing. Waste solutions and disposable laboratory items should be flushed with sodium hydroxide to inactivate any residual DFP. These frozen neutrophil pellets, though not viable, can also be a source of DNA for genetic analyses.

KEY CONCEPTS

Estimated Yield From Whole Blood

In general, 1–2 × 10^6 neutrophils can be isolated per milliliter of whole blood from a normal subject with a normal white blood cell count.

HISTOCHEMICAL ANALYSIS OF NEUTROPHILS

Clinical Indications and Implications

Owing to their unique morphology, microscopic examination of neutrophil preparations with a differential stain (Wright stain) or a histochemical stain (Kaplow stain) to evaluate for myeloperoxidase, remains an essential element of neutrophil study, and can provide valuable insight into some genetic immunodeficiencies.

Principle and Interpretation of Laboratory Assessment

In Wright stain, the nucleus of a segmented neutrophil is normally multilobed (usually 3 to 5 lobes), and each lobe is connected by a narrow filament (Fig. 95.1, A). The nuclear chromatin is coarsely clumped with purple staining. Nucleoli are generally not present. In a band neutrophil, the nucleus is horseshoe shaped, with no indication of constriction into lobes. The pink-violet staining of the cytosol is associated with numerous, evenly distributed, specific granules; occasionally a dark-staining primary granule may be present. Kaplow stain[19] identifies the myeloperoxidase-containing primary granules as dark blue granules uniformly distributed throughout the cytosol (see Fig. 95.1, B). Neutrophils (and platelets) from patients with Chédiak-Higashi syndrome have giant primary granules that are pathognomonic for the disease (see Fig. 95.1, C).[20] Myeloperoxidase staining of Chédiak-Higashi neutrophils is very distinctive, with staining localized to the discrete giant primary granules (see Fig. 95.1, D). Neutrophils from a patient with specific granule deficiency exhibit primarily bilobed nuclei (pseudo–Pelger-Huët anomaly) with a paucity of specific granule staining in the cytosol (see Fig. 95.1, E).[21] Staining of the myeloperoxidase granules of neutrophils from a patient with specific granule

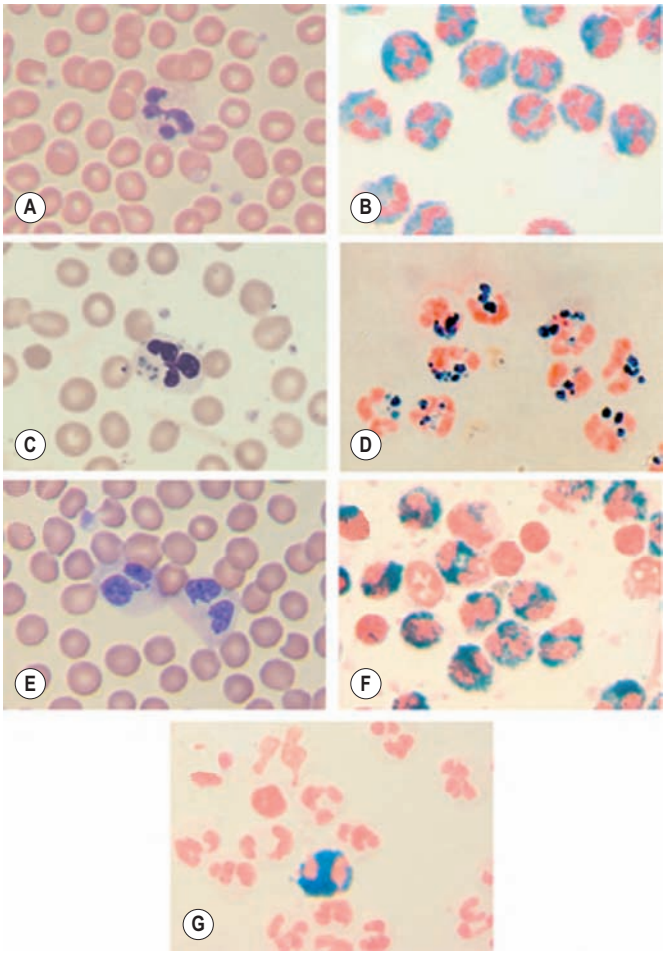


FIG. 95.1 Neutrophils stained with Wright stain and Kaplow stain. (A, C, and E) Blood smears stained with Wright stain. (B, D, F, and G) Neutrophil preparations stained with Kaplow stain. (A and B) Neutrophils from a normal individual. (C and D) Neutrophils from a patient with Chédiak-Higashi syndrome. (E and F) Neutrophils from a patient with specific granule deficiency. (G) Neutrophils from a patient with myeloperoxidase deficiency. The *blue positive-staining cell* is an eosinophil.

deficiency appears normal, since the defect is primarily associated with the granules (see Fig. 95.1, F). Neutrophils from a patient with myeloperoxidase deficiency fail to stain for myeloperoxidase.[22] However, eosinophil peroxidase stored in the eosinophilic granules continues to stain positive (see Fig. 95.1, G).

ANALYSIS OF GRANULE CONSTITUENTS

Clinical Indications and Implications

The granules of the neutrophils can be distinguished by their specific contents. Deficiency of only one granule constituent can be associated with a specific genetic defect, such as myeloperoxidase deficiency; alternatively, deficiency of multiple constituents of a certain granule can be associated with deficiency of an entire pool of granules, such as specific granule deficiency. Both enzymatic assays and immunoassays are available to determine the cellular content of many of these granule constituents.

Principles and Interpretation of Laboratory Assessment

The cellular content of neutrophils can be determined by solubilization of a neutrophil pellet with 0.2% Triton X-100, followed by sonication to disrupt the cells and generation of a homogeneous lysate. Analysis of the lysate using commercial immunoassays can identify deficiencies of certain granule contents. Diagnosis of myeloperoxidase deficiency can be confirmed by analysis of neutrophil lysates. Similarly, deficiency of both lactoferrin and neutrophil gelatinase (matrix metalloprotein-9 [MMP-9]) is indicative of specific granule deficiency,

NEUTROPHIL ADHERENCE

Clinical Indications and Implications

Adherence of neutrophils to the endothelium is a prerequisite step for migration of neutrophils into the tissues. Neutrophils isolated from patients with leukocyte adhesion defect-1 (LAD-1) who lack the common β_2-integrin subunit (CD18) exhibit abnormal adherence to the endothelium,[23] and therefore are not able to migrate efficiently into the surrounding tissues, often resulting in marked granulocytosis even in the absence of infection.[24] LAD-2 is a milder form of the disease, whereby patients exhibit a defect in fucose metabolism and glycoprotein biosynthesis.[25] Neutrophils from patients with LAD-2 exhibit abnormal expression of the glycoprotein, L-selectin, and fail to roll along the endothelium. However, they do exhibit normal β_2-integrin–mediated adherence.

Principles and Interpretation of Laboratory Assessment

Adherence of neutrophils can be assessed by measuring binding to plastic using a 96-well plate either uncoated or coated with fetal bovine serum or a specific extracellular matrix (ECM) protein (fibrinogen or fibronectin). Alternatively, endothelial cell monolayers harvested from human umbilical veins may serve as a more physiological substrate for the measurement of cell adhesion. Isolated neutrophils are preloaded with the cell permeant acetoxymethyl ester derivative of the fluorescent dye calcein (calcein-AM). Nonspecific esterases in the cytosol cleave the ester linkage, trapping the fluorescent probe in the cytosol. The labeled neutrophils are added to each well and incubated in the absence or presence of phorbol myristate acetate (PMA) to promote adherence through activation of the integrins. At the end of the incubation, the wells are washed three times to remove nonadherent cells. The fluorescence of each well is determined with a fluorescent microplate reader, and compared with the fluorescence of a control well with a fixed number of fluorescent cells. As shown in the left panel of Fig. 95.2, under controlled conditions, fewer than 10% of neutrophils adhere to plastic or to plastic coated with fetal bovine serum or fibrinogen. Treatment of normal neutrophils with PMA for 30 minutes results in the adherence of greater than 90% of the neutrophils under all conditions. This adherence assay is valuable in the diagnosis of patients with leukocyte adhesion deficiency. As shown in the right panel of Fig. 95.2, neutrophils isolated from patients with LAD-1 generally exhibit less than 5% adherence under controlled conditions and do not increase adherence after treatment with PMA.

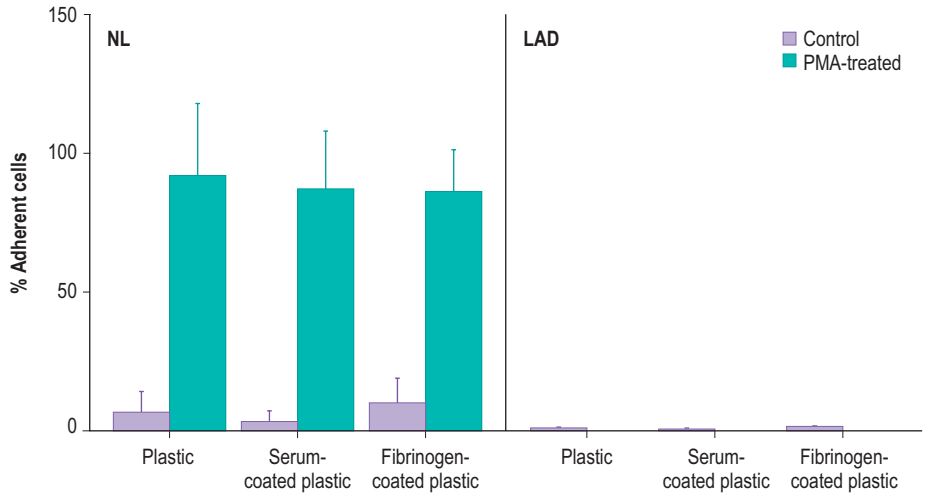


FIG. 95.2 Adherence of Neutrophils to Plastic: Normal versus Leukocyte Adhesion Deficiency. Neutrophils (1×10^7 cells/mL Hanks balanced salt solution [HBSS] without divalent cations) were preloaded with acetoxymethyl ester derivative of calcein (calcein-AM: 5 µg/mL) for 15 minutes at 37°C. The cells were washed twice and resuspended in HBSS/4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) with 2% bovine serum albumin (BSA) at a cell concentration of 2×10^6/mL. The wells of a 96-well plate were coated for 1 hour at 37°C with 32 µL of either buffer alone, fetal bovine serum, or fibrinogen (2.5 mg/mL). The wells were washed three times, and then cells were added to each well (160 µL/well, 3.2×10^5/well). After a 10-minute preincubation at 37°C, phorbol myristate acetate (*PMA*) (100 ng/mL) was added and the plate was incubated for 30 minutes at 37°C. The wells were then washed three times with HBSS/HEPES to remove nonadherent neutrophils. The percentage of adherent cells was determined by the ratio of the fluorescence of the well compared with the fluorescence of a known standard well. The panel (*NL*, normal) on the left represents the data (mean ± standard deviation [SD]) from 22 normal individuals, and the panel on the right represents the data from three patients with leukocyte adhesion deficiency (*LAD*).

NEUTROPHIL CHEMOTAXIS

Clinical Indications and Implications

Neutrophil migration is a prerequisite for neutrophil accumulation at sites of inflammation. Patients with leukocyte chemotactic defects usually show recurrent skin abscesses and occasional life-threatening invasive infections.

Principles and Interpretation of Laboratory Assessment

Chemotaxis in vitro is generally measured using a Boyden chamber. The Boyden chamber includes three components: a lower (chemoattractant) chamber, a nitrocellulose or polycarbonate filter layer, and an upper cell chamber. The lower compartment of the Boyden chamber is filled with a chemoattractant, such as formyl-methionyl-leucyl phenylalanine (fMLF; 10^{-8} M) or interleukin-8 (IL-8; 10^{-8} M). Alternatively, a rapid fluorescence-based measurement of neutrophil chemotaxis that uses a 96-well disposable chemotaxis chamber,[26] can be used and read in a fluorescence microplate reader. The lower chamber contains the chemoattractant and is separated from the cellular compartment by a filter. However, instead of a top chamber, the filter has a hydrophobic mask around each filter site that creates surface tension in the cell suspension, and aligns the suspension on the hydrophilic filter located directly above the chemoattractant chamber. Calcein-labeled neutrophils are placed on top of the filter. The chemotaxis chamber is incubated for up to 60 minutes at 37°C. Nonmigrating neutrophils atop the filter are rinsed off with buffer, and then the plate is read in a fluorescence microplate reader. The number of migrating neutrophils can be determined by comparing the calcein-based fluorescence to a standard well with a known number of fluorescent neutrophils. Less than 5×10^6 fluorescent neutrophils are needed to determine neutrophil chemotaxis using several doses of the chemoattractants (fMLF, IL-8, C5a, and leukotriene B_4). The advantages of this assay are high sensitivity, rapid acquisition and analysis of data, and reduced labor in loading the cell suspension. The 96-well format also allows for multiple comparisons to be made under identical conditions.

Imaging instrumentation is now available to monitor chemotaxis temporally. By acquiring digital images over time and analyzing those images using imaging software, the coordinates of individual cells can be determined. Changes in the distance (and velocity) directed toward the chemoattractant (directed migration) and orthogonal to the direction of the chemoattractant (random migration) can be determined. Tracks of multiple cells can be anchored at the origin and displayed graphically (see Fig. 95.3, top panels). Adding time as a dimension in the analysis of chemotaxis provides a mechanism to evaluate chemotactic and chemokinetic responses in neutrophils simultaneously, and to detect more subtle defects. Using buffer as a chemoattractant, the average cellular velocity vectors parallel and orthogonal to the direction of the chemoattractant are typically equivalent (see Fig. 95.3, bottom panel). When using a chemoattractant (e.g., fMLF, IL-8), typically there is marked increase in the average cellular velocity vector in the direction of the chemoattractant, but little change in the average cellular velocity vector orthogonal to the direction of the chemoattractant (see Fig. 95.3, bottom panel). Moreover, the ratio of the orthogonal vector to the vector in the direction of the chemoattractant equals $\tan \Theta$, where Θ equals the angle of migration, providing another useful parameter to assess the randomness of the migration.

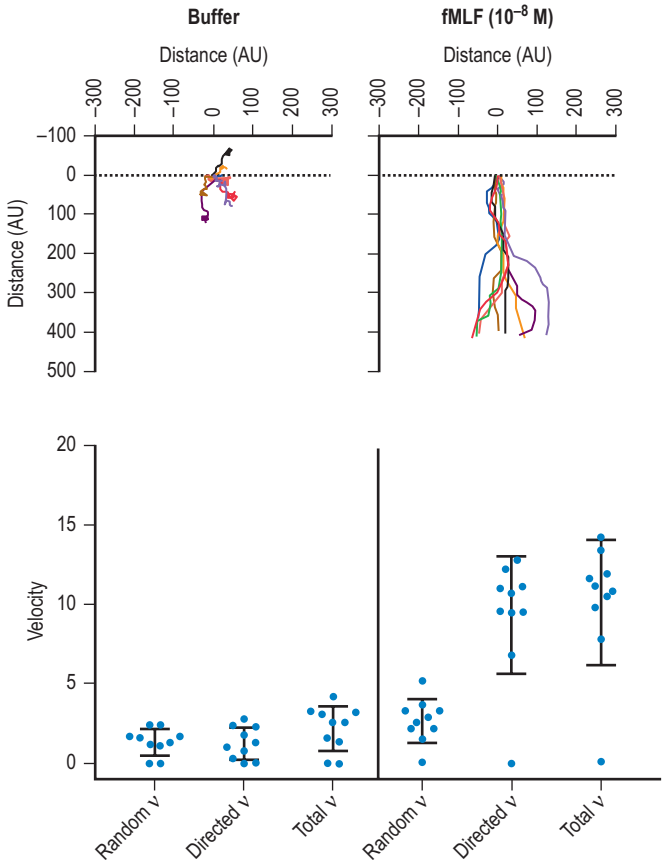


FIG. 95.3 Analysis of Chemotaxis. In the top panel, neutrophils (1 µL of 2 × 10^6 cells/mL in Hanks balanced salt solution [HBSS] with divalent cations) were added to the "Cell" well of EZ-TAXIScan and either buffer (left column) or formyl-methionyl-leucyl-phenylalanine (fMLF; 1 × 10^{-8} M, right column) was added to the "Chemoattractant" well. The cells were incubated for 60 minutes and images were collected every 2.5 minutes. Using the acquired images, 10 randomly chosen cells were electronically tracked and the paths of the cells plotted with their position at t = 0 anchored at the origin. Presented in the bottom panel are scattergrams of the average velocities of the individual cells that were tracked in the top panel.

EXPRESSION OF SURFACE ANTIGENS

Clinical Indications and Implications

The expression of neutrophil membrane antigens is altered in vivo during exudation or after challenge with intravenous endotoxin. Flow cytometric analysis of adhesion molecules on the neutrophil cell surface can indirectly reflect neutrophil adhesion function. Patients with LAD-1 exhibit a deficit in the expression of the common β_2 integrin CD18 that also results in deficiency in surface expression of CD11a, CD11b, and CD11c.[27]

Principles and Interpretation of Laboratory Assessment

The expression of cell-surface antigens is determined on neutrophils stained with specific fluorescent bound monoclonal antibodies (mAbs), and analyzed by flow cytometric analysis. Neutrophils stained with nonspecific isotype antibodies are used to determine the nonspecific background staining (Chapter 93). To determine the expression of circulating neutrophils and avoid artifacts induced by neutrophil isolation, an aliquot of whole blood can be stained with the appropriate antibody before lysis of the erythrocytes. During flow cytometric analysis, the neutrophils are easily differentiated using their forward light scatter and right angle light (side) scatter to gate on the neutrophil population. Since very little blood is needed (100 µL) to evaluate for each surface antigen, neutrophils can be stained with a panel of antibodies to many relevant surface antigens so that a more complete representation expression on neutrophils can be obtained. The panel should include the β_2 integrins (CD11a, CD11b, CD11c, and CD18); the selectin (CD62L); Fcγ receptors I, II, and III (CD64, CD32, and CD16); leukosialin (CD43); the common leukocyte antigen (CD45); and distinct surface markers for the granules—carcinoembryonic antigen-related cell adhesion molecule 8 (CEACAM8, or CD66b), a GPI-anchored glycoprotein family member stored in the specific granules, and lysosomal-associated membrane protein 3 (LAMP-3, or CD63), stored in the azurophilic granules. During exudation, the expression of CD11b and CD18 is increased over that observed in peripheral neutrophils, whereas the expression of CD43 (leukosialin) and CD62L is markedly reduced.

The antibody 7D5[28] recognizes an extracellular epitope of $gp91^{phox}$ and can be used to identify surface expression of $gp91^{phox}$ as well as monitor the mobilization of latent pools of $gp91^{phox}$ stored in the specific granules. Flow cytometric analysis of neutrophils stained with 7D5 can often be used to identify patients with X-linked chronic granulomatous disease (CGD - generally little or no 7D5 staining) and X-linked chronic carriers of CGD (mosaic pattern of staining), particularly in patients where the number of cells available for testing is limited. Flow cytometric analysis of permeabilized neutrophils stained with specific antibodies to either $p22^{phox}$, $p47^{phox}$, $p67^{phox}$, or $p40^{phox}$ has also been used to rapidly identify protein defects in patients with CGD.[29,30] These findings predict the use of permeabilized neutrophils to assess the expression of other intracellular proteins.

The expression of surface antigens can also be used to assess the responsiveness of neutrophils to ligands, such as fMLF and LPS. As shown in Fig. 95.4, neutrophils isolated from a patient who has a genetic defect in IL-1 receptor–associated kinase-4 (IRAK-4)[31] exhibit abnormal regulation of surface antigen expression to LPS but exhibit normal regulation of surface antigen expression to fMLF. Antigen expression can be upregulated because of translocation of latent antigen to the plasma membrane or downregulated due to internalization or shedding of the antigen.

NEUTROPHIL DEGRANULATION

Clinical Indications and Implications

The proteases, acid hydrolases, and inflammatory mediators released from storage granules in the neutrophils can mediate bacterial killing, tissue damage, healing, and immune regulation. Lactoferrin released from specific granules can chelate iron, resulting in a bactericidal or bacteriostatic effect. Elevation of plasma lactoferrin is an indication of intravascular activation and degranulation of neutrophils.

Principles and Interpretation of Laboratory Assessment

Stimulation of neutrophils with various secretagogues can result in the release of granular enzymes into the extracellular fluid. Treatment of the neutrophils with cytochalasin b (5 µg/mL) disrupts microfilament assembly and facilitates the release of both specific and azurophilic enzymes. Since stimulation of neutrophil degranulation is often accompanied by ROS generation

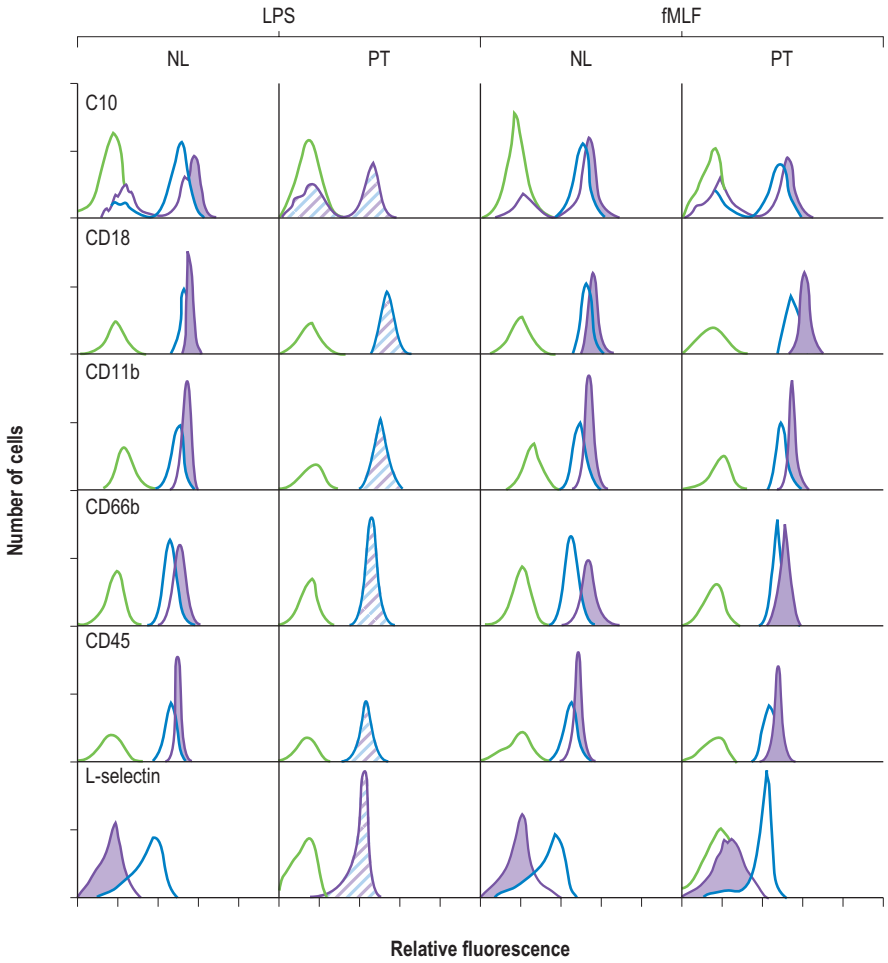


FIG. 95.4 Upregulation of Neutrophil Surface Antigen Expression. Neutrophils (2.5×10^6/mL Hanks balanced salt solution [HBSS] + 10% AB sera) isolated from a normal subject (*NL*) or from a patient with an interleukin-1 receptor–associated kinase-4 (IRAK-4) mutation (*PT*) were treated with either lipopolysaccharide (*LPS*; 100 ng/mL) or formyl-methionyl-leucyl-phenylalanine (*fMLF*; 0.1 μM) for 30 minutes at 37°C. The cells were washed and stained with C10 (an antibody that demonstrates neutrophil heterogeneity), CD18, CD11b (antibodies to β_2 integrins), CD66b (a specific granule marker), CD45 (the common leukocyte antigen), and L-selectin. The *green lines* represent the isotype control, *blue lines* represent control neutrophils, and *purple lines* represent stimulated cells. Differences between control and stimulated cells have been *shaded*. (From Kuhns DB, Long Priel DA, Gallin JI. Endotoxin and IL-1 hyporesponsiveness in a patient with recurrent bacterial infections. *J Immunol*. 1997;158:3959, with permission of the American Association of Immunologists, Inc.)

and oxidative inactivation of enzymes, both the cell supernatant fluid and the cell pellet should be analyzed to determine the percentage of enzyme released. To differentiate degranulation from cell lysis, release of the cytosolic enzyme lactate dehydrogenase should be monitored simultaneously.

The release of azurophilic granules can be assessed by determining the levels of myeloperoxidase or elastase. CD63 is also found in the membrane of azurophilic granules and migrates to the neutrophil surface after stimulation with fMLF in the presence of cytochalasin b. The release of specific granules can be assessed by determination of lactoferrin levels using an enzymatic immunoassay. The carcinoembryonic antigen CD66b is found on the neutrophil surface and the specific granules, and its expression on the surface of the neutrophils is increased after stimulation with fMLF or LPS. The secretory granules usually contain proteins that are translocated into the membrane from the cytosol during degranulation. Detection of the constituents of secretory granules can be assessed by flow cytometric analysis of the change in expression of surface proteins (adhesion molecules), and cytochrome b_{558} of the NADPH oxidase.

GENERATION OF REACTIVE OXYGEN SPECIES

Clinical Indications and Implications

The release of ROS, such as $O_2^{\bullet}$ and H_2O_2, is an important component of the bactericidal machinery of a neutrophil. Neutrophils isolated from patients with CGD have a defect in the NADPH oxidase and are unable to generate ROS, resulting in an O_2-dependent bactericidal defect. The production of ROS has become an important tool to perform risk assessment in patients with CGD. Patients with the lowest ROS generation (<1% of normal generation) have lower survival than patients with higher ROS generation (3%–10% of normal). Moreover, survival in CGD is a continuous function of ROS production, suggesting that therapeutic interventions that result in an increase in ROS generation should incur a survival benefit to patients with CGD.[32]

CLINICAL PEARLS

Reactive Oxygen Species (ROS) in Chronic Granulomatous Disease (CGD)

- Neutrophil ROS production, the primary determinant in diagnosis of patients with CGD, ranges from 0.1% to 27% of that observed in normal individuals.
- In addition, survival in CGD is strongly associated with residual ROS production as a continuous variable, independent of the specific protein defect.
- ROS production is an important and early indicator of overall risk in CGD.
- In addition, small increases (as little as 3%–5% of normal) in residual neutrophil ROS production may confer a survival benefit.
- Careful monitoring with detection of even small increases in ROS could be an important indicator of clinical efficacy during therapeutic intervention.

Principles and Interpretation of Laboratory Assessments

The nitroblue tetrazolium (NBT) test is a qualitative assay of ROS production. Either whole blood or isolated neutrophils are mixed with NBT in a chamber slide and stimulated with PMA for 15 to 30 minutes at 37°C. Once settled onto the slide, the neutrophils are air-dried, counterstained with 0.1% safranin, and examined under a microscope. The NBT test yields a visual record of the reduction of the NBT dye to the insoluble, blue-black deposits of formazan. Normal neutrophils, but not neutrophils from patients with CGD, reduce the yellow dye to black-brown-blue aggregates within cells. Owing to the random inactivation of the X-chromosome (lyonization), X-linked carriers of CGD exhibit both NBT^+ and NBT^- neutrophils. The percentage of NBT^+ neutrophils in X-linked carriers of CGD ranges from 5% to 95%. The drawback of the NBT test is the need for manual counting to obtain an accurate reflection of the percentage of positive cells. Presently the most common alternative to the NBT test is a flow cytometric assay using the dye dihydrorhodamine 123 (DHR-123).[33] Neutrophils are loaded with this non-fluorescent dye and then stimulated with PMA for 15 minutes at 37°C. The H_2O_2 produced oxidizes the dye and results in markedly increased fluorescence, which is detectable by a flow cytometer. The assay is also dependent on an endogenous MPO in the primary granules. Catalase is added to prevent cell-to-cell diffusion of H_2O_2. Since dye is localized to the cytoplasm and catalase is present in the extracellular fluid, the DHR-123 assay detects the intracellular production of ROS. As shown in Fig. 95.5, stimulation of normal neutrophils (see Fig. 95.5, A) with PMA results in a two-log shift in the fluorescence intensity. Neutrophils from an X-linked carrier of CGD (see Fig. 95.5, B) exhibit mosaicism with a negatively stained (abnormal) population and a brightly stained positive (normal) population. Neutrophils from a patient with X-linked CGD lacking $gp91^{phox}$ (see Fig. 95.5, C) express little increases in fluorescence. In addition, neutrophils from a patient with a deficiency in $p47^{phox}$ (see Fig. 95.5, D) exhibit slight increases in fluorescence. The major advantages of the DHR-123 assay are the sensitivity, the signal-to-noise ratio, and the ease of counting a larger number of cells. Moreover, it has been shown that the DHR-123 assay yields reliable results on ethylenediaminetetraacetic acid (EDTA) or heparin-treated blood samples that have been stored overnight. In general, more than 90% of the neutrophils from the control blood samples will exhibit increased DHR-123 fluorescence. For this same reason, however, overnight samples should not be used to rule out X-linked heterozygosity, since a highly lyonized CGD carrier (>90% normal vs. abnormal) could yield similar results.

The production of $O_2^{\bullet}$ can be detected using the reduction of cytochrome c. Because $O_2^{\bullet}$ causes a one-to-one stoichiometric reduction of ferricytochrome c to ferrocytochrome c, the resultant increase in the absorption spectrum at 550 nM can be used to quantify the production of $O_2^{\bullet}$. Superoxide dismutase is added to an identical tube to control for the nonspecific reduction of cytochrome c. However, since cytochrome is not permeable to the cells, the detection of $O_2^{\bullet}$ is limited to that released in the extracellular milieu. Neutrophils isolated from normal volunteers produce 0.42 ± 0.67 nmol/10^6 neutrophils/10 min under resting conditions; treatment with PMA results in production of 35.92 ± 11.92 nmol/10^6 neutrophils/10 min (Fig. 95.6). An estimate of normal $O_2^{\bullet}$ production over 60 minutes can be obtained by reducing the number of neutrophils in the assay to 2×10^5. Neutrophils isolated from patients with CGD produce little, if any, $O_2^{\bullet}$ in response to PMA treatment in 10 minutes (see Fig. 95.6). However, some patients with autosomal forms of CGD have low, but detectable $O_2^{\bullet}$ production in 60 minutes. Neutrophils isolated from X-linked heterozygous carriers of CGD can yield a full spectrum of $O_2^{\bullet}$ production, whereas neutrophils from autosomal recessive carriers of CGD generally yield a normal response (see Fig. 95.6). Although the detection of $O_2^{\bullet}$ by reduction of cytochrome c is useful in the diagnosis of patients with CGD, it cannot be used in the diagnosis of carriers because of the wide spectrum of responses that result from the degree of X-chromosome lyonization.

Studies have shown that $O_2^{\bullet}$ determinations sufficiently reliable to diagnose chronic granulomatous disease (CGD) can be obtained from neutrophils isolated from heparinized whole blood that has been stored overnight. Hence, analyses can be performed on blood samples shipped overnight. A normal control blood sample should accompany the sample to ensure adequate handling during shipment. By 48 h of storage, however, there are marked reductions in the phorbol myristate acetate (PMA) response, and the findings are no longer valid.

An alternative assay to measure reactive oxygen species (ROS) production is luminol-enhanced chemiluminescence. This versatile assay, in addition to its quick and easy setup, offers the capability to test several individuals and stimuli on the same plate using reduced number of cells while providing high sensitivity. The luminescence is read every 1–5 min for up to 2 h, and data are expressed as relative light units (RLUs). Different stimuli exhibit different kinetics (e.g., fMLF induces a rapid respiratory burst that decays quickly, whereas PMA [20–100 ng/mL] induces a peak of luminescence within 5–15 min that slowly decays by 2 h). Results from normal individuals and patients can be assessed simultaneously and monitored kinetically or using the area under the curve (AUC). Luminol-enhanced chemiluminescence is a measure of both intracellular and extracellular ROS production, although it may not detect them with equivalent efficiency. Addition of superoxide dismutase significantly reduces both peak height and AUC after stimulation with PMA, suggesting that at least a portion of the response can be attributed to $O_2^{\bullet}$. Typically, patients with CGD who are deficient in $gp91^{phox}$ have little to no detectable luminescence in this assay; however, as observed with the ferricytochrome c assay, patients with CGD with an autosomal recessive deficiency in $p47^{phox}$ have detectable luminescence that becomes most evident at later readings (40–80 min) (Fig. 95.7).

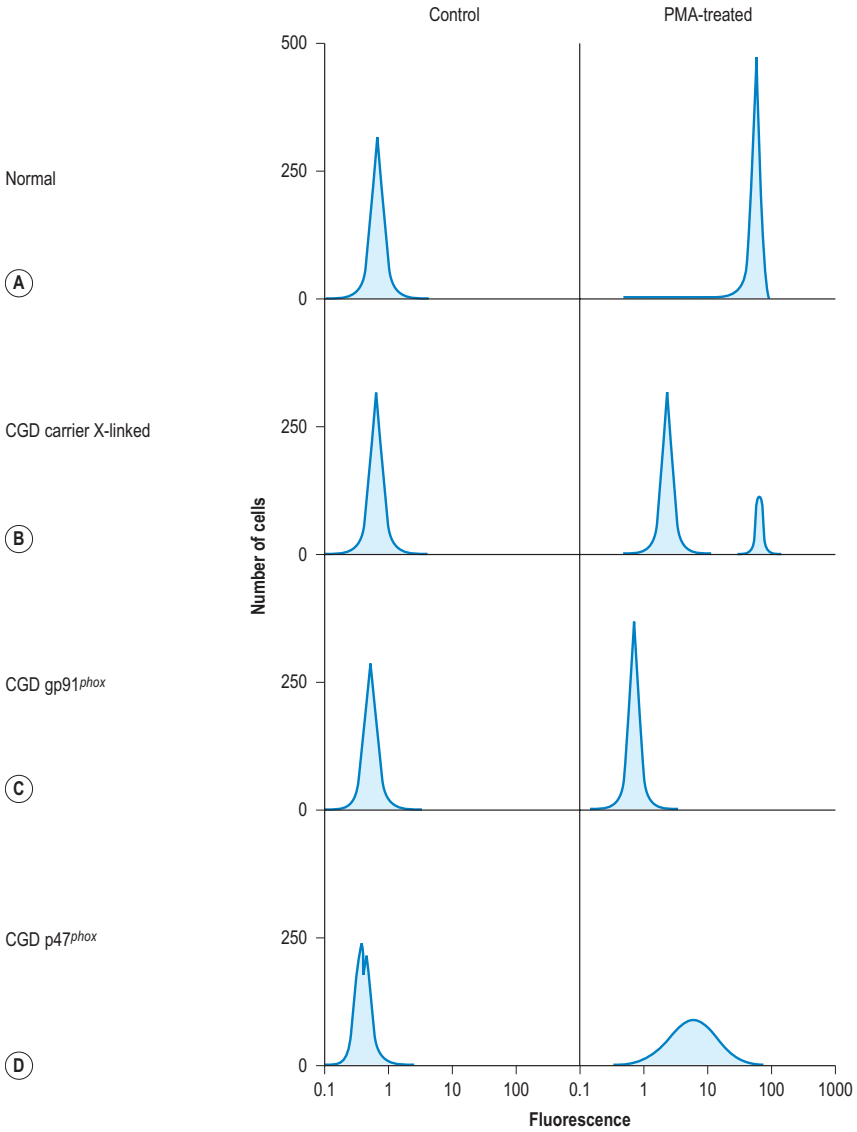
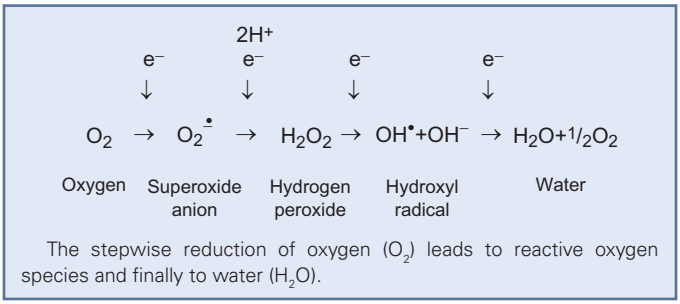


FIG. 95.5 Flow Cytometric Analysis of Dihydrorhodamine (DHR) Staining. Whole blood (1.2 mL with ethylenediaminetetraacetic acid [EDTA] as anticoagulant) collected from a normal individual (A), an X-linked chronic granulomatous disease (*CGD*) carrier (B), a gp91^{phox} CGD patient (C), and a p47^{phox} CGD patient (D) was lysed by using an ammonium chloride–potassium bicarbonate solution. The remaining leukocytes were resuspended in Hanks balanced salt solution (HBSS) and incubated with dihydrorhodamine 123 (DHR-123; 100 µM) and catalase (50 µg/mL) for 5 minutes at 37°C. The cells were then incubated an additional 15 minutes at 37°C with either buffer (control) or phorbol myristate acetate (*PMA*; 400 ng/mL). The cells were immediately analyzed by flow cytometric analysis. Neutrophils were gated using forward light scatter and right-angle light scatter. The analyses presented represent 5000 events within the gated area.

KEY CONCEPTS

Reactive Oxygen Species



The stepwise reduction of oxygen (O_2) leads to reactive oxygen species and finally to water (H_2O).

WESTERN BLOT ANALYSIS OF NADPH OXIDASE PROTEIN SUBUNITS

Clinical Indications and Implications

The NADPH oxidase consists of two membrane components (p22^{phox} and gp91^{phox}), three cytosolic components (p47^{phox}, p67^{phox}, and p40^{phox}), and several guanosine triphosphate (GTP)–binding proteins. CGD is characterized by defects in any of the following components—p22^{phox} (≈5% of patients with CGD), p47^{phox} (≈25% of patients with CGD), p67^{phox} (≈5% of patients with CGD), and gp91^{phox} (remaining 65% of patients with CGD).

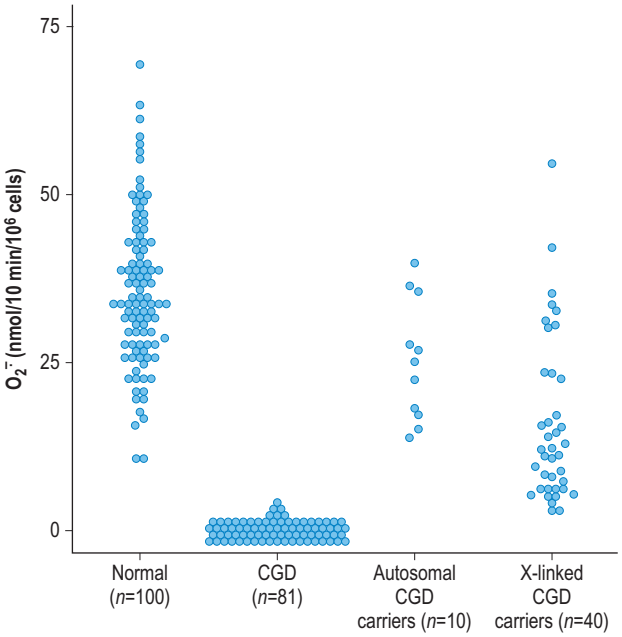


FIG. 95.6 $O_2^{\cdot-}$ Generation from Normal Subjects, Patients With Chronic Granulomatous Disease (*CGD*), and Carriers of CGD. Neutrophils (1 × 10^6/mL Hanks balanced salt solution [HBSS]) were incubated in the presence of 100 µM cytochrome c with phorbol myristate acetate (PMA; 100 ng/mL) for 10 minutes at 37°C. The reaction was terminated by centrifugation at 4°C. Reduction of cytochrome c was monitored at an analytical wavelength of 549.5 nm and a micromolar extinction coefficient of 0.0211. An identical tube containing superoxide dismutase (100 µg/mL) served as a blank.

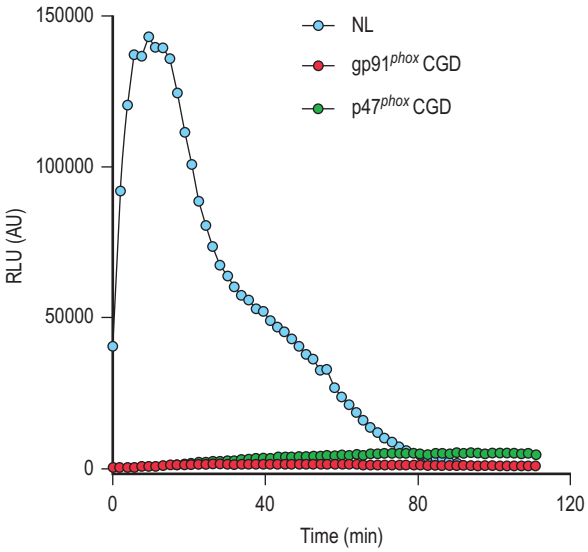


FIG. 95.7 Luminol-Enhanced Chemiluminescence. Neutrophils (1 × 10^5/200 µL) were preloaded with luminol (100 µM) for 10 min at 37°C. At t = 0, either buffer or phorbol myristate acetate (PMA; 100 ng/mL) was added and luminescence was monitored for 2 hours, with readings recorded every 2 minutes. Note that at later readings, neutrophils from a CGD patient with $p47^{phox}$ deficiency had increased luminescence compared with neutrophils from a CGD patient with $gp91^{phox}$ deficiency. *CGD*, Chronic granulomatous disease; *RLU*, relative light unit.

Principles and Interpretation of Laboratory Assessments

The severity of CGD can be related to the specific protein defect. Determination of the specific protein defect in CGD by Western blot analysis also provides direction for determination of the specific genetic defect and enables appropriate genetic counseling for the extended family. A validated normal control is included on each gel for band identification and intensity comparisons. In addition, a control sample from a patient with a known mutation in $gp91^{phox}$ CGD is included on each blot to ensure adequate development of $p22^{phox}$. Typical phox protein band patterns are presented in Fig. 95.8. Patients with CGD having mutations in $p47^{phox}$ are Western blot negative. Patients with CGD having mutations in $p67^{phox}$ are generally Western blot negative; however, we have analyzed one CGD patient with a missense mutation in $p67^{phox}$ yielding a positive Western blot. Because $p22^{phox}$ and $gp91^{phox}$ exist as a membrane complex, patients with a defect in $p22^{phox}$ are generally Western blot negative for both $p22^{phox}$ and $gp91^{phox}$. In contrast, defects in $gp91^{phox}$ yield more variable results, patients with nonsense defects in $gp91^{phox}$ generally exhibit low but detectable levels of $p22^{phox}$, while patients with missense mutations in $gp91^{phox}$ that yield detectable $gp91^{phox}$ protein exhibit proportionately higher levels of $p22^{phox}$. Neutrophils isolated from overnight samples can be used to diagnose $p47^{phox}$

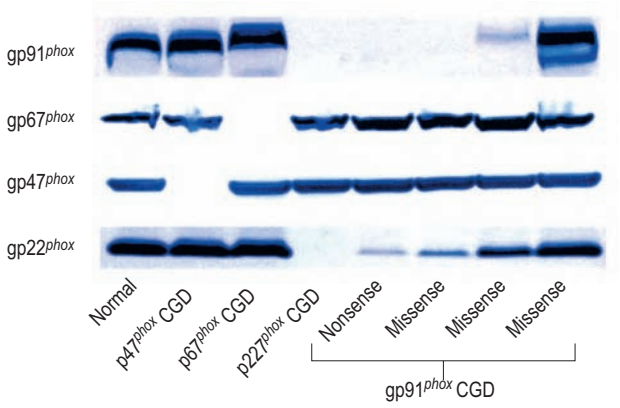


FIG. 95.8 Determination of Nicotinamide Adenine Dinucleotide Phosphate (NADPH) Oxidase Protein Defect by Western Blot Analysis. Frozen, diisopropylfluorophosphate (DFP)–treated neutrophil pellets (5 × 10^6 cells) were resuspended in polyacrylamide gel electrophoresis (PAGE) sample buffer and sonicated to break up the DNA. One million cell equivalents were loaded into each lane (10% PAGE gels for $p47^{phox}$ and $p67^{phox}$, 4% to 12% gradient PAGE gels for $gp22^{phox}$ and $gp91^{phox}$). A validated normal control was run on each gel for quality control. The gels were transferred to nitrocellulose, blocked with 5% powdered milk, and incubated overnight with specific antibodies to the phox proteins. The blots were washed, incubated with peroxidase-labeled secondary antibody, and developed with a color reagent. The blots were scanned for permanent storage, and a composite figure was created by using the relevant bands from each blot. The lanes are identified by the specific protein defect in CGD patients and, for $gp91^{phox}$, the type of mutation. *CGD*, Chronic granulomatous disease.

deficiency because of the stability of the protein. However, detection of other phox protein defects in overnight samples can be more problematic because of proteolysis of $p67^{phox}$ and the $gp91^{phox}$–$p22^{phox}$ complex.

CONCLUSIONS

Neutrophils are a critical cellular component of innate immunity in protecting the host from bacterial and fungal infections. They act as sentinels of the immune system and respond to the changing environment with their surface markers serving as a reflection of these changes. However, in other cells, the surface markers designate what function the cells have been programmed to perform (e.g., Th1 vs. Th2 lymphocytes, M1 vs. M2 monocytes) for neutrophils, surface antigen expression likely reflects their history. Neutrophils are exquisitely sensitive to perturbations, either physical trauma (e.g., shear, *g*-force, ionic stress) associated with cell isolation or physiological perturbations such as exposure to LPS or cytokines as well as transendothelial migration; such changes are reflected in their expression of surface markers, often due to mobilization of latent pools of the antigens stored in the granules (e.g., β_2 integrins), but also due to shedding of antigen by proteolysis (CD62L). These changes appear to be in preparation for the critical role that neutrophils play by internalizing and killing bacterial and fungal pathogens.

Current knowledge of neutrophil biology has developed from the study of genetic immunodeficiencies that target a specific protein and/or pathway in neutrophils. However, additional immunodeficiencies that result from defects in these recently discovered pathways are yet to be determined.

ON THE HORIZON

Developing a better understanding regarding the physiology and fate of neutrophils that survive at sites of inflammation/infection:

- Is the recycling of surface antigens permanent or can the receptors cycle back to their storage granules?
- Are antigens that have been shed replaced?

Some neutrophils appear to return to the vasculature from sites of inflammation:

- Is this reverse migration driven by chemoattraction?
- What are the surface signals that mediate their destination?
- What is the role of these altered neutrophils?
- Are these cells "fast-tracking" to their destination for destruction in the lungs, bone marrow, mucosal surfaces?
- Are the altered neutrophils so transient in circulation that they are lost within the greater population of normal circulating neutrophils?
- Does the presence of neutrophil granule proteins in plasma represent a vestige of their history, representing cells that have undergone NETosis or apoptosis?

ACKNOWLEDGMENT

This project has been funded in whole or in part with federal funds from the National Cancer Institute, National Institutes of Health, under Contract HHSN261200800001E. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the US Government.

REFERENCES

1. Jhunjhunwala S, Alvarez D, Aresta-DaSilva S, et al. Splenic progenitors aid in maintaining high neutrophil numbers at sites of sterile chronic inflammation. *J Leukocyte Biol*. 2016;100:253–60.
2. Bainton DF. The cells of inflammation: a general view. In: Weissman G, ed. *The Cell Biology of Inflammation*. 2nd ed. New York: Elsevier/North-Holland; 1980.
3. Pillay JP, den Braber I, Vrisekoop N, et al. In vivo labeling with ^2H_2O reveals a human neutrophil lifespan of 5.4 days. *Blood*. 2010;116(4):625–7.
4. Borregaard N, Cowland JB. Granules of the human neutrophilic polymorphonuclear leukocyte. *Blood*. 1997;89:3503–21.
5. Volpp BD, Nauseef WM, Clark RA. Two cytosolic neutrophil oxidase components absent in autosomal chronic granulomatous disease. *Science*. 1988;242:1295–97.
6. Matute JD, Arias AA, Wright NAM, et al. A new genetic subgroup of chronic granulomatous disease with autosomal recessive mutations in $p40^{phox}$ and selective defects in neutrophil NADPH oxidase activity. *Blood*. 2009;114:3309–15.
7. Ambruso DR, Knall C, Abell AN, et al. Human neutrophil immunodeficiency syndrome is associated with an inhibitory Rac2 mutation. *Proc Natl Acad Sci U S A*. 2000;97(9):4654–59.
8. Parkos CA, Dinauer MC, Walker LE, et al. Primary structure and unique expression of the 22-kilodalton light chain of human neutrophil cytochrome *b*. *Proc Natl Acad Sci USA*. 1988;85:3319–23.
9. Segal AW, Cross AR, Garcia RC, et al. Absence of cytochrome b_{-245} in chronic granulomatous disease—a multicenter European evaluation of its incidence and relevance. *N Engl J Med*. 1983;308:245–51.
10. Brinkmann V, Reichard U, Goosmann C, et al. Neutrophil extracellular traps kill bacteria. *Science*. 2004;303:1532–35.
11. Lood C, Blanco LP, Purmalek MM, et al. Neutrophil extracellular traps enriched in oxidized mitochondrial DNA are interferogenic and contribute to lupus-like disease. *Nat Med*. 2016;27:146–53.
12. Villanueva E, Yalavarthi S, Berthier CC, et al. Netting neutrophils induce endothelial damage, infiltrate tissues, and expose immunostimulatory molecules in systemic lupus erythematosus. *J Immunol*. 2011;187(1):538–52.
13. Carmona-Rivera C, Kaplan MJ. Induction and quantification of NETosis. *Curr Protoc Immunol*. 2016;115: 14.41.1-14.
14. Gavillet M, Martinod K, Renella R, et al. Flow cytometric assay for direct quantification of neutrophil extracellular traps in blood samples. *Am J Hematol*. 2015;90(12):1155–8.
15. Zuo Y, Yalavarthi S, Shi H, et al. Neutrophil extracellular traps in COVID-19. *JCI Insight*. 2020;5(11). e138999.
16. Kuijpers TW, Tool ATJ, van der Schoot CE, et al. Membrane surface antigen expression on neutrophils: a reappraisal of the use of surface markers for neutrophil activation. *Blood*. 1991;78:1105–11.
17. Böyum A. Isolation of mononuclear cells and granulocytes from human blood. Isolation of monuclear cells by one centrifugation, and of granulocytes by combining centrifugation and sedimentation at 1 g. *Scand J Clin Invest*. 1968;97:77–89.
18. Haslett C, Guthrie LA, Kopaniak MM, et al. Modulation of multiple neutrophil functions by preparative methods or trace concentrations of bacterial lipopolysaccharide. *Am J Pathol*. 1985;119:101–10.
19. Kaplow LS. Simplified myeloperoxidase stain using benzidine dihydrochloride. *Blood*. 1965;26:215–9.
20. Introne W, Boissy REB, Gahl WA. Clinical, molecular, and cell biological aspects of Chediak–Higashi syndrome. *Mol Genet Metab*. 1999;68:283–303.
21. Strauss RG, Bove KE, Jones JF, et al. An anomaly of neutrophil morphology with impaired function. *N Engl J Med*. 1974;290:478–84.
22. Nauseef WM. Diagnostic assays for myeloperoxidase deficiency. In: Quinn MT, DeLeo FR, Bokoch GM, eds. *Neutrophil Methods and Protocols: Methods in Molecular Biology*. Totowa, NJ: Humana Press; 2007:525–30.
23. Buescher ES, Gaither T, Nath J, et al. Abnormal adherence-related functions of neutrophils, monocytes, and Epstein-Barr virus-transformed B cells in a patient with C3bi receptor deficiency. *Blood*. 1985;65:1382–90.
24. Anderson DC, Schmalsteig FC, Finegold MJ, et al. The severe and moderate phenotypes of heritable Mac-1, LFA-1 deficiency: their quantitative

definition and relation to leukocyte dysfunction and clinical features. *J Infect Dis*. 1985;152:668–89.
25. Etzioni A, Frydman M, Pollack S, et al. Recurrent severe infections caused by a novel leukocyte adhesion deficiency. *N Engl J Med*. 1992;327:1789–92.
26. Frevert CW, Wong VA, Goodman RB, et al. Rapid fluorescence-based measurement of neutrophil migration in vitro. *J Immunol Methods*. 1998;213:41–52.
27. Anderson DC, Springer TA. Leukocyte adhesion deficiency: an inherited defect in the Mac-1, LFA-1, and p150,95 glycoproteins. *Annu Rev Med*. 1987;38:175–94.
28. Kawai C, Yamauchi A, Kuribayashi F. Monoclonal antibody 7D5 recognizes the R147 epitope on the $gp91^{phox}$, phagocyte flavocytochrome b_{558} large subunit. *Microbiol Immunol*. 2018;62(4):269–80.
29. Kuhns DB, Hsu AP, Sun D, et al. *NCF1* ($p47^{phox}$)-deficient chronic granulomatous disease: comprehensive genetic and flow cytometric analysis. *Blood Adv*. 2019;3(2):136–47.
30. Sacco KA, Smith MJ, Bahna SL, et al. NAPDH oxidase-specific flow cytometry allows for rapid genetic triage and classification of novel variants in chronic granulomatous disease. *J Clin Immunol*. 2020;40(1):191–202.
31. Kuhns DB, Long Priel DA, Gallin JI. Endotoxin and IL-1 hyporesponsiveness in a patient with recurrent bacterial infections. *J Immunol*. 1997;158:3959–64.
32. Kuhns DB, Alvord WG, Heller T, et al. Residual NADPH oxidase and survival in chronic granulomatous disease. *N Engl J Med*. 2010;363:2600–10.
33. Emmendörffer A, Hecht M, Lohmann-Matthes M-L, et al. A fast and easy method to determine the production of reactive oxygen intermediates by human and murine phagocytes using dihydrorhodamine 123. *J Immunol Methods*. 1990;131:269–75.

Assessment of Human Allergic Diseases

Robert G. Hamilton

Human allergic disease comprises a spectrum of immunoglobulin E (IgE)–mediated immediate-type hypersensitivity reactions that manifest as reactions in skin (urticaria or dermatitis), the respiratory tract (asthma, rhinitis, or sinusitis), eyes (conjunctivitis), the gastrointestinal (GI) tract (abdominal pain, bloating, vomiting, or diarrhea), and, in their most extreme condition, systemic anaphylaxis. These reactions are precipitated by exposure of a genetically predisposed and sensitized (IgE antibody-positive) individual to a variety of environmental substances that are ubiquitous and usually well tolerated by most healthy individuals. This chapter reviews the principles and performance characteristics of analytical methods used in the diagnosis and management of individuals with allergic disease. It examines in vivo and in vitro methods for the quantification of total and allergen-specific IgE and mast-cell tryptase.

BIOLOGICAL PROPERTIES OF IMMUNOGLOBULIN E

In 1921, Prausnitz and Küstner (PK)[1] reported that an intradermal (ID) injection of serum from an allergic individual into the skin of an unsensitized (nonallergic) individual, followed 24 hours later by injection of specific antigen into the same skin site, induced local itching and swelling surrounded by a zone of erythema. This passively transferred allergic reaction, or *PK reaction*, reached a maximum within 10 minutes, persisted for about 20 minutes, and gradually disappeared. In 1967, this antibody was identified as a fifth human immunoglobulin isotype and designated as IgE.[2–4]

Total serum IgE concentrations are the lowest of the five human immunoglobulin isotypes (0 to 0.0001 g/L; 0.004% of the total adult serum immunoglobulin).[5] Its short biological half-life of 1 to 5 days in peripheral blood is primarily the result of a relatively high fractional catabolic rate (71% of the intravascular pool catabolized/day). IgE does not pass the placenta or activate the classical complement pathway. Its reaginic (mast-cell sensitizing) activity is dependent on its ability to bind to the α chain of the high-affinity IgE Fc-ε receptor (α-FcεRI) that resides on the membrane surface of basophils and mast cells.

CLINICAL IMPORTANCE OF TOTAL SERUM IMMUNOGLOBULIN E

The concentration of IgE in the serum is age dependent.[5] Cord serum IgE concentrations are low, usually <2 kU/L (1 kU = 2.44 μg). Serum IgE levels progressively increase in healthy children up to the age of 10 to 15 years and gradually decline from the second to the eighth decades of life. Infants with atopy have an earlier and steeper rise in serum IgE levels during their early years compared with age-matched controls without atopia.[6]

Clinically, a patient's total serum IgE level should be evaluated against reference intervals established with sera from an age-stratified, healthy skin test-negative (nonatopic) population.[6] Many clinical laboratories define a total serum IgE > 100 kU/L (240 ng/mL) as a general demarcation into the atopic region.[7]. Extreme elevations in serum IgE are common in parasitic infections (Chapter 31) and are necessary for the diagnosis of the hyper-IgE (Job) syndrome (Chapter 39). Low total serum IgE levels support the diagnosis of nonallergic (intrinsic) asthma and help exclude allergic bronchopulmonary aspergillosis. There is extensive overlap between IgE levels in atopic and nonatopic populations,[6–9] which means that an elevated serum IgE can be useful in confirming the clinical diagnosis of allergic respiratory or skin diseases; however, a low or normal value does not eliminate the possibility of an IgE-mediated mechanism. Parasitic infections, selected immunodeficiency states (e.g., DOCK-8 deficiency, Job syndrome, Omenn syndrome, and Wiskott-Aldrich syndrome), cancer (Hodgkin disease, bronchial carcinoma), rheumatoid arthritis, liver disease, and atopic dermatitis (eczema) are other disease states that have been associated with a dysregulation of total serum IgE levels. The total serum IgE must therefore be interpreted within the relevant clinical context for each patient.

Because of the overlap between individuals with atopia and those without, total serum IgE measurements have been largely replaced in the routine diagnosis of allergic disease by the quantification of allergen-specific IgE antibody. However, quantification of total serum IgE has remained important for computing the therapeutic dose of anti-IgE. Omalizumab is a recombinant, humanized IgG1-κ monoclonal antihuman IgE Fc biological that specifically binds to the region on the ε heavy chain that interacts with α-FcεR1.[10] It is used to treat moderate to severe persistent allergic asthma and chronic idiopathic urticaria by blocking IgE binding to the α-FcεR1. The binding of omalizumab to IgE in vivo reduces both the number of free IgE molecules able to interact with the α-FcεR1 and the number of α-FcεR1 receptors on the surface of effector cells. The consequence is a reduction in mediator release and allergy symptoms following allergen exposure.

CLINICAL IMPORTANCE OF ALLERGEN-SPECIFIC IMMUNOGLOBULIN E

In contrast to total serum IgE, the presence of allergen-specific IgE antibody on the surface of circulating basophils or skin mast cells or in the serum is highly predictive of an individual's

propensity to exhibit an allergic response following re-exposure to that allergen. Before its identification as a novel immunoglobulin, IgE was only detectable with in vivo bioassays (skin test, bronchial or nasal provocation tests). Purification of IgE myeloma protein and the subsequent production of antisera specific for IgE led to the development of the first in vitro assay (radioallergosorbent test [RAST]) for the detection of allergen-specific IgE antibody in serum.[4,11,12] Since then, nonisotopic autoanalyzer variants based on the original noncompetitive cellulose paper disc solid-phase RAST designs have been widely used in clinical immunology laboratories worlwide.[13]

Historical studies have compared the diagnostic performance (sensitivity and specificity) of in vivo and the in vitro assays in the diagnosis of human allergic disease. These inter-method comparisons have shown that the presence of IgE antibody, as measured by serological immunoassay methods, usually agrees well with the presence of IgE detected in leukocyte and mast-cell histamine release assays, and with provocation tests, such as the skin test, food challenge, and inhalation provocation test.[14] However, these early studies emphasize that the presence of IgE antibody as detected either in vivo or in vitro is at best a confirmatory measurement for sensitization.[15] IgE antibody is necessary, but not sufficient, to identify an individual with allergic disease. IgE antibody presence is one important risk factor in the diagnosis of allergic disease that supports a patient's medical, family, and environmental histories of a temporal association between allergic symptoms and allergen exposure. The clinical importance of differences in diagnostic sensitivity between skin test and serological detection of IgE antibody may be less important for patients with allergies to inhaled (pollen, dust mite, and epidermal) allergens than in those facing life-threatening anaphylactic reactions caused by *Hymenoptera* stings and certain drugs. In these latter cases, skin tests are preferred to serology for the detection of allergen-specific IgE antibodies.[16] Immunoassays of allergen-specific IgE antibody in serum are, however, useful in cases where antihistamines, β-receptor stimulants, or high-dose steroids reduce an in vivo provocation test's measured response as well as in children, pregnant women, and older patients, in whom skin testing may not be well tolerated and when dealing with allergens (e.g., foods, molds) where commercial extracts can be highly variable or labile.[14]

CLINICAL HISTORY

The diagnosis of human allergic disease is driven principally by the patient's clinical history in which objective evidence is collected that an allergic reaction has occurred following exposure to a known or suspected allergen source. During the history, a number of factors need to be considered.[14] These include the patient's symptom characteristics (location, reproducibility, severity, duration, and delay time from appearance following allergen exposure), atopic factors (personal and family histories, age of onset, and the presence of infantile atopic dermatitis), opportunity for sensitization (geography, seasons, duration of exposure, prior exposures, employment, and hobbies), specificity of allergen triggers, and comorbidities (e.g., nasal polyps, recurrent sinusitis, and chronic obstructive pulmonary disease [COPD]). From the history, an a priori or pretest probability or likelihood of allergic disease is derived, and this determines whether or not confirmatory IgE antibody testing for sensitization is warranted.

DIAGNOSTIC METHODS

A combination of in vivo provocation and in vitro laboratory tests may be used to confirm sensitization and provide support for the clinical diagnosis of allergic disease. The actual tests selected depend on the nature of the disease process (e.g., allergic asthma, urticaria/angioedema, rhinitis/sinusitis, or anaphylaxis) and the suspected allergen triggers (e.g., aeroallergens, venoms, drugs, foods). Not only are there a myriad of assay methods, diverse techniques, reagents (extracts and molecules) and grading, interpolation, and interpretation methods, but most importantly, there is a general absence of gold standard methods for defining the presence of allergic disease. Consequently, results of the confirmatory tests need to be viewed as additional risk factors and tests for sensitization, rather than definitive indicators of disease.[15] In the end, the choice of which confirmatory test to use is a matter of clinical judgment.

INITIAL CLINICAL LABORATORY TESTS

Following the collection of a medical history and performance of a physical examination, the patient who is suspected of having allergic disease may undergo several preliminary blood tests. A complete blood count (CBC), and/or a total blood eosinophil count, if performed, should be obtained before any systemic corticosteroids or epinephrine is administered. A normal whole blood eosinophil level ranges from 0 to 500 cells/mm^3. Children generally have higher normal levels (mean 240 cells/mm^3 95% confidence interval [CI] = 0 to 740 cells/mm^3), with peak levels occurring at 4 to 8 years of age. Most clinical laboratories consider a differential white blood cell (WBC) count with an eosinophil proportion >5% to 10% of the total WBC count to be abnormal. Blood, sputum, and nasal secretion eosinophilia is characteristic of asthma, whether or not IgE-mediated allergic processes are present. In a bronchitic sputum specimen, neutrophils predominate. A neutrophilic nasal discharge is characteristic of sinusitis. Other laboratory tests that may be ordered include pulmonary function tests and a chest X-ray or sinus computer tomography (CT) scan.

IN VIVO PROVOCATION TESTING

Both the skin test and nasal/bronchial/GI provocation tests are useful in vivo diagnostic tools for the confirmation of sensitization in the evaluation for allergic disease. They can also help identify offending allergens in an allergy patient's work-up for avoidance, or management with pharmacotherapy, immunotherapy, anti-IgE therapy, or other biological therapy.

Skin Tests

Historically, Guerin and Watson[17] described a three-phase skin response during an immediate-type skin test reaction following the administration of a stimulus (allergen or histamine-positive control). First, a bluish-white area appears that involves the constriction of capillaries that typically disappears within minutes. Second, an erythematous peripheral halo or flare appears as a result of arteriole dilatation. Finally, a circular urticarial papule or wheal is observed, as a result of extravasation of plasma into the skin. The response is generally maximal by 15 to 20 minutes. The immediate "wheal and flare" reaction can be followed by a late-phase reaction 5 to 6 hours later that appears as a poorly defined edema-like reaction that usually disappears by

24 hours. An allergen extract can be administered either by a prick or puncture or by ID injection.[18]

Puncture skin testing involves placing a drop of each test allergen extract or control solutions (histamine and saline) on the skin of the forearm or back and the introduction of allergen into the epidermis by a needle puncture. Importantly, drops are spaced at least 2 cm apart to prevent cross-over contamination that can produce false-positive reactions or difficulty reading each discrete test site because of overlapping erythema. A variety of single-point (23 to 26 gauge), multipoint, and bifurcated needles have been used.[19] An immediate reaction (wheal and erythema) is read at 15 to 20 minutes as it reaches its maximum size. Because of the direct skin irritation with some crude allergen extracts, bleeding at the site of puncture can produce false-positive results.

The ID skin test is 1000 to 30,000 times more sensitive by concentration than the puncture skin test. A 0.02- to 0.05-mL volume of diluted allergen extract or controls (histamine or saline) is injected intracutaneously through a 26- to 27-gauge needle. Importantly, the bevel of the needle needs to face up and injection should be no deeper than the superficial layers of the skin. A 0.02-mL injection will initially produce a superficial 2- to 4-mm-diameter bleb. Like the puncture test, the ID skin test is read at 15 to 20 minutes, when the reaction is maximal. Dilutions of extract >1:1000 weight/volume (w/v) are commonly used to minimize false-positive reactions due to irritation and the potential for systemic reactions, which range from 0.02% to 1.4% of patients tested.[18] Subcutaneous administration of the allergen may lead to a false-negative result. The volume of allergen extract that is injected only slightly influences the size of the wheal-and-flare reaction, whereas concentration is the most important determinant of the final ID skin test result. ID testing allows an investigator to perform a skin test titration to quantitatively determine the patient's skin sensitivity. For serial titration, the same volume (e.g., 0.02 mL) of 3- to 10-fold serial dilutions of allergen extract are injected into different skin sites and the concentration of allergen required to produce a wheal or erythema of a defined mean diameter (e.g., 8-mm wheal) is interpolated.[20] The higher the concentration of allergen required to induce the defined size of wheal or erythema, the less sensitive is the patient to that allergen preparation and/or the lower is the allergenic potency of the extract.

Variables that Influence Skin Test Responses

The quality (composition, potency, heterogeneity) of the allergen extract is the single most important variable that affects skin test performance. Most skin test extracts are non-standardized, and their potency is reported in biological or weight per volume units.[21] Many allergen extracts used in puncture skin testing contain 50% glycerin, which enhances stability. However, glycerin causes skin irritation and false-positive skin test results if used intradermally without dilution. Other factors that influence the skin test response include the area of the body that is tested (back vs. forearm), age of patient (skin wheals increase in size from infancy to adulthood), race (clarity of reading on dark vs. light skin pigmentation) and pre-administered drugs (e.g., antihistamines, tranquilizers, or corticosteroids).

The saline negative control can identify dermatographic patients and trauma-induced reactivity produced by the puncturing device. The histamine positive control (for prick or puncture at 5.43 mmol/L or 1 mg/mL of histamine base) is useful in detecting medication- or disease-induced suppression of the skin test response. It is also used as a quality control reagent to document the reproducibility of technician performance.[22]

Relationship Between Puncture and Intradermal Skin Test Responses

Fig. 96.1 shows the relationship between the ng/mL level of *Dermatophagoides pteronyssinus* (*Dpt*, dust mite) specific IgE antibody in sera from 30 subjects with dust-mite allergy, as measured in serum by an in vitro assay, as well as the ID skin test midpoint *Dpt* allergen extract titer required to produce an 8-mm wheal in the same individual. Using the same *Dpt* extract in both tests, a higher degree of skin sensitivity (i.e., lower titer of antigen required to induce an 8-mm wheal) was strongly correlated ($r^2 = 0.77$; $P < .001$) with higher serum allergen-specific IgE antibody levels in those with the higher levels of skin sensitivity (<10 ng/mL midpoint). Fig. 96.2 shows the strong correlation between the wheal size that is observed in the same patients with *Dpt* allergy receiving the same dust-mite extract by a single puncture skin test and a midpoint ID skin test titration. The maximal diameter and the midpoint perpendicular diameter are averaged to generate an index. A permanent record of the skin reaction can be made by applying adhesive cellulose tape over the wheal-and-flare skin area,

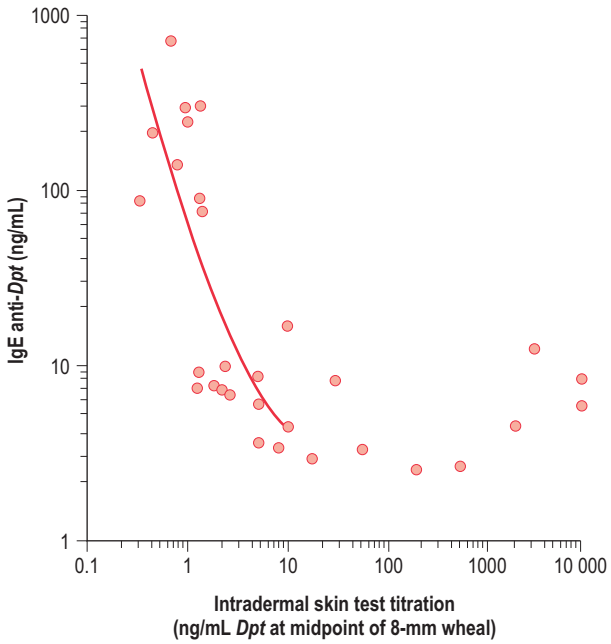


FIG. 96.1 Relationship between immunoglobulin E (*IgE*) anti–*Dermatophagoides pteronyssinus* (*Dpt*) measurements in skin (*x*-axis, intradermal [ID] skin test titration; ng/mL of *Dpt* required to produce an 8-mm wheal) and in the serum (*y*-axis, ng/mL of IgE anti-*Dpt* as measured by IgE antibody serology; sensitivity = 2 ng/mL). These results were obtained by testing the skin and serum of 30 individuals with dust-mite allergy and varying degrees of clinical sensitivity by using the same *Dpt* extract in both IgE antibody serology assay and the ID skin test titration study. A lower "titer" of antigen required to induce an 8-mm wheal (e.g., higher degree of skin sensitivity) was strongly correlated ($r = 0.77$; $P < .001$) with a higher serum IgE antibody level in individuals with the higher level of skin sensitivity (<10 ng/mL midpoint). Less sensitive patients (titers >10 ng/mL *Dpt*) had lower levels of serum antibody (2 to 15 ng/mL) that did not relate well with skin sensitivity.

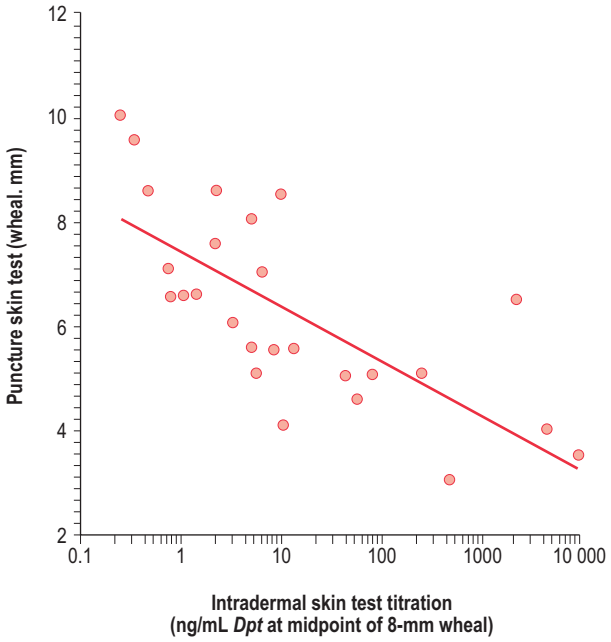


FIG. 96.2 The wheal size in millimeters at a single dose of *Dermatophagoides pteronyssinus* (*Dpt*) allergen administered in a puncture skin test compared with the titer or nanogram per milliliter (ng/mL) of the same *Dpt* allergen obtained in an intradermal (ID) skin test titration on the same 26 dust mite–allergic patients to produce an 8-mm wheal. These data indicate that the wheal size obtained with a single dose of allergen by the less labor-intensive puncture skin test is as predictive ($r^2 = 0.72$; $P < .001$) of relative patient sensitivity as the more technically complex ID skin test titration study, which involves the administration of seven increasing concentrations of the same allergen into different skin sites.

which has previously been outlined with a felt-tip or ballpoint pen. Using a single concentration of allergen, the ID skin test can be graded according to one of several reported systems (Table 96.1).[18] Alternatively, a midpoint titer can be interpolated from a skin test titration, including 3- to 10-fold serial dilutions of the allergen extract. The strong relationship between the size of the intradermal erythema and wheal observed with the mean of 304 duplicate skin tests is shown in Fig. 96.3. This relationship is useful to know because the erythema is difficult to assess in many dark-skinned subjects.

Conjunctival, Bronchial, and Nasal Provocation Tests

Conjunctival, bronchial, and nasal provocation challenges are performed primarily as research procedures to identify a relationship between allergen exposure and a change in the patient's ocular, bronchial, or nasal physiology. Along with puncture skin tests, they are more specific but less sensitive than ID skin tests.

Bronchoprovocation studies with methacholine or histamine are the most commonly performed, particularly in the diagnosis of difficult cases of asthma. The bronchoprovocation procedure involves the administration of either methacholine or histamine via a calibrated nebulizer, starting at doses of 0.03 to 0.1 mg/mL and doubling the concentration up to 10 to 25 mg/mL. Pulmonary function is monitored after each dose. A positive response is typically defined as the concentration of agonist that results in a drop in the forced expiratory volume per second (FEV_1)

TABLE 96.1 Grading System for Puncture and Intradermal Skin Testing Using Histamine as a Reference[a]

Grade or Class	Wheal Size (mm)	Erythema Size (mm)
Skin Testing Grading System[a]		
0	No discernible wheal	
1+	≤3 histamine wheal	
2+	>3 histamine and <13 histamine wheal	
3+	=size of histamine wheal ± 1 mm	
4+	>13 histamine wheal and <23 histamine wheal	
5+	>23 histamine wheal	
Alternative Skin Test Grading System for Intradermal Skin Testing Only Involving Interpretation of Wheal and Erythema Responses[a]		
0	<5	<5
+/–	5–10	5–10
1+	5–10	11–20
2+	5–10	21–30
3+	10–15	31–40
4+	>15 with pseudopods	41–50

[a]Prick/puncture histamine (3–10 mg/mL); intradermal histamine (100 µg/mL). Modified from Norman PS. Skin testing. In: Middleton E, Ellis EF, Reed CE, eds. *Allergy: Principles and Practice*. 2nd ed. St. Louis: CV Mosby; 1982, with permission from Elsevier.

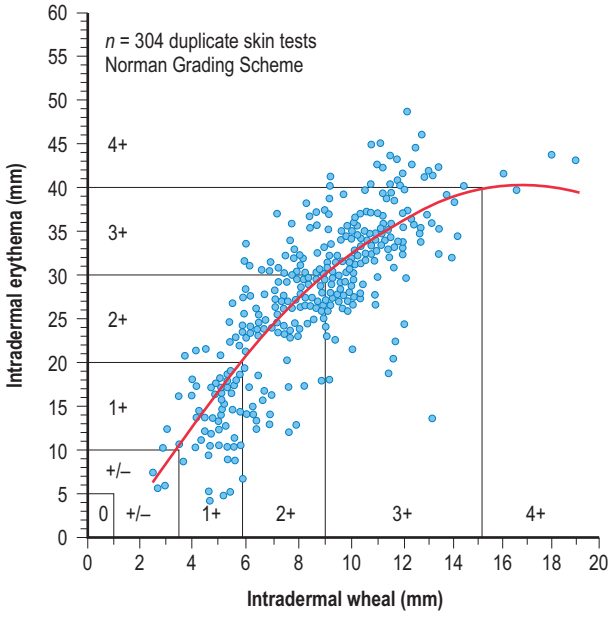


FIG. 96.3 Correlation plot of the mean wheal (*x*-axis) and erythema (*y*-axis) in millimeters for the mean of 304 duplicate intradermal (ID) skin tests to dust mite (*Dermatophagoides pteronyssinus*) obtained in a population of individuals with sensitivity to dust mites. The relationship is highly correlated ($r = 0.82$; $P < .001$) in grades 0 to 3+, indicating that either can be used to judge the degree of ID skin sensitivity. In highly allergic individuals (>35 mm erythema); however, the slope declines dramatically, indicating that wheal size may be more discriminating than erythema.

of 20% or more from the baseline. A baseline value >70% of predicted needs to occur to obtain a valid interpretation.[23]

Nasal provocation involves the controlled administration of buffer (human serum albumin–saline) or increasing concentrations of allergen into one or both sides of a washed nasal passage. The symptoms (e.g., itching, number of sneezes, rhinorrhea, and nasal conjugation) induced and/or concentration of mast-cell and basophil degranulation markers (e.g., histamine, tryptase, prostaglandin D_2, and cysteinyl leukotrienes) and cytokines (interleukin-4 [IL-4], IL-5, IL-10, and IL-13) released into nasal lavage fluids after each allergen dose indicates the relative level of sensitivity of the individual to that allergen. Nasal airway resistance is a less satisfactory endpoint because of high intrinsic variations.[24]

Oral Food Challenge Tests

Oral food challenges remain the definitive test for food hypersensitivity, since in vitro and other in vivo diagnostic methods suffer from suboptimal diagnostic specificity.[25] They are performed to establish an accurate diagnosis of food allergy or when skin tests are not possible because of atopic dermatitis or allergic GI disease or to assess the development of tolerance to a food. Passing a direct food challenge ideally permits reintroduction of the food into the patient's diet. Food-induced GI reactions (e.g., nausea, colic, vomiting, and diarrhea) can occur minutes to hours after the consumption of food allergens by a sensitized individual. Commonly eaten foods that contain potent allergens include cow milk, chicken egg white, cereal grains (wheat, rye, barley, oats), legumes (peanut, soybean, white bean), tree nuts, and fish and seafood.

Diagnosis of food allergy begins with a medical history, which includes an assessment of diet diaries and elimination diets. Skin tests and serological tests for IgE antibody to extracted food allergens and open food challenges with fresh and cooked foods are used to confirm sensitization to suspected foods. No evidence is available to indicate that food-specific IgG or IgG4 antibody levels have any diagnostic value.[26] The double-blind placebo-controlled food challenge (DBPCFC) is considered the definitive diagnostic test for food allergy.[25] The open food challenge is useful as a first-line challenge procedure because it is easier than DBPCFCs. Moreover, it is especially useful when the probability of a negative challenge is high, since a negative open challenge result can obviate the need for a DBPCFC and it reduces anxiety and the risk of bias in infants and young adults. An extensive discussion of open challenges, DBPCFCs, and variables that influence their outcome is presented elsewhere.[25]

In 2001, prospectively collected sera from 100 children and adolescents who had been previously evaluated by a skin test and DBPCFC were analyzed for food-specific IgE antibody.[27] Levels of IgE antibody specific for egg (7 kUa/L), milk (17 kUa/L), peanut (14 kUa/L), and fish (20 kUa/L) were identified, above which clinical reactivity could be predicted with more than 95% certainty. The study concluded that by measuring the concentration of food-specific IgE antibody, a subset of children who were highly likely (>95% probability) to experience a clinical reaction to egg, milk, peanut, or fish could be identified. This study and others subsequently have shown that judicious use of quantitative serological measurements for food-specific IgE antibody may be able to eliminate the need for time-consuming DBPCFCs in children who are suspected of having food allergy.

KEY CONCEPTS

Immunoglobulin E (Reaginic) Antibody Detection

- Allergen-specific immunoglobulin E (IgE) can be detected by the skin test using a puncture or intradermal (ID) administration of allergen or in the serum by laboratory-based immunoassays.
- In general, the ID skin test is more analytically sensitive than a puncture skin test, which is roughly comparable to the best in vitro methods for IgE antibody detection in serum.
- ID skin tests are the diagnostic procedure of choice in the work-up of patients with suspected *Hymenoptera* venom and drug allergy, while both in vitro an in vivo assay methods are complementary for evaluating aeroallergen-related disease.
- Serological analyses of IgE antibody specific for food allergens are often favored over extract-based skin test analyses in part because of more enhanced reagent quality control and pediatric acceptance. However, the double-blind placebo-controlled food challenge (DBPCFC) remains the gold standard for definitive diagnosis of food allergies.
- IgE antibody to allergenic molecules (components and epitopes) can in some cases (e.g., peanut and hazelnut) provide clarity in terms of the specificity of the patient's sensitization profile (genuine vs. cross-reactivity) and relative risk for mild vs. serious systemic reactions.

IN VITRO TESTING

Clinical immunology laboratories worldwide offer serological tests that are useful in the diagnosis and management of human allergic disease. Diagnostic analytes commonly measured in these laboratories include the total serum IgE, IgE antibodies to hundreds of allergen specificities, and mast-cell tryptase. IgG antibody measurements to allergens have not been shown to be predictive of protection. Basophil mediator and activation tests, although rarely offered as clinical tests because of the requirement for fresh blood, are useful investigational methods.

Total Serum Immunoglobulin E

Total serum IgE is currently the only diagnostic allergy analyte regulated under the US Clinical Laboratory Improvement Amendment of 1988 (CLIA-88). The minimum detectable concentration of commercial total serum IgE assays is 0.5 and 1 μg/L. The inter-method agreement of the different commercial IgE assays is excellent (e.g., inter-method coefficients of variation [CVs] typically <15%).[12,28] Nonatopic age-adjusted reference intervals for total serum IgE must be used for normative interpretation.[5]

Total Immunoglobulin E Measurements After Therapeutic Anti– Immunoglobulin E Administration

Omalizumab (anti-IgE) is used as a fourth therapeutic modality to supplement avoidance, pharmacotherapy, and immunotherapy in the management of persistent asthma and urticaria, and off-label for other IgE-mediated states (e.g., allergic bronchopulmonary aspergillosis, pretreatment of food allergy patients receiving immunotherapy). Since its conception, clinicians have desired to quantify total and "free" (uncomplexed) IgE levels in anti-IgE treated patients as a rationale for treatment failures or to justify modification of a patient's dosing regimen to maximize treatment success. A systematic evaluation of the impact of therapeutic anti-IgE on the performance of clinically used total IgE assays showed variable interference that resulted in a 1.9% to 51.9% reduction in accuracy, depending on the assay.[29] Accurate quantitation of the level of uncomplexed or "free" IgE in the serum of treated patients has been more technically difficult and thus is not recommended.[40]

Allergen-Specific Immunoglobulin E

Laboratories in the United States that perform clinical diagnostic allergy testing use assays that have been through rigorous validation, and have achieved unsurpassed intra-assay precision, inter-assay reproducibility, and a high degree of quantification.[12] Their basic design can be traced to the first IgE antibody assay, the RAST, reported by Wide et al. in 1967.[11]

Allergen

The most highly variable component of the IgE antibody assay is the allergen-containing reagent. Allergens are mixtures of molecules, typically proteins, glycoproteins, lipoproteins, or protein-conjugate chemicals or drugs that have been solubilized from a defined source, a portion of which can elicit an IgE antibody response in exposed and genetically predisposed individuals. They possess common properties of stability to heat and digestion because of multiple cysteine linkages. Allergens tend to be abundant in nature, form aggregates or polymers, commonly interact with lipid structures, and serve to defend their biological source. Cross-tabulation of the protein family (PFAM) database ($n = 16{,}230$ protein families) with the Structural Database of Allergenic Proteins (SDAP) identified 130 PFAMs in the Allergenic Family database of allergenic proteins. Thus, importantly, allergens comprise a small fraction of protein families with particular structures and biological functions.[30] The Clinical Laboratory Standards Institute (CLSI) has an established international guidance document that defines the expected performance characteristics of allergenic materials used in immunological assays for human IgE antibodies.[12] It provides a compendium of the genus and species of the allergen specificities of clinical interest, subdividing them into extract and component allergens. They are categorized based on their source, into weed pollen, grass pollen, tree pollen, animal dander, mold, house dust-mite fecal material, parasites, insect venoms, occupational allergens, foods, and drugs. Except for drugs, these extracts are complex heterogeneous mixtures that contain both nonallergenic and allergenic proteins. Some allergens share structural similarity or cross-reactive epitopes, and others possess unique IgE antibody-binding determinants. Extensive allergenic cross-reactivity has also been documented within pollen groups, such as the grasses (June, Brome, Timothy, Perennial Rye, Fescue, Orchard, Red Top, Salt, Sweet Vernal, Velvet). Conversely, other grass pollens, such as those produced by Bermuda Grass, Johnson Grass, and cultivated corn, oat, and wheat, are minimally cross-reactive (allergenically distinct). Variations in the allergenic content of extracted source materials, the extraction process from the raw source material, allergen-reagent manufacturing methods, differential binding to various allergosorbent supports, instability during storage, heterogeneity of internal reference allergen standards, and differences in characterization procedures (antisera, assays) make the production of reproducible allergens for in vitro use a challenge.

Cross-reactivity has also been shown at the allergen component level. There are 10 principal allergen families that show structural similarity and extensive cross-reactivity (Table 96.2).[31,32] The most prominent allergen component family is the pathogenesis-related (PR)–10 family of allergens, also known as the *Bet v 1 homologues*. These are small (17-kDa) proteins in many plant species that transport steroids and exhibit low stability at high pH and in the presence of digestive enzymes. The first allergen in this family was identified from birch pollen (Bet v 1). Others with high amino acid sequence homology include Cor a 1-hazelnut, Mal d 1-apple, Pru p 1-peach, Gly m 4-soybean, Ara h 8-peanut, Aln g 1-alder, Act d 8-kiwi, Api g 1-celery, and Dau c 1-carrot. Other component-based cross-reactivity groups include the profilins, nonspecific lipid transfer proteins, tropomyosins, serum albumins, polcalcins, lipocalins, parvalbumins, storage-binding proteins, and carbohydrate cross-reactive determinants (CCDs) (see Table 96.2). Each of these cross-reactive allergen families is extensively discussed in the *Handbook on Molecular Allergology*.[32]

Calibration

The second attribute that varies among commercially available allergen-specific IgE antibody assays is their calibration algorithm and methodology. Since no internationally recognized polyclonal human IgE antibody reference preparation exists, autoanalyzers in current commercial use employ a heterologous interpolation procedure in which allergen-specific IgE antibody response data are interpolated from a total serum IgE calibration curve. This procedure is valid as long as the total IgE calibrators and the patient's allergen-specific IgE antibody levels dilute out in parallel so that parallelism is maintained. The assays report IgE antibody levels in kUa/L units, using internal total IgE calibrators that are cross-verified and traceable to the World Health Organization (WHO) third IgE international reference preparation. This calibration system allows interpolation of IgE antibody results from a limit of quantitation of 0.1 kUa/L to 100 kUa/L levels of IgE antibody. In terms of quantitation, at least one of the IgE antibody autoanalyzers (ImmunoCAP) has demonstrated equivalence in which 1 kUa/L of chimeric allergen-specific IgE antibody was shown to be equivalent to 1 kU/L (2.4 ng/mL) of total serum IgE.[33]

Single-Plex, Multiallergen, and Multiplex Assays

The autoanalyzers in worldwide clinical use are single-plex assays in which one analyte is measured in a single analysis. In contrast, a multiplex antibody assay allows many specificities of a single antibody isotype to be individually detected and semiquantified in a single analysis. These are distinguished from a multiallergen screening assay, which is a form of single-plex analysis that involves the use of a mix of multiple allergens that is immobilized on the same allergosorbent. The purpose of the multiallergen assay is to simultaneously screen a serum for IgE antibody to a concise number (e.g., 10 to 15) allergens either of the same allergen source type (e.g., foods: chicken egg, cow's milk, peanut, soybean, cod fish) or diverse sources (e.g., respiratory allergens as an aeroallergen mix: pollen from select trees, grasses, weeds, pet epidermals, dust mites, molds).[34] A qualitative result (positive or negative) is generated, and it serves as an efficient single analysis to assess the general atopic status (IgE positivity) of an individual. Clinically, the multiaeroallergen screening assay has a high negative predictive power and thus is used to rule out IgE-mediated allergic disease where the suspicion based on the clinical history is weak.

The availability of unlimited quantities of molecular allergens has made it possible to develop multiplex chip-based microarrays for diagnostic allergy confirmatory testing. The present-day microarray for semiquantification of IgE antibody involves a preactivated glass slide (chip) on which 112 purified allergens are each immobilized in triplicate microdot arrays.[35] IgE anti–cow's milk components as measured in the multiplex immune solid phase chip or ISAC has agreed well with those obtained in a single-plex autoanalyzer using the same allergen specificities.[36]

TABLE 96.2 Principal Allergen Families and Their Associated Biological Functions

Family	Function	Diagnostic Utility and Clinical Features	Examples
Profilins	Actin-binding proteins involved in the dynamic turnover and restructuring of the actin cytoskeleton. Highly conserved, extremely cross-reactive and ubiquitous proteins in pollen and plant foods	Positive skin test and immunoglobulin E (IgE) responses to (nonrelated) pollen species (often including grasses) and plant food extracts are indicative for IgE cross-reactivity to profilins. After proof of sensitization to one profilin, pollen and plant food extracts are of no further use because of their subsequent lack of analytical specificity	Bet v 2 (birch); Phl p 12 (Timothy Grass); Hev b 8 (Natural Rubber Latex); Mal d 4 (apple)
Serum albumins	Highly cross-reactive plasma protein carriers involved in transport of hormones, enzymes, hemin, and fatty acids. Also maintains oncotic pressure	Positive skin test and IgE responses to furred animals indicate IgE cross-reactivity to mammalian serum albumins. Proof of sensitizations to one serum albumin can explain clinical symptoms to rarely or uncooked meat ("cat-pork-syndrome")	Fel d 2 (cat); Can f 3 (dog),
Pathogenesis-related proteins family 10 (PR-10), Bet v 1- homolog	Plant defense proteins; Bet v 1 is a quercetin-3-O-sophoroside-binding molecule (17 kDa); inflammation response proteins	Positive Bet v 1-specific IgE reflects sensitization to fagales tree pollen (i.e., hazel, alder, birch, beech, oak). Thermo and digestion labile Bet v 1-homologs in fruits, legumes, and vegetables. Bet v 1-sensitized individuals can cause predominantly oropharyngeal symptoms after consumption of raw food items (i.e., pip fruits, stone fruits, tree nuts, carrots, soy).	Bet v 1 (birch); Cor a 1 (hazelnut); Mal d 1 (apple); Gly m 4 (soy)
Polcalcin	Cross-reactive, ubiquitous calcium-binding proteins in pollen; involved in calcium regulation	Positive skin test and IgE responses to (nonrelated) pollen species are indicative for IgE cross-reactivity to polcalcins. After proof of sensitization to one, polcalcin pollen extracts are of no further use because of their subsequent lack of analytical specificity	Phl p 7 (Timothy Grass); Bet v 4 (birch); Amb a 10 (Short Ragweed)
Nonspecific lipid transfer proteins (nsLTP, PR-14)	Inflammation response proteins; responsible for shuttling phospholipids and other fatty acids between cell membranes	Primary food allergen after sensitization to peach lipid transfer protein (LTP) (possibly through the skin?); predominantly found in the Mediterranean subjects; variable degree of cross-reactivity between the thermally stable LTPs in fruits and vegetables frequently causing oropharyngeal and sometimes systemic symptoms (i.e., exercise-induced)	Pru p 3 (peach); Ara h 9 (peanut); Cor a 8 (hazelnut)
Lipocalins	A highly heterogeneous group of extracellular proteins within various subfamilies involved in transport of small hydrophobic molecules, such as steroids, bilins, retinoids, and lipids	Positive skin test and IgE responses to more than one or many furred animals are indicative for IgE sensitization and subsequent serological cross-reactivity to lipocalins of a certain subfamily. As a consequence, animal extracts are of no further use because of their subsequent lack of analytical specificity	Fel d 4, 7 (cat); Can f 1, 2, 4, 6 (dog)
Parvalbumins	Calcium-binding proteins; localized in fast-contracting muscles and being involved in calcium signaling	Thermostable and digestion-stable major fish allergen with fairly high, but not complete cross-reactivity and high abundance in almost all fish species. Only limited sequence homology and no cross-reactivity between fish and shellfish calcium-binding proteins	Gad c 1 (cod); Cra c 4, 6 (shrimp)
Tropomyosins	Integral components of actin filament that play a role in regulating muscle contraction. Also regulate actin filament stability in nonmuscle cells	Thermostable and digestion-stable major shellfish allergen with broad cross-reactivity and high abundance in all shellfish species. IgE sensitization is associated with allergic reactions to various panels of shellfish species	Der p 10 (dust mite); Pen m 1 (shrimp)
Seed storage proteins	Heterodimeric, stable and highly abundant proteins involved in nutrient storage: e.g., 2S albumins, 7/8S globulins (vicilins), 11S globulin (legumins)	Thermostable and digestion-stable important primary food allergens in legumes (peanut, soy), tree nuts (hazelnut), capsule fruits, and seeds with limited cross-reactivity within the subfamilies of different species; primary sensitization starts in early childhood and can persist lifelong; high allergen-specific IgE levels are associated with systemic allergic reactions	Ara h 2 and 1 and 3 (peanut); Cor a 14 and 11 and 9 (hazelnut)

Modified from the I/LA20 Guidance Document from the Clinical Laboratory Standard's Institute. Hamilton RG, Matsson PNJ, Chan S, et al. *Analytical Performance Characteristics, Quality Assurance and Clinical Utility of Immunological Assays for Human Immunoglobulin E (IgE) Antibodies of Defined Allergen Specificities*. 3rd ed. I/LA20-A3, International CLSI-Guideline. Wayne, PA: Clinical Laboratory Standards Institute; 2016.

Other groups have tried alternative multiplexing technologies to detect IgE antibodies that have included immobilizing allergen extracts on chips by using Luminex bead–based suspension arrays, employing nanotechnology biosensors, detecting surface plasmon resonance, and using plates equipped to produce electrical pulse–generated chemiluminescence. Single-plex systems have the advantage of greater analytical sensitivity or a lower limit of quantitation, greater precision and accuracy, more established internal and external quality control, and wider global availability of technology. In contrast, multiplex systems provide increased speed of analysis with reduced turnaround times, conservation of sample volume, greater simplicity, and reduced technical and reagent costs.[12]

The quality of allergen-specific IgE antibody results reported from clinical diagnostic allergy laboratories is not uniformly equivalent. In addition to the variability of results for a given serum between assays from different manufacturers as a result of allergen and calibration variance, the positive thresholds used for the same assay by different laboratories varies (e.g., 0.1, 0.35, and 0.7 kUa/L). Physicians requesting IgE antibody

testing thus bear some responsibility for determining the quality of the results they receive. In the United States, testing should be performed in a clinical laboratory that is federally licensed for highly complex immunology clinical testing under CLIA-88 (verified by requesting a copy of the federal laboratory license). The requesting physician should inquire about the assay method used, the source of its reagents, and how assays are quality controlled by the laboratory. As part of the formal record, the assay method used in patient analysis should be indicated on the final report.

Allergen-Specific Immunoglobulin G

Allergen immunotherapy is known to enhance the production of specific IgG "blocking" antibodies.[37] Quantitative measurements of allergen-specific total IgG or IgG subclass antibodies in studies of allergic rhinitis have not correlated with the control of clinical symptoms in individual patients. However, clinically successful immunotherapy is usually accompanied by high serum levels of allergen-specific IgG (typically IgG1 and IgG4) blocking antibodies. Despite a decrease to pre-immunotherapy baseline levels after 2 years of immunotherapy discontinuation, overall functional inhibitory activity, as measured by an IgE-dependent facilitated allergen binding assay, appears to be maintained.[37] For patients with *Hymenoptera* venom allergy, specific IgG antibody measurements have been used as an indicator of effective immunotherapy. Quantitative venom-specific IgG antibody levels may be useful in individualizing the dose and frequency of injections while maximizing the protective effects. However, their clinical utility may be restricted to the first 4 years of venom immunotherapy.[38] In contrast, the presence or levels of IgG antibodies specific for food antigens have shown no correlation with the results of positive DBPCFCs, and they are not indicated in the diagnostic work-up of a patient with suspected food allergy.[26]

CLINICAL PEARLS

Immunoglobulin G Antibody Measurements

- Clinically successful aeroallergen immunotherapy is almost always accompanied by high serum levels of allergen-specific immunoglobulin G (IgG; predominantly the IgG1 and IgG4 subclasses).
- Quantitative venom-specific IgG antibody levels can be of value in individualizing venom doses and frequencies for patients on venom immunotherapy for up to 4 years.
- Food-specific IgG and IgG4 assay results do not correlate with the results of double-blind placebo-controlled food challenges and are not clinically indicated.

Mast-Cell Tryptase

Mast cells have been identified in skin, respiratory, and digestive tract connective tissues and distinguished on the basis of the neutral proteases present in their secretory granules. One group of mast cells contains only tryptase, whereas the other contains both tryptase and chymase.[39] Mast-cell tryptase (MW 134 kDa) is a serine esterase with four subunits, each having an enzymatically active site. A resting mast cell contains 10 to 35 picograms (pg) of tryptase that is stored attached to heparin. When dissociated from heparin, it rapidly degrades into its monomers and loses enzymatic activity. As basophils have ~500-fold less tryptase compared with mast cells, elevated tryptase levels in serum are considered a relatively specific indicator of mast-cell involvement in a clinical reaction. Unstimulated tissue mast cells continually secrete immature protryptase into the tissue, and it diffuses into the circulation to provide a measure of total mast-cell number. α-Protryptase and β-protryptase represent the bulk of the immature tryptase in nonanaphylactic sera. α-Protryptase remains enzymatically inactive, whereas some of β-protryptase is autoprocessed from the proform within the mast cell into the mature enzyme by a dipeptidase where it is stored in granules. Only upon activation of mast cells are both the pro and mature forms of tryptase secreted in parallel with prestored histamine and newly generated vasoactive mediators. Total tryptase in serum from healthy humans ranges from 1 to 11.4 μg/L (average 3 to 5 μg/L). Mature tryptase is normally undetectable (<1 μg/L) in serum from healthy individuals who have no history of anaphylaxis during the preceding hours. Elevated levels of total tryptase (>11.4 μg/L) can be detected in serum 1 to 4 hours after the onset of systemic anaphylaxis with hypotension. Baseline levels of >20 μg/L are detected in most individuals with systemic mastocytosis. Recommended serum collection times for tryptase quantification range from 30 minutes to 4 hours after the onset of an acute event. Because serological tests for mature tryptase are not widely available, it is important to compare an acute event total tryptase level (within 4 hours) with a baseline total tryptase 24 hours after all signs and symptoms of the event have subsided. Reported mature tryptase levels in postmortem cases of fatal anaphylaxis have ranged from 12 to 150 μg/L in all nine fatalities caused by *Hymenoptera* venom and in six of eight food-induced fatalities.

ON THE HORIZON

New trends in laboratory methodology for the assessment of human allergic diseases include the following:

- Although there is a transition from the use of crude allergen extracts to allergenic components and epitopes in immunoglobulin E (IgE) antibody serological assays, extracts remain the principal allergen reagent source for the foreseeable future because of their comprehensive nature.
- Multiplexing platforms are being increasingly developed to allow rapid simultaneous detection of IgE antibodies to many allergenic components using microliter quantities of serum. However, due to their semiquantitative nature, fixed panel of specificities, and lower analytical sensitivity, single-plex assays will remain the most cost-effective and widely used assay format for the foreseeable future.
- Qualitative "point of care" IgE antibody assays to rapidly assess sensitization to aeroallergens during the patient's visit will be slow to be adopted because of their limited/fixed allergen menus, lower diagnostic sensitivity, and concern about potential patient misinterpretation of results.

The presence of IgE antibody whether detected by in vivo or in vitro assays is simply a confirmatory measurement for sensitization. Evidence of sensitization supports a patient's medical, family, and environmental histories of a temporal association between allergic symptoms and allergen exposure. For patients with allergies to inhaled allergens, both in vivo and in vitro diagnostic assays have equivalent utility in identifying sensitization. However, for those facing life-threatening anaphylactic reactions caused by venom and drug allergens, in vivo IDST and provocation methods are favored. In cases where antihistamines, β-receptor stimulants, or high-dose steroids reduce

an in vivo provocation test's measured response, as well as in children, pregnant women, and older patients, in whom skin testing may not be well tolerated and when dealing with allergens (e.g., foods, molds) where commercial extracts can be highly variable or labile, serological methods have distinct advantages. Above all, the presence of allergen-specific IgE antibody is necessary, but not sufficient, to identify an atopic state and is thus considered only one of the important risk factors that must be considered in the diagnosis of human allergic disease.

REFERENCES

1. Prausnitz C, Kustner H. Studine uber die Ueberemfindlichkeil. *Zentralbl Bakteriol Mikrobiol Hyg*. 1921;86:160 [in German].
2. Ishizaka K, Ishizaka T. Physiochemical properties of reaginic antibody. I. Association of reaginic activity with an immunoglobulin other than gamma A or gamma G globulin. *J Allergy*. 1967;37:169.
3. Johansson SGO. Raised levels of a new immunoglobulin class (IgND) in asthma. *Lancet*. 1967;2:951.
4. Hamilton RG. The science behind the discovery of IgE. *J Allergy Clin Immunol*. 2005;115:648.
5. Hamilton RG. Human immunoglobulins. In: O'Gorman MRG, Donnenberg AD, eds. *Handbook of Human Immunology*. 2nd ed. Boca Raton, FL: CRC Press; 2008. pp. 63–106.
6. Barbee RA, Halomen M, Lebowitz M, et al. Distribution of IgE in a community population sample: correlations with age, sex and allergen skin test reactivity. *J Allergy Clin Immunol*. 1981;68:106.
7. Kleine-Tebbe J, Poulsen LK, Hamilton RG. Quality management in IgE-based allergy diagnostics. *J Lab Med*. 2016;40:81–96.
8. Wittig HJ, Belloit J, DeFillippi I, et al. Age-related serum IgE levels in healthy subjects and in patients with allergic disease. *J Allergy Clin Immunol*. 1980;66:305.
9. Dati F, Ringel KP. Reference values for serum IgE in healthy nonatopic children and adults. *Clin Chem*. 1982;28:1556.
10. Saini SS, Bindslev-Jensen C, Maurer M, et al. Efficacy and safety of omalizumab in patients with chronic idiopathic/spontaneous urticaria who remain symptomatic on H1 antihistamines: a randomized, placebo-controlled study. *J Invest Dermatol*. 2015;135:67–75.
11. Wide L, Bennich H, Johansson SGO. Diagnosis by an in vitro test for allergen specific antibodies. *Lancet*. 1967;2:1105.
12. Hamilton RG, Matsson PNJ, Chan S, et al. *Analytical Performance Characteristics, Quality Assurance and Clinical Utility of Immunological Assays for Human Immunoglobulin E (IgE) Antibodies of Defined Allergen Specificities. I/LA20-A3, International CLSI-Guideline*. 3rd ed. Wayne, PA: Clinical Laboratory Standards Institute; 2016.
13. Hamilton RG, Oppenheimer J. Serological IgE analyses in the diagnostic algorithm for allergic disease. *J Allergy Clin Immunol Pract*. 2015;3:833–40.
14. Adkinson NF Jr, Hamilton RG. Clinical history-driven diagnosis of allergic diseases: utilizing in vitro IgE testing. *J Allergy Clin Immunol Pract*. 2015;3:871–76.
15. Hamilton RG. Allergic sensitization is a key risk factor for but not synonymous with allergic disease. *J Allergy Clin Immunol*. 2014;134:360–1.
16. Hamilton RG. Diagnostic in vivo and in vitro methods in insect allergy. In: Freeman T, Tracy J, eds. *Stinging Insect Allergy: A Clinician's Guide*. NY: Springer Science; 2017. pp. 85–99.
17. Guerin B, Watson RD. Skin tests. *Clin Rev Allergy*. 1988;6:211.
18. Mirela Chiriac A, Bousquet J, Demoly P, et al. In vivo methods for study and diagnosis of allergy: skin tests, techniques and interpretation. In: Adkinson NF Jr, Bochner BS, Burks AW, eds. *Middleton's Allergy: Principles and Practice*. 8th ed. Philadelphia, PA: Elsevier Saunders; 2014.
19. Tversky JR, Chelladurai Y, McGready J, et al. Performance and pain tolerability of current diagnostic allergy skin prick test devices. *J Allergy Clin Immunol Pract*. 2015;3:888–93.
20. Turkeltaub PC, Rastogi SC, Baer H, et al. A standardized quantitative skin test assay of allergen potency and stability. Studies on the allergen dose-response curve and effect of wheal, erythema and patient selection on assay results. *J Allergy Clin Immunol*. 1982;70:343.
21. Slater JE, Menzies SL, Bridgewater J, et al. The US Food and Drug Administration review of the safety and effectiveness of nonstandardized allergen extracts. *J Allergy Clin Immunol*. 2012;129:1014–19.
22. McCann WA, Ownby DR. The reproducibility of the allergy skin test scoring and interpretation by board-certified/board-eligible allergists. *Ann Allergy Asthma Immunol*. 2002;89:368–71.
23. Cockcroft DW. Bronchial challenge testing. In: Adkinson NF Jr, Bochner BS, Burks AW, eds. *Middleton's Allergy: Principles and Practice*. 8th ed. Philadelphia, PA: Elsevier Saunders; 2014.
24. Scadding GW, Hansel TT, Durham SR. Nasal provocation testing. In: Adkinson Jr NF, Bochner BS, Burks AW, eds. *Middleton's Allergy: Principles and Practice*. 8th ed. Philadelphia, PA: Elsevier Saunders; 2014.
25. Wood RA. Oral food challenge testing. In: Adkinson NF Jr, Bochner BS, Burks AW, eds. *Middleton's Allergy: Principles and Practice*. 8th ed. Philadelphia, PA: Elsevier Saunders; 2014:1357.
26. Stapel SO, Asero R, Ballmer-Weber BK, et al. Testing for IgG4 against foods is not recommended as a diagnostic tool: EAACI Task Force Report. *Allergy*. 2008;63:793–6.
27. Sampson HA. Utility of food specific IgE concentrations in predicting symptomatic food allergy. *J Allergy Clin Immunol*. 2001;107:891.
28. Participants summary report. College of American pathologists, diagnostic allergy (SE) survey, Cycle B. Northfield, IL: College of American Pathologists; 2015.
29. Hamilton RG. Accuracy of US Food and Drug Administration cleared IgE antibody assays in the presence of anti-IgE (omalizumab). *J Allergy Clin Immunol*. 2006;117:759–66.
30. Radauer C, Bublin M, Wagner S, et al. Allergens are distributed into few protein families and possess a restricted number of biochemical functions. *J Allergy Clin Immunol*. 2008;121:847–52.
31. Kleine-Tebbe J, Matricardi PM, Hamilton RG. Allergy work-up including component-resolved diagnosis: how to make allergen-specific immunotherapy more specific. *Immunol Allergy Clin North Am*. 2016;36:191–203.
32. European Academy of Asthma Allergy and Clinical Immunology working group. Matricardi P, Kleine-Tebbe J, Ollert M, et al, editors. *Handbook on Molecular Allergology. Pediatr Allergy Immunology*. 2016;23:1–250.
33. Kober A, Perborn H. Quantitation of mouse-human chimeric allergen-specific IgE antibodies with ImmunoCAP technology. *J Allergy Clin Immunol*. 2006;117:S219 [Abstract 845].
34. Merrett J, Merrett TG. Phadiatop—a novel antibody screening test. *Clin Allergy*. 1987;17:409–16.
35. Hiller R, Laffer S, Harwanegg C, et al. Microarrayed allergen molecules: diagnostic gatekeepers for allergy treatment. *FASEB J*. 2002;16:414–6.
36. Martínez-Aranguren R, Lizaso MT, Goikoetxea MJ, et al. Is the determination of specific IgE against components using ISAC 112 a reproducible technique? *PLoS ONE*. 2014;9:e88394.
37. James LK, Shamji MH, Walker SM, et al. Long-term tolerance after allergen immunotherapy is accompanied by selective persistence of blocking antibodies. *J Allergy Clin Immunol*. 2011;127:509–16.
38. Golden DBK, Lawrence ID, Hamilton RG, et al. Clinical correlation of the venom-specific IgG antibody level during maintenance venom immunotherapy. *J Allergy Clin Immunol*. 1992;90:386–91.
39. Craig CS, Schwartz LB. Tryptase and chymase: markers of distinct types of human mast cells. *Immunol Res*. 1989;8:130.
40. Hamilton RG. Monitoring allergic patients on omalizumab with free and total serum IgE measurements. *J Allergy Clin Immunol Pract*. 2016;4:366–8.

Selected CD Molecules and Their Characteristics

Thomas A. Fleisher

CD Molecule	Predominant Distribution	Identity/Function
CD1a–e	Thymocytes, subset of lymphocytes, antigen-presenting cells	MHC class I–like molecules; presentation of nonpeptide antigens to T cells; thymic T-cell development
CD2	T cells	Binds to LFA-3; receptor for CD48; T-cell activation; adhesion
CD3	T cells	T-cell signaling complex; associated with TCR
CD4	T-cell subset	TCR coreceptor; interacts with MHC class II molecules on antigen-presenting cells; identifies T cells with helper function; signal transduction
CD5	Most T cells, thymocytes, B-cell subset	Binds to CD72; regulation of cell proliferation/activation; identifies B-1 B-cell subset
CD6	Thymocytes, T cells, B-cell subset	Binds CD166 (ALCAM); adhesion; mediates binding of developing thymocytes with thymic epithelial cells; thymic development; T-cell activation
CD7	Pluripotent hematopoietic cells, thymocytes, T cells	T-cell and NK-cell development
CD8	T-cell subset	TCR coreceptor; interacts with MHC class I molecules on antigen-presenting cells; identifies T cells with cytotoxic function
CD10	B cells	Neutral endopeptidase; enkephalinase; B-cell development; common acute lymphoblastic leukemia antigen (CALLA), neutrophils
CD11a	Leukocytes	α-chain of LFA-1; pairs with CD18; interacts with ICAM; adhesion and cellular migration
CD11b	Monocytes, granulocytes, NK cells	α-chain of complement receptor type 3 (CR3); pairs with CD18; adhesion molecule
CD11c	Monocytes, granulocytes, NK cells	α-chain of complement receptor type 4 (CR4); pairs with CD18; adhesion molecule
CD13	Hematopoietic stem cells, immature and mature myeloid and monocyte elements	Appears before CD33 during myeloid differentiation
CD14	Granulocytes, monocytes/macrophages	Receptor for LPS/LPB complex; myeloid differentiation antigen; cell activation
CD15	Neutrophils, eosinophils, monocytes, basophil subset	Sialyl Lewis X antigen, plays a role in cell adhesion, defective in leukocyte adhesion deficiency type 2
CD16a,b	NK cells, monocytes/macrophages, neutrophils	FcγRIIIA and FcγRIIIB (low-affinity IgG receptors-type III); phagocytosis; ADCC
CD18	Leukocytes	β-chain of β_2-integrin molecules, including LFA-1, CR3, and CR4; pairs with CD11a, b, and c
CD19	B cells	BCR coreceptor; signal transduction; complexes with CD21, expressed pre-B cells
CD20	B cells	Role in B-cell activation/differentiation
CD21	B cells; follicular dendritic cells	Complement receptor type 2 (CR2): C3d receptor; B-cell coreceptor subunit; EBV receptor
CD22	B cells	Associates with BCR; signaling; regulation of B-cell activation; adhesion
CD23	B cells, macrophages, eosinophils, platelets, follicular dendritic cells	FcεRII; (low-affinity IgE receptor), activated B cells
CD24	Leukocytes	Heat-stable antigen; costimulation; adhesion
CD25	Activated T cells and B cells	α-chain of IL-2 receptor, low-affinity IL-2 binding; signaling for cell proliferation/differentiation
CD26	Activated T cells and B cells; macrophages	Dipeptidyl peptidase; role in extracellular adhesion; cell activation
CD27	T cells; B-cell subset	Costimulation; T-cell proliferation, memory B cells
CD28	T cells	Binds B7-1 (CD80) and B7-2 (CD86); T-cell costimulation; signal transduction
CD29	Leukocytes	Integrin β_1-chain; pairs with CD49a-CD49f to form VLA-1-VLA-6 integrins, respectively; adhesion; signal transduction; development
CD30	Activated B cells and T cells	Binds to CD153; T-cell activation/regulation/differentiation, Reed–Sternberg cells
CD31	Monocytes, granulocytes, platelets, endothelial cells, B cells Recent thymic emigrants	PECAM-1; binds to CD38; adhesion; signal transduction
CD32	B cells, monocytes/macrophages, granulocytes, eosinophils	FcgRII (low-affinity IgG receptor-type II) phagocytosis; ADCC; B-cell regulation
CD33	Myeloid progenitors, granulocytes	Binds sialic acid
CD34	Hematopoietic progenitor cells; capillary endothelium	Mucosialin; binds to CD62L; adhesion

Continued

CD Molecule	Predominant Distribution	Identity/Function
CD35	Leukocytes; erythrocytes	Complement receptor 1 (CR1); C3b and C4b receptor; phagocytosis
CD36	Monocytes/macrophages; endothelium; platelets	Binds oxidized LDL; scavenger receptor; binds apoptotic cells; adhesion and endocytic receptor; GPIIIb; platelet adhesion and aggregation
CD38	NK cells; T-cell and B-cell subsets; monocytes	Binds CD31; cyclase; hydrolase; cell activation
CD40	B cells; antigen-presenting cells	Binds CD154 (CD40 ligand); B-cell proliferation, differentiation, and survival; T-cell costimulation
CD41	Megakaryocytes/platelets	Glycoprotein IIb; a IIb integrin chain; binds fibronectin, fibrinogen, von Willebrand factor, thrombospondin; extracellular adhesion; platelet aggregation
CD43	Leukocytes (except resting B cells)	Leukocyte sialoglycoprotein; may bind CD54; signal transduction; adhesion; antiadhesion
CD44	Leukocytes; memory T cells; erythrocytes	Binds hyaluronan (H-CAM), collagen, fibronectin, laminin, osteopontin; extra- and intercellular adhesion; T-cell costimulation; leukocyte homing
CD45	Leukocytes (pan-leukocyte marker)	Protein tyrosine phosphatase; cell differentiation; lymphocyte signal transduction and activation
CD45RA	Naive T-cell marker (in conjunction with CD62L); B cells, monocytes	CD45 isoform
CD45RB	B-cell and T-cell subsets, monocytes/macrophages, granulocytes	CD45 isoform
CD45RO	Memory and activated T cells; B cells, monocytes/macrophages	CD45 isoform
CD46	Hematopoietic cells	Membrane cofactor protein (MCP); binds to C3b and C4b and regulates complement pathway
CD47R	Leukocytes; endothelium	Integrin-associated protein (IAP); leukocyte migration, extravasation, and activation
CD48	Leukocytes (not neutrophils)	Binds CD2; adhesion; costimulation
CD49a–f	Various distributions	Integrin a1-6 chain; binds to CD29 to form VLA-1 to VLA-6; binds extracellular matrix components such as fibronectin, laminin, collagen (CD49D binds VCAM-1, fibronectin, MAdCAM-1, invasin); lymphocyte homing; extracellular adhesion; embryonic development
CD50	Thymocytes, B cells, T cells, monocytes, granulocytes	ICAM-3; adhesion
CD51	Platelets/megakaryocytes, granulocytes, monocytes, T cells	Integrin α-chain; associates with CD61; binds vitronectin, fibronectin, fibrinogen; extracellular adhesion; T-cell costimulation; epithelial inflammatory response
CD52	Leukocytes	GPI linked, signaling, defined by CAMPATH-1
CD54	Broad distribution; increased on activated leukocytes	ICAM-1; binds LFA-1; adhesion; leukocyte transendothelial migration, rhinovirus receptor
CD55	Hematopoietic cells and some nonhematopoietic cells	Decay accelerating factor (DAF); binds complement fragment C3b; regulation of complement activation
CD56	NK cells, NK-T cells	NKH-1; adhesion
CD57	NK cells, T-cell subset, B cells, monocytes	Oligosaccharide expressed on cell surface glycoproteins
CD58	Hematopoietic cells and non-hematopoietic cells	LFA-3; binds CD2; adhesion; lymphocyte coactivation
CD59	Hematopoietic cells and non-hematopoietic cells	Binds complement components C8 and C9 and regulates assembly of complement membrane attack complex
CD61	Platelets/megakaryocytes, macrophages	Integrin β_3 subunit; associates with CD41 or CD51
CD62E	Endothelium	ELAM-1 or E-selection; binds sialyl-Lewis X; adhesion; mediates rolling interaction of neutrophils on endothelium and neutrophil extravasation
CD62L	B cells, T cells, monocytes, NK cells	LECAM-1, LAM-1 (or L-selectin); binds CD34 and GlyCAM; adhesion; mediates rolling interactions with endothelium and call extravasation
CD62P	Platelets/megakaryocytes, endothelium	P-selectin; binds sialyl-Lewis X; mediates interaction of platelets with neutrophils and monocytes; mediates rolling interaction of neutrophils with endothelium
CD64	Monocytes/macrophages, mature neutrophils	FcγR1 (high-affinity IgG receptor)
CD65	Neutrophils, eosinophils, basophils monocyte subset, CD56 bright NK cells	During myeloid differentiation, it appears after myeloperoxidase, appears to be the ligand for CD62L
CD66c	Differentiating myeloid cells peaks at promyelocyte stage, subpopulation of monocytes	Member of the carcinoembryonic antigen (CEA) family
CD68	Monocytes/macrophages	Macrosialin; early activation antigen; role in phagocytic activity
CD69	Activated T cells, B cells, macrophages, NK cells	Early activation antigen; costimulation
CD70	Activated B cells and T cells; macrophages	Binds to CD27; costimulation
CD71	Activated leukocytes, erythroid precursors	Transferrin receptor; cell activation
CD72	B cells	Ligand for CD5; B-cell activation and differentiation; costimulation
CD73	B-cell and T-cell subsets	Ecto-5′ nucleotidase; allows nucleoside uptake
CD74	MHC class II expressing cells	MHC class II–associated invariant chain; involved in antigen processing and peptide presentation in antigen-presenting cells
CD77	Germinal center B cells	Entering apoptosis

CD Molecule	Predominant Distribution	Identity/Function
CD79a,b	B cells	Iga, Igb; components of BCR complex that mediate signal transduction
CD80	Monocytes/macrophages, dendritic cells, activated B cells	B7-1; ligand for CD28 and CTLA-4; T-cell interaction with antigen-presenting cells; costimulation
CD81	Lymphocytes	Associates with CD19 and CD21 to form B-cell coreceptor complex; costimulation; adhesion
CD86	Monocytes, activated B cells	B7-2; ligand for CD28 and CTLA-4; T-cell interaction with antigen-presenting cells; costimulation
CD87	Granulocytes; monocytes/macrophages, activated T cells	Urokinase plasminogen activator receptor
CD88	Granulocytes, macrophages, mast cells	Complement component fragment C5a receptor
CD89	Monocytes/macrophages, granulocytes, B-cell and T-cell subsets	FcαR (IgA receptor)
CD91	Monocytes	α_2-macroglobulin receptor
CD94	NK cells, T-cell subset	Inhibits killing
CD95	Broad distribution	Fas or APO-1; induces apoptosis after being bound by Fas ligand
CD97	Activated B and T cells, monocytes, PMN	Activation antigen
CD102	Resting lymphocytes, monocytes, endothelial cells	ICAM-2; binds LFA-1 (CD11a/CD18); adhesion; T-cell costimulation
CD103	Intraepithelial lymphocytes, T-cell subset	αE integrin; T-cell development and costimulation
CD104	Epithelial cells, Schwann cells	β_4-integrin; epidermal adhesion to basement membrane
CD105	Endothelial cells, bone marrow cell subset, activated macrophages	Endoglin; adhesion
CD106	Endothelial cells	VCAM-1; ligand for VLA-4; lymphocyte adhesion; embryonic development
CD107a,b	Epithelial cells, subsets of monocytes, granulocytes, and lymphocytes	LAMP-1 and LAMP-2; adhesion
CD110	Hematopoietic stem cells, megakaryocytes, platelets	Thrombopoietin receptor; megakaryocyte proliferation and differentiation
CD114	Granulocytes	G-CSF receptor; regulates granulopoiesis
CD115	Monocytes/macrophages	M-CSF receptor; cell differentiation
CD116	Monocytes, neutrophils, eosinophils	GM-CSF a chain receptor; cell differentiation
CD117	Hematopoietic progenitors, mast cells	c-*kit*; stem cell factor (SCF) receptor; hematopoietic cell differentiation
CD118	Broad distribution	Type 1 interferon (interferon-α/β) receptor
CD119	Broad distribution	Interferon-γ receptor
CD120a	Hematopoietic cells, nonhematopoietic cells, myeloid cells	TNF receptor-type I signal transduction; apoptosis
CD120b	Hematopoietic cells, nonhematopoietic cells, myeloid cells	TNF receptor-type II; signal transduction; apoptosis
CD121a	Thymocytes, T-cell subset, fibroblasts, epithelial cells, and brain cells	IL-1 receptor-type I; signal transduction
CD121b	T-cell subset, myeloid cell subsets	IL-1 receptor-type II
CD122	NK cells, T-cell and B-cell subset	IL-2 and IL-15 receptor β chain; signal transduction; regulation of lymphocyte development, differentiation, activation, and proliferation
CD123	Bone marrow stem cells, granulocytes, monocytes, megakaryocytes	IL-3 receptor α chain; cell development and differentiation
CD124	Mature B cells and T cells, hematopoietic precursor cells	IL-4 receptor; signal transduction; lymphocyte development, activation, differentiation, and proliferation
CD125	Eosinophils, basophils, B-cell subset	IL-5 receptor; eosinophil and B-cell growth and differentiation
CD126	Activated B cells, plasma cells, T cells, granulocytes, monocytes/macrophages; also expressed on epithelial cells, fibroblasts, hepatocytes, and neural cells	IL-6 receptor α chain; regulation of B-cell and T-cell differentiation and function; hematopoiesis
CD127	Bone marrow lymphoid precursors, pro-B cells, T-cell precursors, T-cell subset, monocytes	IL-7 receptor α chain; signal transduction; B-cell and T-cell proliferation and differentiation
CD128	Neutrophils, basophils, T-cell subset	IL-8 receptor; neutrophil activation and migration
CD129	T cells	IL-9 receptor α chain; T-cell proliferation
CD130	Broad distribution	IL-6 receptor β chain (with CD126); signal transduction
CD131	Lymphocytes, granulocytes, monocytes	IL-3, IL-5, and GM-CSF receptor; common β chain; signal transduction; see CD123 and CD125
CD132	Lymphocytes	Common γ chain of high-affinity receptor for IL-2 (with CD25 and CD122), IL-4 (with CD124), IL-7 (with CD127), IL-9 (with CD129), and IL-15 (with CD122) receptors; signal transduction
CD134	Activated T cells	OX-40 antigen of TNFR superfamily (binds OX-40 ligand); T-cell–B-cell interaction and T-cell costimulation
CD135	Lymphoid and myeloid cell progenitor subsets	Flt3 ligand receptor; development of myeloid and lymphoid progenitors
CD137	Activated T cells	4-1BB; binds 4-1BB ligand and extracellular matrix components; T-cell–B-cell interaction and T-cell costimulation; extracellular adhesion; signal transduction

Continued

CD Molecule	Predominant Distribution	Identity/Function
CD138	B-cell subset, plasma cells, epithelial cells	Syndecan-1; binds interstitial matrix proteins; B cell–matrix interactions
CD140a,b	Endothelial cells	PDGF receptor α and β chain; embryonic development; signal transduction; chemotaxis
CD141	Endothelium	Thrombomodulin (binds thrombin); regulates coagulation
CD142	Endothelium	Tissue factor; binds plasma factors VII/VIIa; hemostasis, coagulation, and angiogenesis
CD143	Endothelium	Angiotensin-converting enzyme (ACE); binds angiotensin 1; regulates blood pressure
CD144	Endothelium	VE-cadherin; cell–cell adhesion; maintenance of endothelium integrity
CD146	Activated T-cell subset	Mel-CAM, adhesion molecule during development
CD150	T-cell and B-cell subsets	Surface lymphocyte activation marker (SLAM); B cell–T cell interaction; costimulation
CD151	Not defined	PETA-3; regulates platelet aggregation and mediator release
CD152	Activated T cells	CTLA-4; binds B7-1 (CD80) and B7-2 (CD86); T-cell costimulation–negative signal
CD153	T cells	CD30 ligand; T-cell activation, differentiation, and regulation
CD154	Activated T cells	CD40 ligand; T-cell costimulation
CD156a	Leukocytes, B cells	Transmembrane glycoprotein; disintegrin and metalloproteinase domain (ADAM) family member; leukocyte adhesion and protease function; infiltration of myelomonocytic cells
CD156b	Broad distribution	TNF-α-converting enzyme (TACE); disintegrin and metalloproteinase domain (ADAM) family member; cleaves TNF and transforming growth factor-α from cell surface, thereby releasing soluble form
CD158	NK cells	Killer immunoglobulin-like receptors (KIR); family of molecules that inhibit NK cytotoxic activity
CD159a	NK cells	NKG2A (killer cell lectin-like receptor)
CD161	NK cells	Natural killer cell receptor-P1; target cell recognition; NK cell activation
CD162	Granulocyte and T-cell subsets	P-selectin glycoprotein ligand-1 (PSGL-1); adhesion
CD166	Activated T cells, B cells	ALCAM; binds CD6; T-cell activation; thymocyte development
CD167a	Epithelial cells	Discoidin domain receptor 1 (DDD1); tyrosine kinase receptor; binds to collagen; cell–cell contact and adhesion
CD178	Activated T cells; various tissue cells	FAS ligand (ligand for CD95); binding to FAS triggers apoptosis
CD179a	Pro-B and pre-B cells	VpreB; forms surrogate light chain with CD179b; early B-cell differentiation
CD179b	Pro-B and pre-B cells	λ5; forms surrogate light chain with CD179a; early B-cell differentiation
CD180	B cells	RP105; toll-like receptor family; regulates B-cell recognition and signaling of LPS
CD183	Effector/memory T cells, NK cells, eosinophils	CXCR3 receptor for interferon-inducible chemokines IP10, Mig, and I-TAC; chemotactic migration of effector T cells into areas of inflammation
CD184	Leukocytes; hematopoietic progenitors	CXCR4 receptor for chemokines such as stromal cell-derived factor 1 (SDF-1) (fusin); chemotaxis; HIV-1 coreceptor
CD195	Broad distribution; myeloid cells, lymphocytes, T lymphocytes, neurons, epithelium, endothelium	CCR5 receptor for chemokines such as macrophage inflammatory proteins, MIP-1a and MIP-1b, and RANTES; chemotaxis; HIV-1 coreceptor
CDw197	Lymphoid tissues, B cells, T-cell subset	CCR7 chemokine receptor; chemotaxis; T-cell homing and migration
CD201	Endothelial cells	Protein C receptor; coagulation
CD203c	Mast cells, basophils	Member of the ectonucleotide pyrophosphatase/phosphodiesterase enzymes
CD204	Myeloid cells, monocytes/macrophages	Macrophage scavenger receptor-1 (MSR1); mediates binding, internalization, and processing of various negatively charged macromolecules
CD206	Dendritic cells, macrophages, myeloid cells, endothelial cells	Mannose receptor, C-type 1; binds microorganisms; phagocytosis
CD207	Dendritic cells, Langerhans cells	Langerin; mannose receptor; phagocytosis and internalization of antigen for processing
CD208	Dendritic cells	DC-LAMP
CD209	Dendritic cells	DC-SIGN
CDw210	Broad distribution	I1-10 receptor α and β chain
CD212	T cells, NK cells	IL-12 receptor β_1 chain
CD213a1, a2	Lymphocytes, bronchial epithelial and smooth muscle cells	IL-13 receptor α_1 and α_2 chains
CDw217	Activated T-cell subset	IL-17; cytotoxic T-lymphocyte–associated serine esterase 8; stimulates cell activation; induces osteoclast differentiation factor (ODF)
CD220	Broad distribution	Insulin receptor; stimulates glucose uptake
CD221	Broad distribution	Insulin-like growth factor 1 receptor; cell signaling; activation, and differentiation
CD222	Broad distribution	Mannose-6-phosphate receptor; insulin-like growth factor 2 receptor
CD226	NK cells, platelets, monocytes, T-cell subset	Platelet and T-cell activation antigen 1 (PTA1); adhesion
CD233–241	Erythrocytes	Various erythrocyte membrane antigens, including blood group–associated glycoproteins
CD242	Erythrocytes	ICAM-4
CD246	T cells	TCR or CD3 ζ chain; associated with TCR and CD3; couples TCR recognition with T-cell signaling
CD247	T cells	T cells ζ chain of the TcR
CD252	Activated B cells	OX40 ligand
CD253	Activated T cells	TRAIL, death receptor
CD254	Activated T cells, LN and BM stroma	RANK ligand
CD256	Monocytes, macrophages	APRIL, binds TACI and BCMA
CD257	Activated monocytes	BLyS, BAFF, binds TACI, BCMA, BAFFR, induces B-cell proliferation

CD Molecule	Predominant Distribution	Identity/Function
CD261	Activated T cells, peripheral leukocytes	TRAIL-R2, DR5, death receptor
CD262	Peripheral lymphocytes	TRAIL-R1, DR4, death receptor
CD263	Peripheral lymphocytes	TRAIL-R3, DcR1, death receptor
CD264	Peripheral lymphocytes	TRAIL-R4, DcR2, death receptor
CD265	Broad distribution	RANK
CD267	B cells, activated T cells	TACI
CD268	B cells	BAFFR, binds BLys, mature B-cell survival
CD269	Mature B cells	BCMA, binds APRIL and BAFF, B-cell survival and proliferation
CD275	B cells, dendritic cells, monocytes	ICOSL, costimulation, cytokine production
CD278	Activated T cells	ICOS, T-cell costimulation

ADCC, Antibody-dependent cellular cytotoxicity; *ALCAM*, activated leukocyte cell adhesion molecule; *APRIL*, a proliferation-inducing ligand; *BCMA*, B-cell maturation antigen; *BCR*, B-cell receptor for antigen; *DC*, dendritic cell; *CTLA*, cytotoxic T lymphocyte antigen; *EBV*, Epstein-Barr virus; *ELAM*, endothelial leukocyte adhesion molecule; *G-CSF*, granulocyte–colony-stimulating factor; *GM-CSF*, granulocyte macrophage–colony-stimulating factor; *GPI*, glycosyl-phosphatidylinositol; *HIV*, human immunodeficiency virus; *ICAM*, intercellular adhesion molecule; *ICOS*, inducible T-cell costimulator; *Ig*, immunoglobulin; IL, interleukin; *LAM*, leukocyte adhesion molecule; *LAMP*, latent membrane protein; *LDL*, low-density lipoproteins; *LECAM*, lymphocyte endothelial cell adhesion molecule; *LFA*, lymphocyte function antigen; *LPB*, lipopolysaccharide binding protein; *LPS*, lipopolysaccharide; *MAdCAM-1*, mucosal addressin cell adhesion molecule-1; *M-CSF*, macrophage colony-stimulating factor; *MHC*, major histocompatibility complex; *NK cells*, natural killer cells; *PDGF*, platelet-derived growth factor; *PECAM*, platelet endothelial cell adhesion molecule; *PETA*, platelet-endothelial cell tetraspan antigen; *PMN*, polymorphonuclear neutrophil; *SIGN*, specific intercellular adhesion molecule–grabbing nonintegrin; *TACI*, transmembrane activator and CAML interactor; *TCR*, T-cell receptor for antigen; *TNF*, tumor necrosis factor; *TNFR*, tumor necrosis factor receptor; *TRAIL*, TNF-related apoptosis inducing ligand; *VCAM*, vascular cellular adhesion molecule; *VLA*, very late antigen.

This list was adapted from the results of the Eighth International Workshop on Human Leukocyte Differentiation Antigens (HLDA8) held in Adelaide, Australia, in December 2004 (Proceedings of the 8th International Workshop on Human Leukocyte Differentiation Antigens. December 12–16, 2004. Adelaide, Australia. *Cell Immunol.* 2005;236[1–2]:1–187). Engel P, Boumsell L, Balderas R, et al. CD nomenclature 2015: human leukocyte differentiation antigen workshops as a driving force in immunology. *J Immunol.* 2015;195:4555–4563.

Laboratory Reference Values

Thomas A. Fleisher

TABLE A2.1 Immunoglobulin Levels (Age-Related Reference Ranges; see Chapter 5)

AGE OF HEALTHY DONORS	IGG G/L	IGG_1 G/L	IGG_2 G/L	IGG_3 G/L	IGG_4 G/L	IGA G/L	IGA_1 G/L	IGA_2 G/L	IGM G/L
0 to <5 months	1.0–1.34	0.56–2.15	≤0.82	0.07.6–8.23	≤0.198	0.07–0.37	0.10–0.34	0.004–0.055	0.26–1.22
5 to <9 months	1.64–5.88	1.02–3.69	≤0.89	0.119–0.740	≤0.208	0.16–0.50	0.14–0.41	0.015–0.062	0.32–1.32
9 to <15 months	2.46–9.04	1.60–5.62	0.24–0.98	0.173–0.637	≤0.220	0.27–0.66	0.20–0.50	0.028–0.070	0.40–1.43
15 to <24 months	3.13–11.70	2.09–7.24	0.35–1.05	0.219–0.550	≤0.230	0.36–0.79	0.24–0.58	0.039–0.077	0.46–1.52
2 to <4 years	2.95–11.56	1.58–7.21	0.39–1.76	0.170–0.847	0.004–0.491	0.27–2.46	0.16–1.62	0.013–0.311	0.37–1.84
4 to <7 years	3.86–14.70	2.09–9.02	0.44–3.16	0.108–0.949	0.008–0.819	0.29–2.56	0.17–1.87	0.011–0.391	0.37–2.24
7 to <10 years	4.62–16.82	2.53–10.19	0.54–4.35	0.085–10.26	0.010–1.087	0.34–2.74	0.21–2.21	0.014–0.480	0.38–2.51
10 to <13 years	5.03–15.80	2.80–10.30	0.66–5.02	0.115–10.53	0.010–1.219	0.42–2.95	0.27–2.50	0.026–0.534	0.41–2.55
13 to <16 years	5.09–15.80	2.89–9.34	0.82–5.16	0.200–10.32	0.007–1.217	0.52–3.19	0.36–2.75	0.047–0.551	0.45–2.44
16 to <18 years	4.87–13.27	2.83–7.72	0.98–4.86	0.313–0.976	0.003–1.110	0.60–3.37	0.44–2.89	0.066–0.543	0.49–2.01
≥18 years	7.67–15.90	3.41–8.94	1.71–6.32	0.184–10.60	0.024–1.210	0.61–3.56	0.50–3.14	0.097–1.560	0.37–2.86

Immunoglobulin (Ig) levels were assessed in serum by nephelometry, and data were statistically analyzed for the mid-95% confidence interval. For total IgG and IgG subclass quantitation, data from 156 pediatric and 92 adult donors were used; for total IgA quantitation, data from 201 pediatric and 99 adult donors were used; for IgA subclasses, data from 119 pediatric and 99 adult donors were used; and for IgM quantitation, data from 212 pediatric and 401 adults were used at Mayo Medical Laboratories.

TABLE A2.2 Total Serum IgE (IU/mL)

Age	Gender	Geometric Mean	Upper 95% Confidence Limit
6–14 years	M	42.7	527
	F	43.3	344
15–24 years	M	33.6	447
	F	18.6	262
25–34 years	M	16.8	275
	F	16.6	216
35–44 years	M	21.7	242
	F	19.3	206
45–54 years	M	19.2	254
	F	13.3	177
55–64 years	M	21.3	354
	F	11.7	148
65–74 years	M	21.2	248
	F	11.5	122
>75 years	M	18.4	219
	F	9.2	124
6–75 years	all M	22.9	317
	all F	14.7	189

Data were generated using individuals with negative skin prick tests (i.e., house dust mite, Bermuda grass, tree mix, weed mix, mold mix).
From Barbee RA, Halonen M, Lebowitz M, Burrows B. Distribution of IgE in a community population sample: correlations with age, sex, and allergen skin test reactivity. *J Allergy Clin Immunol.* 1981;68(2):106–11, with permission.

TABLE A2.3 Lymphocyte Immunophenotype (see Chapter 93): Adult Reference Range Expressed as 95% Confidence Intervals

Surface Antigens	Percent Positive	Cells/MM³
T cells		
CD3	55.3–87.7	651–2804
CD3/TcR alpha/beta	53.0–83.6	585–2716
CD3/TcR gamma/delta	0.6–16.6	11–288
CD3/CD4	27.9–55.	370–1336
CD3/CD8	13.6–46.2%	185–1024
CD4/CD8 ratio	0.81–4.00	
CD3/CD4–/CD8– TcR alpha/beta	0.2–1.2	3–36
CD3/CD4–/CD8– TcR gamma/delta	0.6–12.6	10–202
CD3/CD4/CD45RO	12.0–34.	231–668
CD3/CD4/CD45RA	5.0–31.4	66–914
CD3/CD4/CD62L+/CD45RA+	6.2–26.9	87–796
CD3/CD4/CD62L+/CD45RA–	9.1–28.6	166–544
CD3/CD4/CD62L–/CD45RA–	3.7–12.9	80–262
CD3/CD4/CD62L–/CD45RA+	0.1–5.5	1–132
CD3/CD4/CD31/CD45RA	2.6–25.3	47–527
CD3/CD4/CD45RS-/CXCR5+	1.3–6.6	28–123
CD3/CD4/CD25/FOXP3	1.3–5.5	27–122
CD3/CD8/CD45RO	2.2–14.	19–309
CD3/CD8/CD45RA	2.7–18.0	45–564
CD3/CD8/CD62L+/CD45RA+	2.3–18.2	37–484
CD3/CD8/CD62L+/CD45RA–	1.2–7.6	19–175
CD3/CD8/CD62L–/CD45RA–	2.0–12.4	47–383
CD3/CD8/CD62L–/CD45RA+	0.9–13.	17–274
CD3/CD56+ or CD16	2.1–18.0	56–448
B cells		
CD19	3.8–18	79–399
CD20	3.8–18	79–399
CD20/CD27	0.8–4.8	18–120
CD20/CD27/sIgM+	0.3–2.5	6–65
CD20/CD27/sIgM–	0.3–2.2	6–49
CD20/CD38	3.2–16.	66–358
CD20/CD10	0.1–2.6	1–64
NK cells		
CD56+ or CD16/CD3–	7.3–33.4	126–641

Data generated in the Flow Cytometry Section, Immunology Service, DLM, CC, NIH, Bethesda, MD. The 95% confidence interval for the WBC is 4300–9200/mm³.

TABLE A2.4 Age-Dependent Lymphocyte Immunophenotype Reference Range (80% Confidence Interval)

	T CELLS					
	CD3		CD4		CD8	
Age	Percent positive	Cells/mm³	Percent positive	Cells/mm³	Percent positive	Cells/mm³
0–3 months	53–84	2500–5500	35–64	1600–4000	12–28	560–1700
3–6 months	51–77	2500–5600	35–56	1800–4000	12–23	590–1600
6–12 months	49–76	1900–5900	31–56	1400–4300	12–24	500–1700
1–2 years	53–75	2100–6200	32–51	1300–4300	14–30	620–2000
2–6 years	56–75	1400–3700	28–47	700–2200	16–30	490–1300
6–12 years	60–76	1200–2600	31–47	650–1500	18–35	370–1100
12–18 years	56–84	1000–2200	31–52	530–1300	18–35	330–920

	CD4 T-CELL SUBPOPULATIONS					
	CD4/CD45RA		CD3/CD4/CD45RO		CD4/HLA-DR	
Age	Percent CD4 positive	Cells/mm³	Percent CD3/CD4 positive	Cells/mm³	Percent CD4 positive	Cells/mm³
0–3 months	64–95	1200–3700	2–22	60–900	2–6	40–180
3–6 months	77–94	1300–3700	3–16	120–630	2–10	60–280
6–12 months	64–93	1100–3700	5–18	160–800	2–11	50–260
1–2 years	63–91	1000–2900	7–20	210–850	2–11	70–280
2–6 years	53–86	430–1500	9–26	220–660	3–12	50–180
6–12 years	46–77	320–1000	13–30	230–630	3–13	40–120
12–18 years	33–66	230–770	18–38	240–700	4–11	30–100

	CD8 T-CELL SUBPOPULATIONS					
	CD8/CD45RA		CD3/CD4⁻/CD45RO		CD8/HLA-DR	
Age	Percent CD8 positive	Cells/mm³	Percent CD3/CD4-positive	Cells/mm³	Percent CD8 positive	Cells/mm³
0–3 months	80–99	450–1500	1–9	30–330	2–20	20–160
3–6 months	85–98	550–1400	1–7	30–290	3–17	30–170
6–12 months	75–97	480–1500	1–8	40–330	4–27	40–290
1–2 years	71–98	490–1700	2–12	60–570	6–33	60–600
2–6 years	69–97	380–1100	4–16	90–440	7–37	70–420
6–12 years	63–92	310–900	4–21	70–390	6–29	40–270
12–18 years	61–91	240–710	4–23	60–310	5–25	30–180

	B CELLS AND NK CELLS			
	CD19		CD3⁻/CD16⁻56⁺	
Age	Percent positive	Cells/mm³	Percent positive	Cells/mm³
0–3 months	6–32	300–2000	4–18	170–1100
3–6 months	11–41	430–3000	3–14	170–830
6–12 months	14–37	610–2600	3–15	160–950
1–2 years	16–35	720–2600	3–15	180–920
2–6 years	14–33	390–1400	4–17	130–720
6–12 years	13–27	270–860	4–17	100–480
12–18 years	6–23	110–570	3–22	70–480

Data generated by Shearer WT, Rosenblatt HM, Gelman RS, et al. Lymphocyte subsets in healthy children from birth through 18 years of age: the Pediatric AIDS Clinical Trials Group P1009 study. *J Allergy Clin Immunol.* 2003;112(5):973–980.

Additional pediatric lymphocyte reference intervals can be found in the following references:

- Piatosa B, Wolska-Kusnierz B, Pac M, et al. B cell subsets in healthy children: reference values for evaluation of B cell maturation process in peripheral blood. *Cytometry B Clin Cytom.* 2010;78(6):372–381.
- Schatorje EJH, Gemen EFA, Driessen GJA, et al. Paediatric reference values for the peripheral T cell compartment. *Scan J Immunol.* 2012;75(4):436–444.
- van Gent R, van Tilburg CM, Nibbelke EE, et al. Refined characterization and reference values of the pediatric T- and B-cell compartments. *Clin Immunol.* 2009;133(1):95–107.